

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

DMITRY TAUBINSKY

University of California at Berkeley and NBER

and

ALEX REES-JONES

The Wharton School, University of Pennsylvania and NBER

First version received August 2016; Editorial decision September 2017; Accepted November 2017 (Eds.)

This article shows that accounting for variation in mistakes can be crucial for welfare analysis. Focusing on consumer under-reaction to not-fully-salient sales taxes, we show theoretically that the efficiency costs of taxation are amplified by differences in under-reaction across individuals and across tax rates. To empirically assess the importance of these issues, we implement an online shopping experiment in which 2,998 consumers purchase common household products, facing tax rates that vary in size and salience. We replicate prior findings that, on average, consumers under-react to non-salient sales taxes—consumers in our study react to existing sales taxes as if they were only 25% of their size. However, we find significant individual differences in this under-reaction, and accounting for this heterogeneity increases the efficiency cost of taxation estimates by at least 200%. Tripling existing sales tax rates nearly doubles consumers' attention to taxes, and accounting for this endogeneity increases efficiency cost estimates by 336%. 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 behavioural welfare analysis.

Key words: Tax salience, Rational inattention, Deadweight loss, Welfare analysis.

JEL Codes: C9, D0, H0

1. INTRODUCTION

When incentive schemes are complex, or when decision-relevant attributes 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 *et al.*, 2009), shipping and handling charges (Hossain and Morgan, 2006), energy prices (Allcott, 2015), and out-of-pocket insurance costs (Abaluck and Gruber, 2011). Such studies typically estimate the “average mistake”, usually because inferring mistakes at the individual level is difficult or impossible with available data.

The editor in charge of this paper was Botond Koszegi.

Correspondingly, policy analysis building on these results often studies a representative agent committing the “average mistake”, and thus assumes that mistakes are homogeneous.

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

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

We present a series of results that generalize the canonical [Harberger \(1964\)](#) formula for the efficiency costs of taxation. We find that the efficiency cost of imposing a small tax in a previously untaxed market is increasing in the mean of the weights that the marginal consumers place on the tax when making purchasing decisions—thus, as in CLK, homogeneous under-reaction reduces efficiency costs.¹ However, we additionally show that inefficiency is increasing in the variance of misreactions to a degree of equal quantitative importance. The result arises because variation in mistakes across consumers generates misallocation of products. When under-reaction to the tax is homogeneous, the product is always purchased by those consumers who value it the most, and thus the market preserves the efficient sorting that is obtained with fully optimizing consumers. However, when consumers vary in their misreaction, purchasing decisions depend on both their valuation of the good *and* on their propensity to ignore the tax, thus breaking the efficient sorting property. The consequences of misallocation are particularly stark when supply is inelastic relative to demand and thus the equilibrium quantity purchased is relatively unaffected by taxation—a situation in which efficiency costs are low when consumers optimize perfectly but can be substantial in the presence of varying mistakes.

When evaluating “small” taxes, the mean and variance of marginal consumers’ misreaction—together with the price elasticity of demand—are sufficient statistics for computing efficiency costs. When considering increasing pre-existing taxes, however, accounting for how misreaction changes with the tax rate is crucial. If increases in the tax rate increase attention, and thus “debias” consumers, the distortionary effects of tax increases can be substantially higher than would otherwise be expected under the hypothesis that attention is exogenous. Intuitively, this is because consumers act as if prices have increased not only by the salient portion of

1. This result is derived in the absence of income effects, or under the assumption that the purchases in question constitute a small share of the budget. We maintain these assumptions throughout most of the article, as we have in mind products whose prices are small relative to consumers’ total earnings. However, CLK show that with income effects, under-reaction can sometimes generate larger efficiency costs when consumers make a big-ticket purchase due to the over-estimation of their remaining budget.

the new tax, but also by a portion of the existing tax that they had previously ignored, but now do not.²

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

Our experiment studies the behaviour of 2,998 consumers—approximately matching the U.S. adult population on household income, gender, and age—drawn from the forty-five U.S. 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 those in the consumer’s city of residence, or (3) high sales taxes that are triple those in the consumers’ city of residence. Decisions in the experiment are incentive compatible: study participants use a \$20 budget to potentially buy one of the randomly chosen products, and purchased products are shipped to their homes.

We begin our empirical analysis by estimating the average amount by which study participants under-react to taxes. Following CLK, we measure under-reaction by estimating the implicit weight placed on taxes, denoted by θ . This measure constitutes a sufficient statistic for welfare analysis when mistakes are homogeneous. In the standard-tax condition, we estimate an average θ of 0.25: study participants react to the taxes as if they are only 25% of their size. This result is quantitatively similar to that of CLK, who find an average θ of 0.35 in an analysis of grocery store purchases and an average θ of 0.06 in an analysis of demand for alcoholic beverages. Our estimates fall within the confidence intervals of this previous work, and our design affords significantly greater statistical power.

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 increase in weight placed on the tax is significant at least at the 5% level, and provides initial evidence that consumers are more attentive to higher taxes. Complementing this evidence, we also show that consumers are on average more likely to under-react to taxes on particularly cheap products (priced below \$5), than they are to taxes on more expensive products (priced above \$5).

Having established variation of misreaction across tax rates, in the second part of our empirical analysis we focus on variation of misreaction across consumers. This analysis is directly motivated by the efficiency cost formulas that we derive, which show that the efficiency cost of a small tax t on a product sold at price p depends on the variance of under-reaction 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 $\text{Var}[\theta|p, t]$. We bound this statistic through a novel combination of a “self-classifying” survey question and experimental behaviour, in a way that requires no assumptions about truth-telling or metacognition. Our estimates of the bound imply that for taxes that are the size of those observed in the U.S., 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 article relates to three distinct literatures. First, beyond extending and generalizing the existing work on tax salience (*e.g.* CLK, [Finkelstein, 2009](#); [Feldman and Ruffle, 2015](#); [Feldman et al., 2015](#)), the article broadly contributes to a growing theoretical and empirical literature in “behavioral public economics” (see [Chetty \(2015\)](#) for a review, and [Mullainathan et al.](#)

2. Moreover, increases in the tax rate can also affect the variance of under-reaction, which in turn affects efficiency costs.

(2012) and [Farhi and Gabaix \(2015\)](#) for general theoretical frameworks). Some of our own previous work on corrective taxation in energy markets has emphasized the importance of welfare estimates that are robust to heterogeneous bias ([Allcott and Taubinsky, 2015](#); [Allcott et al., 2014](#)).³ This article focuses on an importantly different domain and is the first, to our knowledge, to explicitly formalize the welfare-relevant statistics of mistake variation and to empirically measure those statistics. These results have immediate applications to the literature on tax misunderstanding;⁴ however, our framework for analysing variation in mistakes is broadly portable, and can serve as a template for empirical analysis of other psychological biases, and in other domains of behaviour.

Second, our experimental findings are also relevant to the growing literature on firm and consumer interactions in markets with shrouded attributes ([Gabaix and Laibson, 2006](#); [Veiga and Weyl, 2016](#); [Heidhues et al., 2017](#)). The predictions of these models rely on particular assumptions about the heterogeneity of attention to the shrouded attributes, as well as how the inattention depends on the size of the shrouded attribute. Our estimates can thus help guide the quantitative predictions of these models.⁵

Third, our work contributes to the literature on boundedly rational value computation (see *e.g.* [Gabaix, 2014](#); [Woodford, 2012](#); [Caplin and Dean, 2015a](#); [Chetty, 2012](#)). To the best of our knowledge, our result that consumers under-react less to higher tax rates provides one of the first experimental demonstrations in a naturalistic setting of imperfect processing of a financial attribute responding to economic incentives.⁶

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

2. THEORY

This section analyses the tax policy implications of variation in consumers' inattention to or misunderstanding of tax instruments. Specifically, we generalize [Harberger's \(1964\)](#) canonical formulas for the efficiency costs of taxation, as well as CLK's formulas for the case of homogeneous consumers. The formulas we develop transparently highlight the importance of

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

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

5. [Veiga and Weyl \(2016\)](#), for example, show that a monopolist's shrouded attribute strategy will depend on the covariance between inattention to the shrouded attribute and household income.

6. Results on this general topic are mixed. [Abeler and Jäger \(2015\)](#) find that study participants under-react to complex changes in an experimental income tax, but that this under-reaction does not depend on the magnitude of the change. [Feldman et al. \(2015\)](#) find no statistically significant evidence that experimental subjects attend differently to an 8% and a 22% sales tax, although their confidence intervals admit effect sizes of the magnitude documented in this article. In contrast, [Hoopes et al. \(2015\)](#) find that taxpayers pay more attention to capital gains information when the payoffs to doing so are higher. Interestingly, [Feldman and Ruffle \(2015\)](#) find asymmetric attention to comparable taxes and rebates. In tests of boundedly rational decision-making more broadly, [Caplin and Dean \(2015b\)](#) and (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 \(2011, 2015\)](#) show that consumers pay more attention to energy costs when gasoline prices are higher.

accounting for the variation of mistakes across both consumers and tax sizes. The results can be immediately applied to questions about optimal Ramsey or Pigouvian taxes—which we summarize in Section 2.5 and elaborate on in [Online Appendix B](#)—and also apply more broadly to consideration of any kind of imperfectly understood policy instrument. All proofs are contained in [Online Appendix C](#).

2.1. Set-up

2.1.1. Consumer and producer behaviour. *Consumers:* There is a unit mass of consumers who have unit demand for a good x and spend their remaining income on an untaxed composite good y (the numeraire). A person's utility is given by $u(y) + vx$, where $x \in \{0, 1\}$ denotes whether or not the good is purchased, and v is the person's utility from x . Let Z denote the budget (assumed identical across consumers), p the posted price of the product, and t the tax set by the policymaker.⁷ We assume throughout that $Z > p + t$.

A fully optimizing consumer chooses $x = 1$ if and only if $u(Z - p - t) + v \geq u(Z)$. However, we allow consumers to not process the tax fully. Instead, a consumer chooses $x = 1$ when $u(Z - p - \theta t) + v \geq u(Z)$, where θ —which may covary with v or be endogenous to t —denotes how much the consumer under- (or over-) reacts to the tax.⁸

Because we make minimal assumptions about the distribution of θ , this modelling 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 a constant fraction θ of its size ([Gabaix and Laibson, 2006](#); [DellaVigna, 2009](#)).
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, $\theta = \hat{t}/t$.
4. Rounding heuristics.
5. Forgetting about the tax.
6. Any combination of the above biases.

In practice, multiple mechanisms are likely to be in play. Existing data provides little guidance on which mechanisms are the most important (CLK) or on the shape of the distribution of θ . [Gabaix's](#) (2014) anchoring and adjustment model of attention, for example, predicts that each consumer will have a $\theta \in [0, 1]$, with that value depending on the size of the tax. Other theories of inattention may predict binary attention $\theta \in \{0, 1\}$. Incorrect beliefs and rounding heuristics can generate a variety of different values of θ , with instances in which $\theta > 1$.

We develop our theoretical and empirical framework to be robust to all of these possible mechanisms. Instead of defining θ in relation to a specific mechanism, we define it by the behaviour that these mechanisms generate: a difference in willingness to pay depending on the presence of a tax. 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

7. Note that we are assuming here that the policymaker is using a tax instrument with only one level of salience. See [Goldin \(2015\)](#) for a model in which the policymaker can combine tax instruments of differing salience to raise revenue in the least distortionary way possible.

8. We also assume that $Z > p + \theta t$ for all θ , by virtue of our assumption that $Z > p + t$.

between θ and v other than that their joint distribution $F_t(v, \theta)$ generates smooth, downward-sloping aggregate demand curves,⁹ that $\theta \geq 0$ and is bounded, and that the marginal distribution of v does not depend on t . By allowing the distribution of θ to depend on t we capture the possibility that attention to taxes may depend on the tax rate. With minor abuse of notation, we define $E[\theta|p, t]$ and $Var[\theta|p, t]$ to be the mean and variance of θ of consumers who are indifferent about purchasing the product at (p, t) .

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

To focus our analysis on mistakes arising solely from incorrect reactions to the sales tax, we assume that (1) in the absence of taxes, consumers optimize perfectly and (2) consumers' utility depends only on the final consumption bundle (x, y) .¹⁰ Welfare analysis under these two assumptions and our choice-based definition of θ is an application of [Bernheim and Rangel's \(2009\)](#) approach to welfare analysis: we view choice in the presence of taxes as provisionally suspect, and we use consumer choice in the absence of taxes as the welfare-relevant frame. We relax the first assumption in [Online Appendix B](#), following models such as those in [Lockwood and Taubinsky \(2017\)](#) and [Farhi and Gabaix \(2015\)](#).

Producers: We define production identically to CLK: price-taking firms use $c(S)$ units of the numeraire y to produce S units of x . The marginal cost of production is weakly increasing: $c'(S) > 0$ and $c''(S) \geq 0$. The representative firm's *profit* at pretax price p and level of supply S is $pS - c(S)$. Producers optimize perfectly so that the supply function for good x is implicitly defined by the marginal condition $p = c'(S(p))$. Let $\varepsilon_{S,p} = -\frac{\partial S}{\partial p} \frac{p}{S(p)}$ denote the price elasticity of supply. We define $\varepsilon_{D,t}^{TOT} = -\frac{d}{dt} D(p, t) \cdot \frac{p+t}{D}$ to be the total percentage change in equilibrium demand (taking into account changes in producer prices) caused by a 1% change in the tax.¹¹

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

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

9. The smoothness assumption may be violated in situations where these mechanisms follow threshold rules and the thresholds are homogeneous. For example, if a positive mass of consumers always rounds a tax that is greater than 7.5% to 10%, and rounds a tax smaller than 7.5% to 5%, then there would be a point of non-differentiability in the demand curve at a 7.5% tax. Relatedly, if all consumers either completely ignore or fully attend to the tax, and if the tax threshold at which they start paying attention is the same for all consumers, non-differentiability in the demand curve may similarly be generated. However, as long as the thresholds applied for rounding or for paying attention are smoothly distributed across consumers, as in the [Chetty et al. \(2007\)](#) model, the resulting demand curve will be smooth.

10. Assumption 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 generate a third-order, and thus negligible, efficiency cost ([Chetty et al., 2007](#)).

11. For clarity, we remind the reader that all elasticities with respect to the tax are elasticities given behavioural biases, not the rational elasticities.

from this consumer. Excess burden is given by:

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

where $\pi_0 - \pi_1$ is the change in producer profits, p_0 is the equilibrium market price in the absence of taxes, and $p(t)$ is the equilibrium price at tax t . That is, excess burden is the sum of the change in consumer surplus and producer surplus minus government revenue. With quasilinear utility and fixed producer prices (*i.e.* perfectly elastic supply), this is simply $\int_i (v_i - p_0)(x_i^*(p_0, t) - x_i^*(p_0, 0))$: the loss in surplus that accrues from discouraging transactions in which the value of the product v exceeds its marginal cost of production.

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

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

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

2.2. Direct efficiency costs

For the results presented in the body of the article, we assume that u is linear (*i.e.* no income effects are present), but we discuss the implications of income effects at the end of the section, and in more detail in [Online Appendix A.2](#).

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

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

Proposition 1 provides a general formula for the (direct) excess burden of a small tax t when consumers are arbitrarily heterogeneous. When $\text{Var}[\theta|p, t] = 0$, the formula reduces to the formula provided in CLK, which shows that the excess burden of the tax is proportional to $E[\theta|p, t]$. In the simple framework without income effects, the more the consumers ignore the tax, the less the consumers are discouraged from purchasing the product because of the tax, and thus the smaller the excess burden. The formula, as written, does not feature the covariance between θ and v or between θ and elasticities. However, we note that those covariances determine which consumers are on the margin, and are thus incorporated into our $E[\theta|p, t]$ and $\text{Var}[\theta|p, t]$ terms.

The general formula illustrates that it is not just how much people under-react to the tax on average that matters, but also the variance of marginal consumers' under-reactions. To take a

stark example, suppose that $E[\theta]=0.25$ for consumers on the margin. When all consumers are homogeneous with $\theta=0.25$, equation (2) shows that the excess burden from a marginal increase in the tax is $(0.25)tD(p,t)\frac{\varepsilon_{D,t}^{TOT}}{p+t}$; 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 marginal consumers have $\theta=1$ while 75% have $\theta=0$, so that $E[\theta]=0.25$ and $Var[\theta]=(0.75)(0.25)$. In this case, we still have $E[\theta]=0.25$, but equation (2) implies that the excess burden is now *at least* $tD(p,t)\frac{\varepsilon_{D,t}^{TOT}}{p+t}$, since $\varepsilon_{D,t} \geq \varepsilon_{D,t}^{TOT}$. Interestingly, this is greater than or equal to the inference that would be made by an analyst who assumes that consumers optimize perfectly and thus uses the tax elasticity of demand as a sufficient statistic for calculating excess burden.

The intuition for this result is that heterogeneity in consumers' mistakes creates a market failure that is conceptually distinct from the effect of a homogeneous mistake. If consumers are homogeneous in their under-reaction 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 under-reaction, 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 under-react to taxes the most. There is thus an additional efficiency cost from an inefficient match between consumers and products.¹²

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

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

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

Corollary 1 shows that when supply is inelastic—and thus the equilibrium quantity produced by the market does not change—the excess burden of a small tax t depends only on the variance of bias and the price elasticity of demand. Intuitively, this is because all of the efficiency cost is generated by misallocation, the extent of which is proportional to the variance of θ —which quantifies the extent of individual differences—and the price elasticity of demand—which determines how much the individual differences translate to different purchase decisions. That efficiency costs can be significant even when supply is inelastic is in sharp contrast to standard results in public finance that efficiency costs should be zero if taxes do not distort the equilibrium quantity. More generally, the results imply that when consumers are heterogeneous in their under-reaction, efficiency costs will be significantly higher than in the standard model when supply is relatively inelastic compared to demand.¹³

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

13. Empirical work on how the supply elasticities compare to demand elasticities is scarce and has not settled on a range. Studies that estimate pass-through of salient consumption taxes (those included in the upfront

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

Proposition 2. *The excess burden of imposing a small tax (so terms of order t^3 or higher are negligible) in a previously untaxed market is*

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

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

Assumption A. For each θ in the support of the distribution F , the conditional distribution $F(v|\theta)$ has a differentiable density function.

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

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

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

price of the good) find that the pass-through to the final, after-tax price—given by $\frac{\varepsilon_{S,p}}{\varepsilon_{S,p} - \varepsilon_{D,p}}$ —ranges from 19% to 48% (Benzarti and Carloni, 2016). Studies that estimate pass-through of not-fully-salient sales taxes into the after-tax price—given by $\frac{\varepsilon_{S,p} - (1 - E[\theta|p, t]) \varepsilon_{D,p}}{\varepsilon_{S,p} - \varepsilon_{D,p}}$, find estimates ranging from 70% to 100% (Besley and Rosen, 1999; Doyle and Samphantharak, 2008).

2.3. Indirect efficiency costs: the consequences of debiasing

In this section, we provide “Harberger-type” formulas for the efficiency costs (or benefits) of changing the distribution of θ . We keep the tax fixed, and we consider a family of distributions $F_n(\theta, v)$ that are smooth functions of n for all θ, v . We think of n as the “nudge parameter”, and we ask how the excess burden of a tax changes as we shift this parameter by a small amount from n to $n + \Delta n$. The formulas here serve as an intermediate step to the final formulas that we derive in Section 2.4, but we also view them to be of independent interest as a novel extension of the standard public finance toolbox. We provide results under two additional assumptions:

Assumption B. $F_n(h(\theta, n), v) = F_0(\theta, v)$, where h is differentiable in θ and n , and $\frac{\partial}{\partial n} h$ is bounded.

Assumption C. The terms $t^{k+1} \frac{\partial^k}{\partial p^k} D$ are negligible for all $k \geq 2$.

Assumption B requires that the nudge smoothly changes the distribution of θ . Assumption C is a variation of the standard Harberger formula assumption that the term $t(\Delta t)^k D_{pp}$ is negligible, but is a slightly stronger requirement on how small t or D_{pp} needs to be.

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

The additional structure provided by Assumptions B and C essentially ensures a good fit from a quadratic approximation for the efficiency costs corresponding to each θ type, and that the price elasticities of demand are not too different across the θ types.

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

Proposition 4. Suppose that producer prices are fixed ($\varepsilon_{S,p} = \infty$), and that Assumptions A–C are satisfied. Then

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

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

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

2.4. Total efficiency costs

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

Proposition 5. Consider two taxes t_1 and $t_2 = t_1 + \Delta t$. Suppose that producer prices are fixed ($\varepsilon_{S,p} = \infty$) and that Assumptions A–C are satisfied for the family of distributions F_t indexed by the tax t . Suppose also that for $k \geq 2$, the terms $t(\Delta t)^k D_{pp}$ and $(\Delta t)^k$ are negligible. Then

$$EB(t_2, F_{t_2}) - EB(t_1, F_{t_1}) \approx - \left(t_1(\Delta t) + \frac{(\Delta t)^2}{2} \right) \left(E[\theta|p, t_2]^2 + \text{Var}[\theta|p, t_2] \right) D_p \quad (3)$$

$$- \left(\frac{t_1^2}{2} \right) \left(E[\theta^2|p, t_2] - E[\theta^2|p, t_1] \right) D_p \quad (4)$$

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

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

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

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

To illustrate the formula in the corollary, suppose that θ is homogeneous, so that $\text{Var}[\theta|p, t] = 0$. In this case, efficiency costs are not simply given by $\left(t_1(\Delta t) + (\Delta t)^2/2 \right) \frac{D}{p+t_1} E[\theta|p, t_1] \varepsilon_{D,t}$, as would be prescribed by Proposition 3. There are additional efficiency costs, arising from the nudge effect, given by $\frac{t_1(\Delta t)}{4} \frac{D}{p+t_1} (E[\theta|p, t_2]^2 - E[\theta|p, t_1]^2) \varepsilon_{D,p}$. In the simple case of $\text{Var}[\theta|p, t] = 0$, these additional efficiency costs correspond to the fact that the value of the product to the marginal consumer under t_2 is not simply $p + E[\theta|p, t_1](t_1 + \Delta t)$, as it would be if taxes did not change under-reaction, but is instead $p + E[\theta|p, t_2](t_1 + \Delta t)$. That is, in contrast to the standard model, the value of the product to the marginal consumer is a convex, rather than a linear function of the tax when $E[\theta|p, t]$ is increasing in t .

2.5. Extensions and optimal tax implications

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

In formalizing the implications of our excess burden calculations for optimal taxes, the results in the Appendix generate several new insights. First, when there are no other market frictions and taxes are used only to meet a fixed revenue requirement, the optimal tax system may deviate from the canonical Ramsey inverse elasticity rule in several ways. If people under-react 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.¹⁴

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 , under-reaction 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 therefore cannot achieve the socially optimal consumption of x . Our general message about the importance of taking into account the misallocation arising from heterogeneity in θ is thus particularly relevant in the presence of other market frictions.

Income effects: We have thus far assumed that $u(y)$ is linear, imposing an absence of income effects. This is a reasonable assumption for small-ticket items for which p and t are small relative to income. Relaxing this assumption complicates our analyses, but follows the same principles as the baseline excess burden formula without income effects. As we show in [Online Appendix A.2](#), the formulas we derive in the body of the article still hold in the presence of income effects when either (1) the taxed product is a small share of consumers' expenditures or (2) the taxed product is purchased on a reasonably frequent basis, and the consumer can observe his budget in between the purchases. Thus, for common household commodities, we believe that our results hold robustly in the presence of income effects.

However, for infrequent, large-ticket purchases there can still be efficiency costs when consumers ignore the tax fully. This can occur when a consumer spends more money than he realizes on the product in question, and then consumes inefficiently too little y in the future after he is surprised by a smaller budget. For large-ticket purchases, this process of budget adjustment can become quantitatively important, and we note that this process is not incorporated into the analyses presented here. For related discussion, see [Reck \(2014\)](#).

14. We perform these calculations under the assumption that there are no cross-price effects. While this assumption is common in excess burden analyses, it can be reasonably viewed as limiting. However, the broad concepts developed in this article apply even when this assumption is relaxed. When people homogeneously under-react to a tax on one product, shifting that tax will dampen the, *e.g.* substitution to other products. Heterogeneity similarly creates additional misallocation through the cross-price effect, as the people substituting will sometimes be the "wrong" ones.

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

2.6. Identification from aggregate demand data

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

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

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

3. EXPERIMENTAL DESIGN

Platform: The experiment was implemented through *ClearVoice Research*, a market research firm that maintains a large and demographically diverse panel of participants over the age of 18. This platform is frequently used by firms that ship products to consumers to elicit product ratings, but is additionally available to researchers for academic use (for other examples of research using this platform, see, e.g. [Benjamin et al., 2014](#); [Rees-Jones and Taubinsky, 2016](#)). Two key features of this platform make it appropriate for our experimental design. First, ClearVoice provides samples that match the U.S. population on basic demographic characteristics. Second, ClearVoice maintains an infrastructure for easily shipping products to consumers, which facilitates an incentive-compatible online-shopping experiment.

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

Each consumer was randomly assigned to one of three arms: (1) the “no-tax arm”, (2) the “standard-tax arm”, and (3) the “triple-tax arm”. The standard- and triple-tax arms were

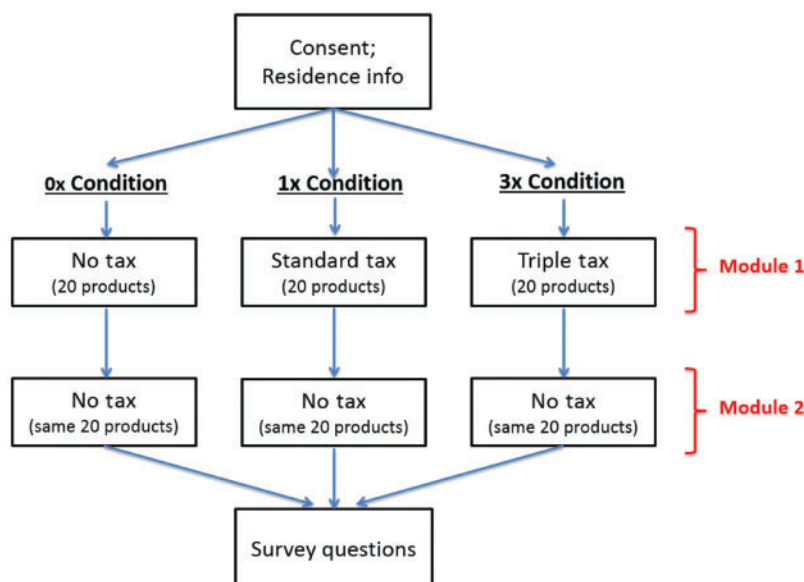


FIGURE 1

Experimental design

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

implemented to provide within-subject comparisons of purchasing decisions with and without taxes. The no-tax arm was implemented to identify any order effects on valuations over the course of the experiment and to help test for demand or anchoring effects.¹⁵

Each module consisted of a series of shopping decisions involving twenty 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 twenty products were used in each module and in each arm of the experiment. The order in which the twenty products were presented was randomly determined, and independent between the two modules.

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 forty-five states are all represented in our final sample. Prior to learning the details of the experiment, consumers were asked to report their state, county, and city of residence. To correctly determine the money spent in the experiment, this information was matched to a data set of tax rates in all cities in the U.S.¹⁶

This design is closely related to several recent experimental studies of tax salience (*e.g.* Feldman and Ruffle, 2015; Feldman *et al.*, 2015), but differs in important ways. Our design

15. An additional goal of the no-tax arm was to identify the distribution of random shocks to valuations between module 1 and module 2, and to combine this with data from the other two arms to deconvolve the distribution of individual θ parameters from the distribution of measurement error. Ultimately, the variance of the measurement error we encountered was too high to permit a well-powered deconvolution of this type.

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

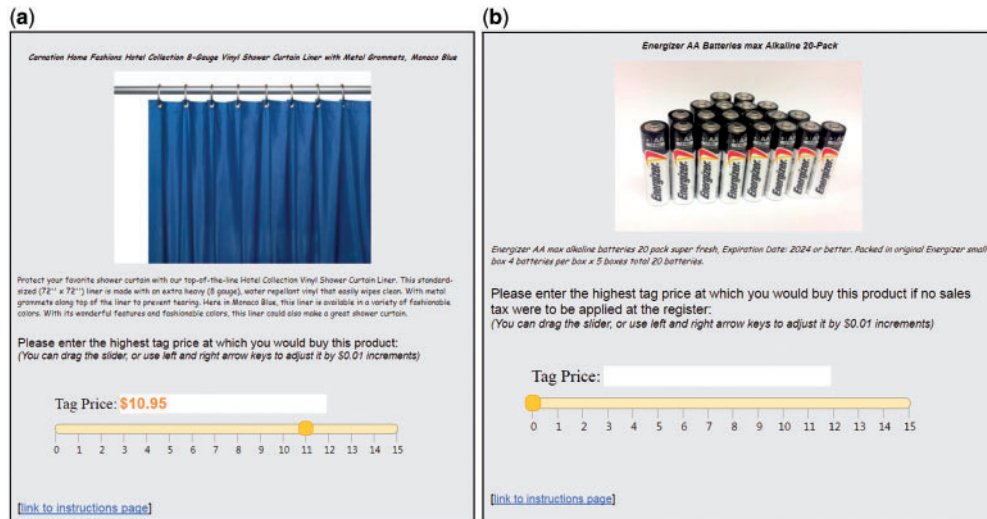


FIGURE 2

Decision format

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.

combines within-subject manipulation of tax rates with a pricing mechanism that elicits full and precise demand curves. This design, combined with our unusually large sample size, allows us to infer the sufficient statistics of our general welfare formulas—an exercise not possible with previous experimental designs.

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

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

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

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

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

To make this experimental shopping experience as close as possible to the normal shopping experience and to enable tests for incorrect beliefs, consumers were not told what tax rate applies in their city of residence. Once consumers read the instructions (and answered the comprehension questions), they were never reminded of the taxes again in the tax modules. In contrast, the no-tax modules emphasized the absence of taxes to ensure that choices in those models reflect consumers’ true willingness to pay for the products.

Product selection: To arrive at the final list of twenty household products, we began with a list of seventy-five potential items in the \$0–\$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 twenty items that had unimodal distributions of valuations and had the least censoring at \$0 and \$15. [Online Appendix F](#) lists the products, prices, and Amazon.com product descriptions that were displayed to study participants.

Incentive compatibility: Decisions in the experiment were incentive compatible. All study participants who passed the necessary comprehension questions (described below) had a 1/3 chance of being selected to receive a \$20 budget; accounting for the probability of failing the comprehension check, this chance was approximately 1/4. Participants were informed of this incentive structure prior to making any decisions, but they did not know if they received the budget until they completed the experiment. If they did not receive the budget, they simply received a compensation of \$1.50 and no products from the study. Consumers who were selected to receive the \$20 budget had one out of the forty decisions (from modules 1 and 2 combined) selected to be played out. Outcomes were determined using the Becker–DeGroot–Marshak (BDM) mechanism. A random tag price, between \$0 and \$15, was drawn. If the randomly generated price was below the maximum tag price the consumer was willing to pay, then the product was sold to the consumer at that tag price p , and a final amount of $p(1 + \tau)$ (where τ is the experimentally induced tax rate) was subtracted from this consumer’s budget. The product was shipped to the consumer by ClearVoice, and the remainder of the budget was included in experimental compensation. 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 want to buy the product.

Comprehension questions: It is important to ensure that study participants understand the experimental tax rate that applies to their decisions, so that the appearance of under-reaction is not generated by a simple failure to read experimental instructions. In both module 1 and module 2, we thus gave study participants a multiple-choice comprehension question designed to confirm their understanding of the applicable experimental tax rate. This question presents an item being purchased for a \$5 tag price, and asks the respondent to choose the amount of money that

would be deducted from their budget from several tag-price/tax combinations. In both modules, the quiz question appeared on the same screen as the instructions for that module. Subjects who fail these questions are generally excluded from our analyses; however, we demonstrate that our main analyses are robust to alternative treatments of these subjects in Section 4.7.

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

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

4. QUANTIFYING UNDER-REACTION ACROSS DIFFERENT TAX SIZES

4.1. *Sample selection, demographics, and balance*

In this section, we discuss the creation of our final sample for analysis. We then analyse the demographic properties and balance of that sample.

A total of 4,328 consumers completed the experiment. For our primary analyses, we restrict our sample to the 3,066 respondents who correctly answered the instruction-comprehension questions regarding the tax rate that applied in both module 1 and module 2. Unsurprisingly, the 29% of consumers who failed these comprehension questions do not react to the differences in taxes across conditions. Thus, while these respondents would contribute to evidence of under-reaction to sales taxes, we believe the misoptimization these consumers exhibit is likely due to misunderstanding of our experimental manipulation. This type of misunderstanding is conceptually distinct from misunderstanding a given tax rate and is not the object of interest in our theoretical analysis.¹⁷

Out of the remaining 3,066 consumers, thirty consumers were not willing to buy at any 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 thirty consumers. We thus exclude them from analysis as well. We additionally exclude ten consumers who reported living in a state with no sales tax.¹⁸

In part due to our pre-test 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 twenty-eight consumers did not provide an answer to this question upon encountering it. We exclude these consumers as well, leaving us with a final sample size of 2,998.

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. Experimental recruitment was targeted to generate a final sample approximating the gender,

17. We also included questions to check if participants understood the BDM mechanism. In total, 78% of participants passed those comprehension questions, and we show in [Online Appendix E.7](#) that our results are robust to restricting to this sample. We are far less concerned about potential misunderstanding of the pricing mechanism for two reasons. First, participants were clearly instructed that it was in their best interest to always truthfully report the maximum tag price at which they would be willing to buy the product. Second, most forms of "strategic" price reporting do not confound estimates of θ . While subjects might report a threshold for purchase that is not their true willingness to pay, this threshold should be a function of final price. Differences in the reported threshold across conditions may still be interpreted as evidence of differential weighting of posted price and sales taxes.

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

TABLE 1
Demographics by experimental arm

	All	No tax	Std. tax	Triple tax	<i>F</i> -test <i>p</i> -value
Age	50.49 (14.63)	50.80 (14.28)	50.43 (14.84)	50.20 (14.82)	0.66
Household income (\$1,000s)	63.04 (56.29)	61.86 (55.00)	63.67 (58.21)	63.78 (55.77)	0.68
Household size	2.40 (1.52)	2.40 (1.60)	2.38 (1.50)	2.42 (1.46)	0.86
Married	0.34	0.33	0.36	0.34	0.35
Male	0.48	0.47	0.51	0.47	0.15
Education					
Highschool degree or higher	0.95	0.94	0.95	0.95	0.74
College degree or higher	0.41	0.41	0.40	0.40	0.84
Post-graduate education	0.17	0.18	0.16	0.16	0.49
Ethnicity					
Asian	0.03	0.03	0.04	0.03	0.88
Caucasian	0.77	0.76	0.77	0.78	0.57
Hispanic	0.03	0.04	0.03	0.03	0.47
African American	0.07	0.08	0.07	0.08	0.83
Other	0.02	0.03	0.02	0.02	0.39
Tax rate in city of residence	7.32 (1.15)	7.36 (1.16)	7.31 (1.13)	7.30 (1.15)	0.36
<i>N</i> (Final sample)	2,998	1,102	982	914	
Comprehension test pass rate (%)	71	78	70	65	< 0.01

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

income, and age distribution of the U.S. As a result, our sample—which is 48% male, has a median income of \$50,000, and average age of 50—is similar to the U.S. population on these basic demographics. Despite this favourable comparison, we caution the reader that the nature of recruitment into the ClearVoice panel likely induces selection on unobservable characteristics.

We find no evidence of selection on demographic covariates across experimental arms. We fail to reject the null hypotheses of equality of the demographics in Table 1 when comparing Arm 1 versus Arm 2 (*F*-test, $p=0.49$), Arm 2 versus Arm 3 (*F*-test, $p=0.94$), or Arm 1 versus Arm 3 (*F*-test, $p=0.36$).

In contrast to the demographic results, there are statistically significant cross-arm differences in the likelihood that consumers pass the comprehension questions regarding the tax rate that apply in the experiment. The likelihoods of correctly answering both comprehension questions are 78%, 70%, and 65% in the no-tax, standard-tax, and triple-tax arms, respectively.¹⁹ The null hypothesis of equal pass rates is rejected for any pair of arms at the 5% significance level. The

19. To provide further detail, the fraction of people correctly identifying the applicable tax rate in module 1 was 81%, 82%, and 66%, in the no-tax, standard-tax, and triple-tax arms, respectively. In module 2, the corresponding rates were 87%, 80%, and 84%. Conditional on correctly answering the module 1 question, the likelihood of correctly answering the module 2 question was 96%, 86%, and 97%, respectively. Note, in particular, that while the module 1 question was of approximately equal difficulty in the no-tax and standard-tax arms, the likelihood of answering both module 1 and module 2 questions correctly was significantly higher in the no-tax arm. We believe this is because the tax rate, and thus the correct answer to the comprehension question, changed in one arm, but not the other.

differential selection introduced by these differing pass rates introduces a potentially important confound to cross-arm inference. However, we will show in Section 4.7 that our primary results are robust to both worst-case assumptions about differential selection and to the reinclusion of those failing the test.

4.2. Summary of behaviour

We begin with a graphical summary of the data. Figure 3 provides a summary of the demand curves as functions of before- or after-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 experimental arm, m denotes the module, and p the before-tax price. Because there are twenty 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 .

Panel (a) of Figure 3 shows that consumers do react to sales taxes in module 1, as their willingness to buy at a given before-tax price is decreasing in the size of the sales tax. However, as shown in panel (b), consumers do not react to taxes as much as perfect optimization would require. In this panel, we construct the demand curves that would be expected if consumers reacted to the taxes fully, and find substantially larger differences than those observed in panel (a). As demonstrated in panel (c), this discrepancy results in differences in demand curves across treatment arms when they are plotted as a function of after-tax price: consumers are willing to buy at higher final prices in the presence of larger taxes.

While consumers in the different treatment arms behave differently in module 1, panel (d) shows that all treatment arms exhibit similar demand patterns in module 2. This pattern is confirmed by several statistical tests. For our first test, we compute an average pre-tax price $\bar{p}^i = \frac{1}{20} \sum_k p^{ik}$ for each consumer i , and then compare the distributions of \bar{p}^i . Kolmogorov–Smirnov tests find no differences in the \bar{p}_i between the no-tax and standard-tax arms ($p=0.73$), between the no-tax and triple-tax arms ($p=0.29$), and between the standard-tax and triple-tax arms ($p=0.50$).²⁰ OLS and quantile regressions comparing the average willingness to pay in module 2 across experimental arms similarly detect no differences (see [Online Appendix E.1](#)). Since all three treatment arms face the same no-tax environment in module 2, this similarity of demand behaviour is reassuring: it suggests that the willingness to pay elicited in module 2 is not contaminated by earlier cross-arm differences, as could arise in the presence of anchoring or demand effects.²¹

4.3. Econometric framework

We now present our baseline econometric framework for studying how under-reaction to taxes varies by experimental condition and by observable characteristics. Let p_1^{ik} be the highest tag price a subject i is willing to pay in module 1 for product k , and define p_2^{ik} analogously for module 2.

20. By contrast, the corresponding p -values for module 1 are $p=0.12$, $p<0.001$, $p<0.001$, respectively. Note that these tests are less powerful than our measures of reaction to taxes in Section 4.4, which make use of within-subject identification provided by both modules.

21. By “anchoring” we mean that consumers might under-report willingness to pay in the standard and triple-tax arms due to the psychological influence of previously reporting a lower module 1 price. By “demand effects” we mean that consumers might react more strongly to the absence of taxes in module 2 of the experiment because they perceive this to be an experiment about how they are supposed to choose “differently” in the different modules. Either of these confounds would lead the module 2 demand curves to differ. This would bias our estimates of $E[\theta]$, since they rely on within-subject comparisons of module 1 to module 2 prices.

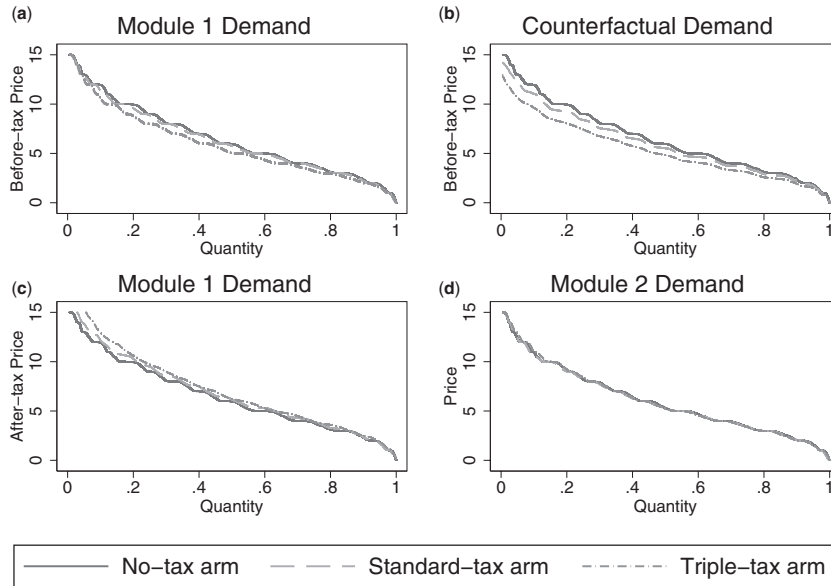


FIGURE 3

Average demand curves in the first and second stages of the experiment

Notes: This figure plots demand curves from the first and second modules of the experiment, averaging across all twenty products. In the first stage, consumers face either no taxes, standard taxes, or triple their standard taxes. In the second stage, consumers in all three arms face no additional taxes. To construct the figures, we start with the demand curves, denoted $D_k^{C,m}(p)$, for each product k , $C \in \{0x, 1x, 3x\}$ denotes the no-tax, standard-tax, or triple-tax experimental arm, m denotes the module (stage), and p the relevant price. The average demand curves are calculated as $D_{avg}^{C,m}(p) := \frac{1}{20} \sum_k D_k^{C,m}(p)$. Panel (a) plots average demand as a function of the after-tax prices in module 1. For comparison, panel (b) plots the counterfactual average demand in module 1 that would be expected if consumers react to taxes fully. We reconstruct the demand curves by assuming that if a fraction $D(p)$ of consumers are willing to buy at price p in the no-tax arm, then a fraction $D\left(\frac{p}{1+\tau}\right)$ of consumers are willing to buy at a (before-tax) price p when facing tax rate τ . Panel (c) plots demand as a function of the after-tax prices in module 1. Panel (d) plots demand as a function of the tax-free prices in 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 under-reaction 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, $\frac{y_{ik}}{\tau_i}$ provides a noisy estimate of θ_{ik} because study participants' reported values for the product fluctuate. Furthermore, this measure may be biased if average perceived values of the products vary between module 1 and module 2 even in the absence of tax changes. This phenomenon—which we refer to as *order effects*—is commonly found in pricing experiments (see, e.g. Andersen *et al.*, 2006; Clark and Friesen, 2008), and the no-tax arm of the experiment was designed to allow us to identify and econometrically accommodate these effects. In this arm, we find that participants' valuations declined by an average of 42 cents from module 1 to module 2 ($p < 0.001$). Our econometric approach incorporates these order effects and allows them to depend on any estimated covariates, but we assume that order effects do not vary by experimental arm. This assumption, labelled A1 below for reference, allows us to extrapolate the estimated order effects from the no-tax arm to the other tax arms, in which the identification of order effects would otherwise be confounded with the variation in tax rates between module 1 and 2.

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

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

$$E[y_{ik}|X] = \log(1 + \theta_{ik}\tau_i) + \beta X_{ik}$$

$$E[\theta_{ik}|X_{ik}] \approx E\left[\frac{\log(1 + \theta_{ik}\tau_i)}{\tau_i}|X_{ik}\right] = \alpha X_{ik}$$

The model above implies the following moment conditions:

$$E[X'_{ik}y_{ik}] = X'_{ik}\beta X_{ik} \quad \text{for no-tax arm} \quad (5)$$

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

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 Equations (5) and (6) for either the standard or triple-tax arm separately, the system of equations is exactly identified. When pooling data from multiple treatment arms, we will assume that Equation (6) holds independently for each arm, but with a common α . The system is thus over-identified, and we use the two-step generalized method of moments (GMM) estimator to obtain an approximation to the efficient weighting matrix. We will often condition on $p_2^{ik} \geq \underline{p}$ (typically $p_2^{ik} \geq 1$)—*i.e.* focusing analysis on those with non-negligible willingness to pay—as a means of increasing precision. Because most of our analysis takes p_2/p_1 as an object of interest, noisiness in responses can generate dramatic variation in this quantity when valuations approach zero. All of our point estimates are robust to the inclusion of all data.

Although our approach may seem somewhat complicated, we show in [Online Appendix E.6](#) that all of our main results are robust to a simpler approach, using OLS to regress y_{ik} on the tax rate. As we elaborate in that Appendix, however, we prefer our GMM approach as it avoids the need to assume that mean under-reaction is constant across tax sizes within an experimental arm—an assumption that our results refute.²²

4.4. Average under-reaction to taxes by experimental arm

Table 2 presents our estimates of average θ in each arm using the econometric framework presented in Section 4.3. We provide estimates using all data, as well as conditioning on $p_2^{ik} \geq 1$ and $p_2^{ik} \geq 5$. Across all specifications, we estimate an average θ of approximately 0.25 in the standard-tax arm and an average θ of approximately 0.5 in the triple-tax arm.²³ Due to advantages of our design,

22. Note, also, that in principle, we could have used $\frac{p_2^{ik} - p_1^{ik}}{\tau_i p_1^{ik}}$ instead of y_{ik} as the dependent variable. We prefer our approach 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 $\frac{p_2^{ik} - p_1^{ik}}{\tau_i p_1^{ik}}$ for each arm.

23. Note that the relevant statistic in our welfare formula is the average θ of marginal consumers, $E[\theta|p, t]$. In contrast, the estimates presented here are the iterated expectation $E[E[\theta|p, t]] = E[\theta]$, averaging this value across different

TABLE 2
Average θ (weight placed on tax) by experimental arm

	1 All	2 $p_2 \geq 1$	3 $p_2 \geq 5$
Std. tax avg. θ	0.261** (0.111)	0.250*** (0.095)	0.226** (0.094)
Triple tax avg. θ	0.481*** (0.045)	0.475*** (0.039)	0.535*** (0.041)
Observations	59,960	58,478	32,810
Difference p -value	0.03	0.01	<0.001

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

these estimates are notably more precise than those of prior work and strongly reject the null hypotheses that consumers completely neglect or completely attend to taxes.²⁴ 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 $p_2^{ik} \geq 5$.²⁵

4.5. Further tests of endogenous attention

Our baseline results suggest that consumers attend more to higher taxes. However, several important caveats apply. Consumers might overreact to the triple tax if they are surprised by the unusual scenario (Bordalo *et al.*, 2017). This suggests that a measurement of average θ shortly after a real or experimentally induced tax change might overestimate the degree of attention that would be realized after the resolution of surprise. On the other hand, our estimates of average θ

possible margins. We show in Section 4.5 that, because market prices are slightly higher than the median price at which consumers are on the margin, and because consumers pay more attention to larger taxes, the average θ of consumers on the margin at existing market prices is similar, but slightly higher, than the unconditional averages $E[\theta]$ reported here.

24. CLK’s estimates of θ are calculated by drawing estimates from several different regressions, and standard errors are not reported. To approximate the relevant standard errors for comparison to our own, we apply the delta method using the reported standard errors of each input estimate and assuming no covariance between them. This results in an estimated standard error for θ of 0.18 in the grocery store experiment and of 0.67 in the observational study of demand for alcoholic beverages, compared to point estimates of 0.35 and 0.06, respectively.

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

in the triple-tax arm may underestimate long-run attention because it may take time for people to update their heuristics in a modified decision environment.

A complementary analysis that could test for long-run response would be to estimate whether consumers are more attentive in states with larger sales taxes. However, since the variation in tax rates across states is substantially lower than the tripling of taxes considered in our experiment, such an analysis would require a sample size that is approximately 45 times larger than ours to be well-powered. Unfortunately, such a test cannot currently be feasibly implemented with a lab-in-field approach like our own.²⁶

An alternative and better-powered approach to testing endogenous response to stake size makes use of variation in willingness to pay rather than tax rates. Since the total tax is determined by $t = \tau p$, variation in either tax rates or maximal acceptable purchase prices may be used to generate variation in stakes.

We operationalize this test by dividing all consumers (from all three arms) into three bins corresponding to their module 2 valuation ($p_2^{ik} < 5$, $p_2^{ik} \in [5, 10)$, and $p_2^{ik} \geq 10$), and then estimating an average θ for each bin. Note that we partition consumers using module 2 prices to avoid endogeneity issues arising from the fact that the module 1 prices will depend on a person's attention to the tax.²⁷ 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 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 versus first bin and 0.15 for third versus first bin, in both the triple-tax arm or pooled analysis. We do not detect a difference for average θ between the second and third bin, although we also cannot reject a moderate one. This suggests that attention may not increase linearly with price and that consumers employ different attention strategies for very low-price products below \$5 versus moderate-price products above \$5.

This analysis is consistent with attention increasing in the absolute tax $p\tau$. However, this result could also be obtained if consumers willing to pay the most for the products are also the most attentive. Columns (4)–(6) report an analogous test, ruling out this possibility through the inclusion of individual fixed effects (Online Appendix D.2 formally documents how we modify our GMM strategy). While estimates of attention are somewhat lower than in the first three columns, we again find greater inattention when $p_2^{ik} < 5$ than when $p_2^{ik} \geq 5$.

In summary, our findings are consistent with attention allocation that is endogenous to stake size, whether variation in stake size is generated through experimentally manipulated tax rates or through naturally occurring variation in the prices at which consumers are marginal.

This finding becomes important when comparing average attention found in our experiment to the attention predicted to occur at existing market prices. Subjects in our experiment valued the

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

27. As an alternative approach to accounting for the endogeneity of module 1 prices and θ , Amazon.com prices may be used as an instrument for module 1 willingness to pay. Such an approach is ill-powered compared to our preferred specification—point estimates indicate similar patterns of endogenous inattention, but we cannot reject the null hypothesis of exogeneity. Results of this approach are reported in Online Appendix E.4.

TABLE 3
Average θ (weight placed on tax) for different product valuations

	1 Standard	2 Triple	3 Pooled	4 Standard	5 Triple	6 Pooled
Middle p_2 bin	−0.097 (0.147)	0.117** (0.054)	0.123** (0.054)	−0.077 (0.101)	0.097** (0.038)	0.104*** (0.038)
High p_2 bin	0.115 (0.185)	0.147** (0.074)	0.154** (0.074)	0.168 (0.152)	0.069 (0.053)	0.072 (0.053)
Std. tax cons.	0.266* (0.140)		0.156 (0.098)			
Triple tax cons.		0.402*** (0.054)	0.395*** (0.054)			
Individual fixed effects	No	No	No	Yes	Yes	Yes
Observations	40,651	39,378	58,478	40,651	39,378	58,478

Notes: This table displays GMM estimates of the relationship between average θ and the valuation of the good considered, applying the methodology discussed in Section 4.3. θ is defined as the “weight” that consumers place on the sales tax, with $\theta = 0$ corresponding to complete neglect of the tax and $\theta = 1$ corresponding to full optimization. Columns (1)–(3) estimate the model $\bar{\theta}_{ik} = \alpha_0^{1x} \mathbf{1}_{1x} + \alpha_0^{3x} \mathbf{1}_{3x} + \alpha_{p_2 \in [5,10]} \mathbf{1}_{p_2 \in [5,10]} + \alpha_{p_2 \geq 10} \mathbf{1}_{p_2 \geq 10}$. We assume that $\alpha_{p_2 \in [5,10]}$ and $\alpha_{p_2 \geq 10}$ do not change across the standard- and triple-tax arms, but we allow for different baseline values α_0^{1x} and α_0^{3x} . Columns (4)–(6) control for individual fixed effects, estimating the model $\bar{\theta}_{ik} = \theta_i + \alpha_{p_2 \in [5,10]} \mathbf{1}_{p_2 \in [5,10]} + \alpha_{p_2 \geq 10} \mathbf{1}_{p_2 \geq 10}$. We model the two moment conditions for each arm separately, and we use the two-step GMM estimator to approximate the efficient weighting matrix for the over-identified model. All specifications condition on module 2 price (p_2) being greater than 1. Standard errors, clustered at the subject level, are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

considered products somewhat lower, on average, than the prices posted on Amazon.com (average Amazon.com price: \$10.15; average module 2 willingness to pay: \$6.09). As we document in [Online Appendix E.3](#), consumers who are marginal at market prices have average values of θ approximately 0.1 higher than other consumers. When extrapolating the quantitative estimates of this article into new settings, differences in marginal valuations between our experiment and the setting of interest must be similarly accommodated.

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 under-reaction, 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.”

On average, consumers’ beliefs are very accurate. In total, 52% of consumers know their tax rate exactly, 74% are within 0.5 percentage points, and 85% are within 1 percentage point. The average of beliefs is 7.05%, while the average actual tax rate of consumers in the study is 7.32%, indicating almost no mean bias.²⁸

To provide a graphical summary of how perceived beliefs vary with the actual tax rate, we construct Figure 4 which plots average perceived taxes for each of twenty quantiles of actual taxes. The best-fit regression line in the figure has an intercept of 0.28 percentage points

28. Although the participants were asked 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. We additionally exclude four obvious outlier values that are above 100.

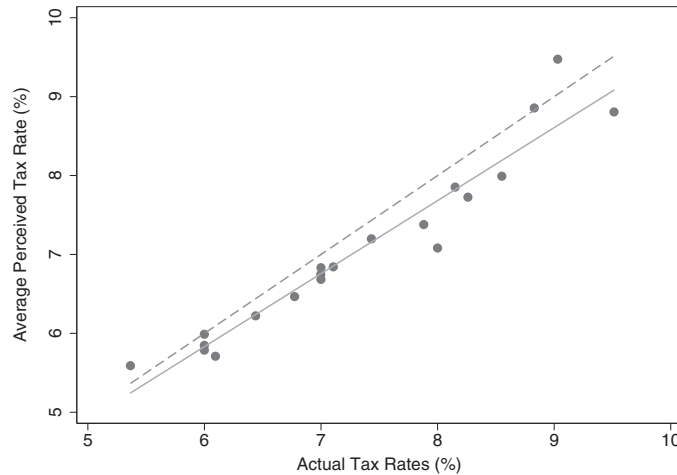


FIGURE 4

Perceived versus actual sales tax rates

Notes: This figure plots the relationship between the actual tax rates subjects face and the tax rates that they believe apply. To construct the figure, we first divide the actual tax rates into twenty quantiles. We then plot the average belief versus the average actual tax rate for each of the quantiles. The dashed 45-degree line represents the counterfactual of correct beliefs.

(s.e. = 0.44), which is not statistically different from 0 ($p=0.53$), and a slope of 0.93 (s.e. = 0.06), which is not statistically different from 1 ($p=0.22$). We conclude that incorrect beliefs are a negligible source of the consumer mistakes that we document, consistent with CLK's survey results from consumers in a California store.

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

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

4.7. Robustness to selection on comprehension questions

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

TABLE 4
Average θ (weight placed on tax) by experimental arm, re-including subjects who failed comprehension checks

	1 All	2 $p_2 \geq 1$	3 $p_2 \geq 5$
Std. tax avg. θ	0.064 (0.108)	0.107 (0.086)	0.146* (0.085)
Triple tax avg. θ	0.276*** (0.040)	0.292*** (0.032)	0.376*** (0.033)
Observations	84,460	82,009	44,918
Difference p -value	0.03	0.02	<0.01

Notes: This table replicates Table 2, but does not exclude study participants who failed comprehension checks. Standard errors, clustered at the subject level, reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

the experimental tax rate that would apply in comprehension questions before both module 1 and module 2. While we prefer specifications with these subjects excluded as a matter of principle, we note that the main results of Tables 2 and 3 qualitatively replicate when re-including these subjects. Estimates of average θ are systematically lower in these analyses since individuals who do not know the experimental tax rate do not respond to it. However, as demonstrated in Tables 4 and 5, analyses including these subjects similarly demonstrate substantial inattention to taxes, with greater attention among those facing triple taxes and in cases where valuations are comparatively high.²⁹ In summary, while we were concerned *ex ante* about the possibility of selection induced by our screening criteria, *ex post* it appears that our primary results are robust to this concern.

5. QUANTIFYING THE VARIATION OF UNDER-REACTION ACROSS CONSUMERS

Having established that under-reaction varies across tax rates, we now turn to the measurement of variation in θ across individuals.

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

Informally, the idea behind our approach is to partition study participants into subgroups with different average θ 's based on self classifications. We then compute the variance of the subgroup means, which provides a lower bound for the total variance. We divide subjects into subgroups using our “self-classifying” survey question, which we *ex ante* selected as most promising to be strongly associated with under-reaction, and which indeed turned out to be our most predictive measure *ex post*. In this section, we begin by presenting the details of our self-classifying survey

29. As an alternative approach, in Online Appendix D.1, we derive a tight lower bound for the difference between average θ in the triple- and standard-tax arms under relatively mild assumptions about the selection process. When implementing the lower bound, we find that we can reject no difference between the triple- and standard-tax conditions at the 10% significance level ($p = 0.08$). We reject this difference at the 5% significance level ($p = 0.04$) when conditioning on module 2 price $p_2^{ik} \geq 1$, and at the 1% significance level ($p < 0.01$) when conditioning on $p_2^{ik} \geq 5$.

TABLE 5
Average θ (weight placed on tax) for different product valuations, re-including subjects who failed comprehension checks

	1 Standard	2 Triple	3 Pooled	4 Standard	5 Triple	6 Pooled
Middle p_2 bin	−0.003 (0.133)	0.138*** (0.047)	0.145*** (0.047)	0.023 (0.096)	0.116*** (0.035)	0.124*** (0.034)
High p_2 bin	0.159 (0.169)	0.194*** (0.063)	0.201*** (0.063)	0.330** (0.138)	0.148*** (0.049)	0.150*** (0.049)
Std. tax cons.	0.117 (0.126)		0.037 (0.087)			
Triple tax cons.		0.221*** (0.046)	0.214*** (0.045)			
Individual fixed effects	No	No	No	Yes	Yes	Yes
Observations	54,503	54,988	82,009	54,503	54,988	82,009

Notes: This table replicates Table 3, but does not exclude study participants who failed comprehension checks. Standard errors, clustered at the subject level, reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

question. We then present our methodology in Section 5.2 and implement an estimate of the lower bound in Section 5.3.

5.1. The self-classifying survey question

The self-classifying survey question asked consumers in the standard- and triple-tax arms the following: “Think back to Section 1, where you made your first twenty decisions about tag prices. In that section, there was a sales tax that you would have to pay if you bought an item from that section. If there was no sales tax in Section 1, would you choose higher tag prices for the products?” The possible answers to the question were “Yes”, which we code as $R = H$; “Maybe a little”, which we code as $R = M$; and “No”, which we code as $R = L$. Table 6 summarizes participants’ responses to the survey question. In the standard-tax arm, 6% of participants answered “Yes”, 56% answered “Maybe a little”, and 38% answered “No”. Participants in the triple-tax arm were more likely to say “Yes” or “Maybe” than participants in the standard-tax arm (Ranksum test, $p < 0.01$).³⁰

Responses to this question are strongly associated with experimental behaviour. To estimate an average θ for each survey response, we employ a similar methodology as in Section 4.3. Because this survey question was not asked in the no-tax arm, we make the additional Assumption A2 that if survey responses R are predictive of behaviour, it is solely because they are correlated with θ :

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

A2 implies that for the standard- and triple-tax arms,

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

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

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

TABLE 6
Distribution of self-classifying survey responses

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

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

TABLE 7
Average θ (weight placed on tax) conditional on self-classifying survey response

	(1) Standard	(2) Triple
“Yes” average θ	1.103*** (0.277)	0.936*** (0.103)
“A little” average θ	0.436*** (0.110)	0.622*** (0.048)
“No” average θ	-0.172 (0.139)	0.047 (0.056)
Observations	40,651	39,378

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

Table 7 shows that this survey question has a striking degree of predictive power. The table shows that the average θ is not statistically different from 0 for consumers who answer “No”, is in the neighbourhood of 0.5 for consumers who answer “Maybe a little”, and is in a neighborhood of 1 for consumers who answer “Yes”. Table 7 thus shows that under Assumption A2, there are stark differences in θ between different consumers. Moreover, the predictive power of the survey question suggests that, consistent with models of bounded rationality and deliberate attention, people are aware of the mistakes they make in responding to sales taxes.

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

5.2. A lower-bound for the variance of mistakes: theory

Let R be the random variable of study participants' responses to the survey question, which can take on the values $R=H$, $R=M$ or $R=L$.³¹ We now create new random variables $\phi := \frac{\log(1+\theta\tau)}{\tau}$, $\mu := E[\phi|p, \tau]$, $\bar{\phi} := E[\phi|R=r, p, \tau]$. In words, ϕ is the approximation to θ that we obtain from our log-transformed data. The variable μ is the average of ϕ for all consumers who are marginal at price p and tax rate τ . And the variable $\bar{\phi}$ takes on three different values for consumers marginal at price p and tax rate τ : amongst the marginal consumers with $R=r$, it is the average of ϕ for those consumers. For short-hand, we set $\bar{\theta}_r := E[\bar{\phi}|R=r]$; this is the average ϕ across all consumers with $R=r$ (without conditioning on a price–tax pair).

Proposition 6.

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

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

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

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

Proposition 6 shows that $E_{p_1, \tau}[Var[\theta|p_1, \tau]]$ can be bounded from below by the significantly easier-to-estimate expression in Equations (9)–(11). The expression in Equations (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 Equations (9)–(11), the expression corrects for the fact that $Var[\bar{\theta}_R]$ would overestimate $E_{p_1, \tau}[Var[\theta|p_1, \tau]]$ if all individual differences in θ were due simply to variation in (p_1, τ) .

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

The idea behind the proof of Proposition 6, which is contained in the [Online Appendix](#), is as follows. First, we show that $E_{p_1, \tau}[Var[\theta|p_1, \tau]] \geq E\left[Var\left[\frac{\log(1+\theta\tau)}{\tau} \middle| \tau, p_1\right]\right]$, 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 $E_{p_1, \tau}[Var[\theta|p_1, \tau]] \geq E[Var[\bar{\phi}|p_1, \tau]]$. Third, we arrive at the final quantity in Equations (9)–(11) through an application of the Cauchy–Schwarz inequality.

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

Although in principle one could attempt to use self-classifying survey questions to estimate the statistic in Equation (8), in practice it is estimable to a far lower degree of precision than the statistic in Equations (9)–(11).³²

5.3. A lower bound for the variance of mistakes: estimation

A challenge in estimating the lower bound from Proposition 6 is estimating the terms $E(\mu|R=r)$. Because our data set is finite, we cannot obtain an estimate of each $\mu(p_1, \tau)$ for each pair (p_1, τ) . Instead, we partition the price–tax space into small cells of positive measure, and estimate an average value of $\frac{\log(1+\theta_{ik}\tau_i)}{\tau_i}$ within each cell. Formally, let $\{p_j\}_{j=1}^{15}$ denote the fifteen cells $[0, 1], [1, 2], \dots, [14, \infty)$ and let $\{\tau_j\}_{j=1}^5$ 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

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

As the cell sizes converge to zero, $\tilde{\mu}$ will converge to μ . To estimate the lower bound we simply replace each theoretical moment with its empirical moment counterpart, and we bootstrap the standard errors of the estimators. See [Online Appendix D.3](#) for further details of the empirical implementation.

Table 8 presents the results. The top row displays our estimates of the lower bound: 0.132 for the standard-tax arm and 0.094 for the triple-tax arm. To benchmark these estimates, consider the variances that would arise if consumers fully processed ($\theta = 1$) or completely neglected ($\theta = 0$) the tax. 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 1,000 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 arm³³; in the triple-tax arm, all effect sizes are three times larger, and thus the relative variance of noise is nine times smaller.

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

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

33. The source of the bias is that any noise in our estimates of $\bar{\theta}_r$ or $E[\mu|R=r]$ amplifies our estimates of variance because it involves squares of imperfectly estimated moments.

TABLE 8
Lower bound estimates for the expected conditional variance of θ (weight placed on tax)

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

Notes: This table presents lower bounds for $E_{p_1, \tau}[Var[\theta|p_1 \tau]]$, estimated for both the standard- and triple-tax arms using the methodology of Section 5. θ is defined as the “weight” that consumers place on the sales tax, with $\theta=0$ corresponding to complete neglect of the tax and $\theta=1$ corresponding to full optimization. We compute standard errors and mean bias (Efron, 1982) using the percentile (non-accelerated) bootstrap (with 1,000 iterations), blocked by consumers. We compute approximate 95% confidence intervals using the unadjusted bootstrap, as well as the median bias correcting bootstrap (Efron, 1987).

5.4. *Alternative approaches*

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

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

This approach could, in theory, be modified to deconvolve the variance of measurement error (*i.e.* random fluctuation in BDM valuations) and the variance of misreaction. Indeed, the no-tax arm of our experiment was designed to identify the variance of the measurement error term, so long as two concerns were avoided. As practical considerations, if the variance of measurement error encountered is either large or arm-specific, a deconvolution approach would be ill-powered or unidentified, respectively. As a theoretical concern, the presence of rounding heuristics in BDM responses generates additional variance that confounds the deconvolution (though this issue does not confound first-moment estimates and thus our bounding approach).³⁴

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

5.5. *Summary of empirical results*

To summarize our empirical results: we find substantial evidence of heterogeneous inattention to sales taxes. This heterogeneity is found across tax levels: under standard taxes, average attention is given by $\theta=0.25$, whereas under triple taxes average attention increases to $\theta=0.48$. Furthermore,

34. In practice, about 40% of the decisions in our study are within a few cents of a round number, suggesting that subjects do engage in some rounding behaviour.

this heterogeneity is found across individuals: under standard taxes, $E_{p_1, \tau}[\text{Var}[\theta|p_1, \tau]] > 0.13$ and under triple taxes $E_{p_1, \tau}[\text{Var}[\theta|p_1, \tau]] > 0.09$.

6. FROM EMPIRICAL MAGNITUDES TO WELFARE IMPLICATIONS

We now use the theoretical results from Section 2 to translate the experimental results from summarized in Section 5.5 into their implied welfare consequences. We assess our welfare estimates relative to a benchmark that assumes that misreaction is exogenous and homogeneous, and find that this benchmark substantially understates the welfare costs of taxation.

6.1. Individual differences

To translate the estimates from Section 5.5 into excess burden estimates, we use the formula in Proposition 2, which expresses excess burden in terms of the mean and variance of θ . To provide maximally conservative estimates, we suppose that supply is perfectly elastic because, as shown in Proposition 2, the relative importance of individual differences increases as the elasticity of supply decreases.³⁵ For the illustrative calculations here, we approximate $E[\theta|p, t]$ with our estimate of average θ , and we bound $\text{Var}[\theta|p, t]$ with our lower-bound estimate of $E_{p, t}[\text{Var}[\theta|p, t]]$.

Let EB_{NC} 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.³⁶ Let EB_H be the excess burden that would be computed by an analyst who assumes that θ is homogeneous, and knows the mean θ by, say, estimating D_t/D_p .³⁷ Finally, let EB denote the actual excess burden.

Consider now the implications of heterogeneity for welfare inferences. For the standard-tax arm, $EB_H \approx (0.25)EB_{NC}$. However, by Proposition 2, the actual excess burden is $EB \geq (0.25 + 0.132/0.25)EB_{NC} = (0.78)EB_{NC}$. For the triple-tax arm, $EB_H \approx (0.48)EB_{NC}$. However, by Proposition 2, the actual excess burden is $EB \geq (0.48 + 0.094/0.48)EB_{NC} = (0.68)EB_{NC}$.

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

6.2. Endogenous attention

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

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

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

37. As shown CLK (and replicated in Appendix Proposition A.1 for unit demand), the ratio D_t/D_p identifies θ for homogeneous consumers for small t . As shown in the proof of Online Appendix Proposition A.2 and implicitly used in the result, it is more generally true that $D_t/D_p = E[\theta|p, t]$ for small t .

and exogenous θ . Under the assumption that $F(\theta|p, t)$ is degenerate, Proposition 5 implies that

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

Consider now inferences under the assumption of homogeneous and exogenous θ . Suppose that the analyst computes $E[\theta] = 0.25$ by studying responses to standard taxes. Then assuming exogenous (and homogeneous) θ , the analyst would infer the excess burden of tripling the tax to be $4t^2(0.25)^2 D_p = 0.25t^2 D_p$. In this case, the endogeneity of θ with respect to t implies that the correct estimate is 336% higher.³⁸

7. DISCUSSION

In this article, 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 under-reaction to a not-fully-salient sales tax, this increases the efficiency costs arising from that tax’s distortionary effect on demand. When under-reaction varies with economic incentives, this affects the demand response to new policies and introduces a new channel by which taxes distort behaviour. Estimates from our experimental population suggest that these dimensions of variation exist, are sizable in magnitude, and can starkly affect the welfare analysis of tax policies.

These issues are of course not unique to sales taxes, and arise in any question about tax policy. And more broadly, these issues arise in any setting where the true price of a good is divided into different components of differing salience. The theoretical framework we develop in Section 2 can be easily extended to accommodate related shrouded attributes, and can therefore serve as a template for robust behavioural welfare analysis.

While we believe our theoretical framework is broadly portable, caution is needed when using our experimental estimates to assess welfare in external settings. When implementing our experiment, we devoted significant effort and resources to recruiting a broad and diverse subject population, and to making our experiment as natural as possible despite the unusual presence of a varying tax rate. 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 BDM 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. An alternative experimental design that potentially avoids this worry (at the cost of reduced experimental power) presents “take it or leave it” offers, in which consumers directly indicate whether they would purchase an item at some fixed price. Previous experiments employing this design have found evidence of average inattention to sales taxes (Feldman and Ruffle, 2015; Feldman *et al.*, 2015; Taubinsky, 2017). Furthermore, Taubinsky (2017) replicates the primary empirical estimates of this article under this alternative experimental format.

Second, the population used in our study is likely non-representative. Despite matching the U.S. population on several key observable demographics, unobserved characteristics could

38. The analysis above could be repeated to take the variance of θ into account by substituting our lower-bound variance estimates. This yields very similar results, since the variance lower bounds are very similar—0.132 and 0.094—and are not statistically distinguishable. Using the variance lower bounds to compute the incremental impact on excess burden is justifiable if the *within-bin* variances are not impacted by the size of the tax.

influence selection into our online survey platform. However, were heterogeneity in mistakes not present in the general population, it would not be found in arbitrary subsamples; as such, we do not view these issues as a hindrance to a demonstration that meaningful heterogeneity exists. We view our measurement of these statistics as an initial step, and proof of concept, of a necessary empirical agenda working towards robustly incorporating heterogeneity into behavioural welfare analysis.

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

Acknowledgments. For helpful comments and suggestions, we thank Hunt Allcott, Eduardo Azevedo, Doug Bernheim, Raj Chetty, Stefano DellaVigna, Sarah Beate Eichmeyer, Emmanuel Farhi, Xavier Gabaix, Jacob Goldin, Tatiana Homonoff, Shachar Kariv, Supreet Kaur, Judd Kessler, David Laibson, Erzo F.P. Luttmer, Matthew Rabin, Emmanuel Saez, Andrei Shleifer, Jeremy Tobacman, Glen Weyl, Michael Woodford, Danny Yagan, and audiences at the AEA Annual Meetings, Berkeley, Carnegie Mellon SDS, the CESifo Behavioral Economics Meeting, Columbia, Cornell, Dartmouth, Haas (marketing), Federal Trade Commission, Harvard, the National Tax Association Annual Meetings, New York University, Stanford, Wharton, and Yale. 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.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

REFERENCES

- ABALUCK, J. and GRUBER, J. (2011), "Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program", *American Economic Review*, **101**, 1180–1210.
- ABELER, J. and JÄGER, S. (2015), "Complex Tax Incentives", *American Economic Journal: Economic Policy*, **7**, 1–28.
- ALLCOTT, H., MULLAINATHAN, S. and TAUBINSKY, D. (2014), "Energy Policy with Externalities and Internalities", *Journal of Public Economics*, **112**, 72–88.
- ALLCOTT, H. and TAUBINSKY, D. (2015), "Evaluating Behaviorally-Motivated Policy: Experimental Evidence from the Lightbulb Market", *American Economic Review*, **105**, 2501–2538.
- ALLCOTT, H. (2011), "Consumers' Perceptions and Misperceptions of Energy Costs", *American Economic Review*, **101**(3), 98–104.
- (2015), "Paternalism and Energy Efficiency: An Overview" (NBER Working Paper No. 20363).
- ANDERSEN, S., HARRISON, G. W., LAU, M. I., *et al.* (2006), "Elicitation using Multiple Price List Formats", *Experimental Economics*, **9**, 383–405.
- AUERBACH, A. J. (1985), "The Theory of Excess Burden and Optimal Taxation", in Auerbach, A. and Feldstein, M., (eds), *Handbook of Public Economics*. (Elsevier Science Publishers B. V.) 61–128.
- BENJAMIN, D. J., HEFFETZ, O., KIMBALL, M. S., *et al.* (2014), "Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference", *American Economic Review*, **104**, 2698–2735.
- BENZARTI, Y. and D. CARLONI (2016), "What Goes Up May Not Come Down: Asymmetric Passthrough of Consumption Taxes" (Working Paper).
- BERNHEIM, B. D., and A. RANGEL (2009), "Beyond Revealed Preference: Choice-theoretic Foundations for Behavioral Welfare Economics", *The Quarterly Journal of Economics*, **124**, 51–104.
- BESLEY, T. J. and ROSEN, H. S. (1999), "Sales Taxes and Prices: an Empirical Analysis", *National Tax Journal*, **52**, 157–178.
- BHARGAVA, S. and MANOLI, D. (2015), "Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment", *American Economic Review*, **105**, 3489–3529.
- BORDALO, P., GENNAIOLI, N. and SHLEIFER, A. (2017), "Memory, Attention, and Choice" (NBER Working Paper No. 23256).
- BUSHONG, B., SCHWARTZSTEIN, J. and RABIN, M. (2015), "A Model of Relative Thinking" (Working Paper).
- CAPLIN, A., and DEAN, M. (2013), "The Behavioral Implications of Rational Inattention with Shannon Entropy" (NBER Working Paper No. 19318).

- (2015a), “Revealed Preference, Rational Inattention, and Costly Information Acquisition”, *American Economic Review*, **105**, 2183–2203.
- (2015b), “Revealed Preference, Rational Inattention, and Costly Information Acquisition” (NBER Working Paper No. 19876).
- CHETTY, R., FRIEDMAN, J. N. and SAEZ, E. (2013), “Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings”, *American Economic Review*, **103**, 2683–2721.
- CHETTY, R., LOONEY, A. and KROFT, K. (2007), “Salience and Taxation: Theory and Evidence” (NBER Working Paper No. 13330).
- (2009), “Salience and Taxation: Theory and Evidence”, *American Economic Review*, **99**, 1145–1177.
- CHETTY, R. and SAEZ, E. (2013), “Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients”, *American Economic Journal: Applied Economics*, **5**, 1–31.
- CHETTY, R. (2009), “The Simple Economics of Salience and Taxation” (NBER Working Paper No. 15246).
- (2012), “Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply”, *Econometrica*, **80**, 969–1018.
- (2015), “Behavioral Economics and Public Policy: A Pragmatic Perspective”, *American Economic Review Papers and Proceedings*, **105**, 1–33.
- CLARK, J. and FRIESEN, L. (2008), “The Causes of Order Effects in Contingent Valuation Surveys: An Experimental Investigation”, *Journal of Environmental Economics and Management*, **56**, 195–206.
- DELLAVIGNA, S. (2009), “Psychology and Economics: Evidence from the Field”, *Journal of Economic Literature*, **47**, 315–372.
- DE BARTOLOME, C. A. M. (1995), “Which Tax Rate do People Use: Average or Marginal?”, *Journal of Public Economics*, **56**, 79–96.
- DOYLE, J. J. and SAMPHANTHARAK, K. (2008), “2.00 Dollar Gas! Studying the Effects of a Gas Tax Moratorium”, *Journal of Public Economics*, **92**, 869–884.
- EFRON, B. (1982), “The Jackknife, The Bootstrap, and Other Resampling Plans”, *Society of Industrial and Applied Mathematics CBMS-NSF Monographs*, **38**. doi: 10.1137/1.9781611970319.
- EFRON, B. (1987), “Better Bootstrap Confidence Intervals”, *Journal of the American Statistical Association*, **82**, 171–185.
- FARHI, E. and GABAIX, X. (2015), “Optimal Taxation with Behavioral Agents” (NBER Working Paper No. 21524).
- FELDMAN, N. E., KATUSCAK, P. and KAWANO, L. (2016), “Taxpayer Confusion: Evidence from the Child Tax Credit”, *American Economic Review*, **106**, 807–835.
- FELDMAN, N. E. and RUFFLE, B. J. (2015), “The Impact of Including, Adding, and Subtracting a Tax on Demand”, *American Economic Journal: Economic Policy*, **7**, 95–118.
- FELDMAN, N., GOLDIN, J. and HOMONOFF, T. (2015), “Raising the Stakes: Experimental Evidence on the Endogeneity of Taxpayer Mistakes” (Working Paper).
- FINKELSTEIN, A. (2009), “E-ZTAX: Tax Salience and Tax Rates”, *The Quarterly Journal of Economics*, **124**, 969–1010.
- GABAIX, X. and LAIBSON, D. (2006), “Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets”, *Quarterly Journal of Economics*, **121**.
- GABAIX, X. (2014), “A Sparsity Based Model of Bounded Rationality”, *Quarterly Journal of Economics*, **129**, 1661–1710.
- GLAESER, E. L. and LUTTMER, E. F. P. (2003), “The Misallocation of Housing Under Rent Control”, *American Economic Review*, **93**, 1027–1046.
- GOLDIN, J. and HOMONOFF, T. (2013), “Smoke Gets in Your Eyes: Cigarette Tax Salience and Regressivity”, *American Economic Journal: Economic Policy*, **5**, 302–336.
- GOLDIN, J. (2015), “Optimal Tax Salience”, *Journal of Public Economics*, **131**, 115–123.
- HARBERGER, A. C. (1964), “The Measurement of Waste”, *American Economic Review*, **54**, 58–76.
- HEIDHUES, P., KÓSZEGI, B. and MUROOKA, T. (2017), “Inferior Products and Profitable Deception”, *Review of Economic Studies*, **84**, 323–356.
- HOOPES, J., RECK, D. H. and SLEMMOD, J. (2015), “Taxpayer Search for Information: Implications for Rational Attention”, *American Economic Journal: Economic Policy*, **7**, 177–208.
- HOSSAIN, T. and MORGAN, J. (2006), “...Plus Shipping and Handling: Revenue (Non) Equivalence in Field Experiments on eBay”, *Advances in Economic Analysis and Policy*, **5**, 1–30.
- LIEBMAN, J. B. and ZECKHAUSER, R. (2004), “Schmeduling” (Working Paper).
- LOCKWOOD, B. and TAUBINSKY, D. (2017), “Regressive Sin Taxes” (NBER Working Paper No. 23085).
- MULLAINATHAN, S., SCHWARTZSTEIN, J. and CONGDON, W. J. (2012), “A Reduced-Form Approach to Behavioral Public Finance”, *Annual Review of Economics*, **4**, 1–30.
- RECK, D. H. (2014), “Taxes and Mistakes: What’s in a Sufficient Statistic?” (Working Paper).
- REES-JONES, A. and TAUBINSKY, D. (2016), “Heuristic Perceptions of the Income Tax: Evidence and Implications for Debiasing” (NBER Working Paper No. 22884).
- TAUBINSKY, D. (2017), “Deliberate Inattention to Shrouded Attributes: New Evidence from Consumers’ Over- and Under-Reaction to Sales Taxes” (Working Paper).
- VEIGA, A. and WEYL, G. (2016), “Product Design in Selection Markets”, *Quarterly Journal of Economics*, **131**, 1007–1056.
- WOODFORD, M. (2012), “Inattentive Valuation and Reference-Dependent Choice” (Working Paper).