

NBER WORKING PAPER SERIES

RULES OF THUMB AND ATTENTION ELASTICITIES:  
EVIDENCE FROM UNDER- AND OVERREACTION TO TAXES

William Morrison  
Dmitry Taubinsky

Working Paper 26180  
<http://www.nber.org/papers/w26180>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
August 2019, Revised March 2021

We are especially grateful to Alex Rees-Jones for many helpful conversations and input that led to this project. We are also indebted to Xavier Gabaix for guiding us through the application of his model. We also thank Sandro Ambuehl, Ned Augenblick, Andrew Caplin, Mark Dean, Stefano DellaVigna, Shachar Kariv, Patrick Kline, John Leahy, Daniel Martin, Filip Matejka, Josh Schwartzstein, Amitai Shenhav, and Philipp Strack, as well as conference and seminar participants, for helpful comments. Afras Sial provided excellent research assistance. Taubinsky thanks the Alfred P. Sloan Foundation for financial support. The experiment was approved by the Dartmouth College Committee for the Protection of Human Subjects (CPHS), # STUDY00029784. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by William Morrison and Dmitry Taubinsky. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Rules of Thumb and Attention Elasticities: Evidence from Under- and Overreaction to Taxes  
William Morrison and Dmitry Taubinsky  
NBER Working Paper No. 26180  
August 2019, Revised March 2021  
JEL No. D9,H2

**ABSTRACT**

This paper tests costly attention models of consumers' misreaction to opaque taxes. We report an online shopping experiment that involves shrouded sales taxes that are exogenously varied within consumer over time. Some consumers systematically underreact to sales taxes while others systematically overreact, but higher stakes decrease both under- and overreaction. This is consistent with consumers using heterogeneous rules of thumb to compute the opaque tax when the stakes are low, but using costly mental effort at higher stakes. The results allow us to differentiate between various theories of limited attention. We also develop novel econometric techniques for quantifying individual differences.

William Morrison  
University of California at Berkeley  
530 Evans Hall  
MC #3880  
Berkeley, CA 94720  
wmorrison@berkeley.edu

Dmitry Taubinsky  
University of California, Berkeley  
Department of Economics  
530 Evans Hall #3880  
Berkeley, CA 94720-3880  
and NBER  
dmitry.taubinsky@berkeley.edu

# 1 Introduction

Economists have long theorized that cognitive resources are limited, and that individuals may simplify complex decisions by deliberately using heuristic shortcuts or by processing only a subset of available information (for recent reviews, see Caplin, 2016; Maćkowiak et al., 2018; Gabaix, 2019). This view is in line with the *resource rationality* framework in the cognitive sciences (Lieder and Griffiths, 2019), which recognizes “mental effort as a domain of decision-making” (Shenhav et al., 2017).

For example, when choosing whether or not to buy a product sold for a posted price of \$17.99 and a sales tax rate of 7%, some consumers might reduce the cognitive burden of computing the total after-tax price by instead choosing to ignore the sales tax completely. Other consumers might approximate the sales tax with a rough sense of how much tax they usually pay when they buy ~\$17.99 worth of products, including instances in which not all of the products are subject to the tax. And yet other consumers might approximate the tax to be negligibly less than 10% of \$17.99, which they compute easily by moving the decimal point one digit to the left.

In the first two of these example cases, the consumers *underreact* to sales taxes—they behave as if the taxes are smaller than they are. In the last case, the consumers *overreact*. When purchasing expensive electronics or an automobile, however, consumers may choose to exert more cognitive effort to compute the actual price that they would end up paying, thereby reducing their propensity to both over- and underreact.

Prior literature on sales tax salience has convincingly shown that *on average*, individuals underreact to opaque sales taxes.<sup>1</sup> However, existing results about averages do not preclude that some individuals overreact, and provide little evidence about the degree to which individuals’ misreaction to sales taxes is due to deliberate, and plausibly elastic, use of cognitive shortcuts. In this paper, we provide a series of tests, grounded in models of costly attention, that fill this gap. In doing so, we develop a methodology for testing models of costly attention that could be applied to other domains with opaque attributes that are imperfectly processed by consumers—energy prices (Allcott and Wozny, 2014; Allcott and Taubinsky, 2015), shipping and handling charges (Hossain and Morgan, 2006), various features of health insurance contracts (Handel and Kolstad, 2015; Bhargava et al., 2017; Abaluck and Adams, forthcoming), less significant digits (Lacetera et al., 2012; Shlain, 2019), shrouded financial fees (Heidhues et al., 2017), and add-on charges (Gabaix and Laibson, 2006).

We begin our investigation in Section 2 by formalizing the economic environment and several types of costly attention models. Consumers must decide whether or not to buy a good or service that has both a transparent posted price and an opaque price. Consumers have a prior perception of the post-tax price that they can access costlessly, and which can vary between consumers, as in our example. We consider several formulations of the cognitive costs of updating: the Shannon cost function used in rational inattention models,<sup>2</sup> and the attention-weight adjustment cost function of Gabaix (2014).

---

<sup>1</sup>See Chetty et al. (2009); Goldin and Homonoff (2013); Feldman and Ruffle (2015); Taubinsky and Rees-Jones (2018); Feldman et al. (2018); Bradley and Feldman (2020); Kroft et al. (2020)

<sup>2</sup>This formulation leads to a model that is almost identical to rational inattention models, with one exception: because we allow priors to be heterogeneous, we allow for systematically biased perceptions of the true value. This heterogeneity is necessary to capture individual differences in the tendency to either under- or overreact to the sales tax, which we show are very significant in our data. This clarification is meant only for readers who define *rational* inattention as having systematically unbiased beliefs.

We establish that both types of cost functions have a simple reduced-form representation in our economic setting: both models lead to consumer behavior that looks *as if* the consumer places some (possibly stochastic) weight on the opaque price (e.g., Chetty et al., 2009; DellaVigna, 2009). We call this weight the revealed valuation weight, or just valuation weight for short, because it is easily estimated from observable price variation, as in the reduced-form regression models used to quantify under- and overreaction in applied empirical work. In the context of sales taxes, a valuation weight of, e.g.,  $\theta = 0.4$  means that imposing a sales tax of size  $t$  decreases demand as much as increasing posted prices by  $0.4t$  would decrease demand. In other words,  $\theta = 0.4$  means that consumers are 40 percent as responsive to taxes as fully attentive consumers would be.

The underlying costly attention models discipline the reduced-form valuation weights in several straightforward and economically meaningful ways. First, they imply that if there are individual differences, then these should be persistent across different levels of stakes; e.g., consumers who tend to overreact at moderate stakes should also tend to overreact at higher stakes. Second, the costly attention models imply that the valuation weights should approach one as the stakes increase. In settings such as those of Chetty et al. (2009), where consumers underreact to sales taxes on average, the average underreaction must thus decrease as the stakes increase (e.g., as the sales tax rate increases). A more discerning empirical test, however, is that the higher is the valuation weight at moderate stakes, the smaller is the degree by which it increases when stakes increase. In particular, the valuation weights should decrease for consumers who overreact and increase for consumers who underreact.

We test these predictions in the context of a prominent and policy-relevant domain of behavior: consumer response to sales taxes not included in posted prices. Because the strongest tests of costly attention models concern individual differences in how the sign and magnitude of misreaction are impacted by stakes, we develop a new experimental design in which the size of the tax rate is varied exogenously within consumers over time. Existing empirical work does not feature within-consumer variation in stake size, and thus cannot be used to test these predictions.

Our experiment features 1534 demographically diverse consumers from the forty-five U.S. states with positive sales taxes. The experiment utilizes an online shopping environment with nine different non-tax-exempt household products, such as cleaning supplies. Each consumer encounters three of the nine products in three different types of “stores” at random posted prices. The three different types of stores feature 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 consumer’s city of residence. Each consumer thus encounters  $3 \times 3$  product by store pairs, with each associated to a set of random prices. Decisions in the experiment are incentive-compatible: study participants receive a \$16 budget to potentially buy one of the randomly chosen products in one of the randomly chosen stores, and purchased products are shipped to their homes.

We begin our empirical analysis with a very simple test in Section 4: we estimate average underreaction to taxes of varying size, exploiting both the exogenous variation in prices and the exogenous variation in tax rates. Although this test has been implemented in several prior studies, our analysis is significantly better-powered and is unique in exploiting variation in both prices and tax rate sizes. We find striking evidence that misreaction depends on stakes. The average valuation weight is 0.23 for the smallest price at standard tax

rates—meaning that at these stakes consumers are only 23 percent as responsive to taxes as fully attentive consumers would be. However, the average valuation weight is 0.79 for the largest price at triple tax rates—meaning that at these stakes consumers are 79 percent as responsive to taxes as fully attentive consumers would be. The average increases monotonically in the absolute size of the tax, and in a manner that is invariant to whether the absolute size of the tax is high because the tax rate is high or because the price is high. Direct tests of consumers’ knowledge and computational ability suggest that our findings are more consistent with consumers knowing the tax rate and exerting effort to compute the tax-inclusive price, rather than engaging in information acquisition about the tax rate.

In Section 5 we begin testing our novel predictions about individual differences in responses to stakes. As we discuss in that section, a key challenge for tests of individual differences is that individual-level estimates of the valuation weights will necessarily involve significant measurement error. Thus, making inferences about individual-level differences merely from a distribution of individual-level point estimates would yield confounded conclusions if one took the point-estimates at face-value. This challenge is not unique to our setting, and poses problems for most within-subject experiments seeking to quantify individual differences (including ones where authors choose to report individual-level point estimates nonetheless). To overcome this challenge and generate simple reduced-form tests of our individual-level predictions, we leverage the multiple decisions feature of our design to examine whether people who seem to be most sensitive to taxes on one product react differently to taxes on the *other two* products.

Consistent with real individual differences, we find that consumers who respond the most to standard taxes on one product have a much higher valuation weight on standard taxes on the other two products. Consistent with the prediction that individual differences are persistent across stakes, we also find that consumers who respond to standard taxes the most on one product are more responsive to the triple taxes on the other two products. Our design ensures that these results cannot be confounded by measurement error because the fully random presentation of products and tax environments ensures that measurement errors are independent conditional on the true value of a valuation weight.

We then establish two key empirical results that are consistent with the prediction that valuation weights should get closer to one as stakes increase. First, when the tax rates are tripled, consumers who respond to standard taxes the least on one product exhibit a significantly larger increase in their valuation weights on taxes on the other two products. Second, when we instead use one product to split consumers into groups based on how much they adjust their valuation weight as stakes increase, we find that the least sensitive consumers have much higher valuation weights on taxes on the other two products in both the standard tax regime *and* in the triple tax regime. This second result is consistent with the prediction that the smallest valuation weight changes should occur for consumers with the highest prior perceptions, which translate to high valuation weights in both the standard and high stakes environments.

Having established significant and persistent individual differences in valuation weights, as well as heterogeneous attention responses to higher stakes that are consistent with costly attention models, we ask two key questions in Section 6. First, are the individual differences large enough that some consumers overreact to standard taxes? If so, can we show that some consumers decrease their valuation weights when the stakes increase?

To answer these questions, we develop new econometric techniques for bounding individual differences. First, we develop a new approach that produces a lower bound on the variance of the valuation weights. The approach is in the spirit of instrumental variable corrections that leverage double observations of mismeasured right-hand-side variables in regressions (e.g., Hausman, 2001; Gillen et al., 2019). Second, we develop a concentration inequality approach that uses our point estimates of means and our estimates of variance bounds to form non-parametric bounds on several properties of the distribution of valuation weights.

We find that at standard tax rates, the maximum of the valuation weights must be at least 2.21 (5% confidence bound of 1.55). This implies that at least some consumers overreact to taxes significantly. To our knowledge, our finding of overreaction by some consumers is new to the literature, and runs counter to researchers’ priors. Consistent with the presence of overreaction in costly attention models, we also estimate that some consumers *reduce* their valuation weight by at least 0.94 (5% confidence bound of 0.16) when shopping in the triple tax stores instead of the standard tax store.

Our paper contributes to several literatures. First, our paper contributes to a recent literature that experimentally tests models of costly attention.<sup>3</sup> With the exception of Bartos et al. (2016), these papers utilize abstract information acquisition and problem-solving tasks to provide comprehensive tests of core assumptions of the models. Our paper complements this literature by focusing on a concrete empirical setting, and asking whether the “mistakes” identified by reduced-form empirical work in that domain fit the patterns of costly attention models.

By focusing on the concrete setting of opaque sales taxes, and opaque prices more broadly, our paper also deepens the empirical work in those settings. To our knowledge, empirical work in these settings has not tested predictions about individual differences in attentional responses to stakes. Gabaix (2019) provides suggestive evidence from cross-study comparisons that average underreaction to opaque prices decreases with stakes. This paper continues this line of argument by providing rigorous experimental evidence of this phenomenon, including a series of more nuanced tests of costly attention models that move beyond comparative statics on average underreaction and toward more demanding predictions about attentional responses to stakes at the individual level.

While focusing mostly on normative implications of tax salience and using an experimental design importantly different from ours in key ways, Taubinsky and Rees-Jones (2018, henceforth TRJ) are closest to our work in that they also find that average underreaction decreases when tax rates are saliently increased. However, TRJ cannot quantify how average underreaction varies by pre-tax price, and thus cannot rule out that their results are driven by other possibilities such as consumers overreacting to a surprising change that violates their shopping “norms” (Bordalo et al., 2020). More importantly, the lack of within-consumer variation in tax rates in the TRJ data makes testing core predictions about individual differences in how attention responds to stakes infeasible.

Overall, we test five key predictions in this paper, as summarized in Section 2.4. The experimental design employed by TRJ allows only a partial test of the first prediction, and is infeasible for testing the

---

<sup>3</sup>See, e.g., Gabaix et al. (2006); Bartos et al. (2016); Martin (2016); Dean and Neligh (2019); Ambuehl et al. (2018); Caplin et al. (2020); Carvalho and Silverman (2019).

other four predictions, which are more discerning and novel. Consequently, the results in our paper are uniquely suitable for discerning between various theories of inattention. Unlike TRJ, we can reject models where no one overreacts to the taxes, which rejects the important class of costly attention models with common priors, as well as the bounded-rationality model in Chetty et al. (2007). Unlike TRJ, we can also reject theories in which individuals’ priors about the opaque price are not persistent across stake size, which is crucial for inferring that the decrease in average underreaction after increases in tax rates is indeed due to costly mental adjustment from a heuristic rule-of-thumb, rather than people simply relying on different rule-of-thumb strategies. In Appendix E we provide a detailed comparison to TRJ as well as other studies of tax salience.

Finally, our paper contributes econometric techniques for studying individual differences in the presence of measurement error. While there is a large literature on techniques for addressing measurement error in regression analysis (e.g., Hausman, 2001; Gillen et al., 2019), we introduce techniques for inference about the variance of a noisily measured variable. We then develop concentration inequality approaches to translate bounds on the variance to bounds on several properties of the distribution. Extensions of our approach have been could be used to provide formal statistical evidence for other questions about individual differences. For example, Mueller et al. (2021) adapt our approach to study individual differences in job-finding rates, and other extensions could involve questions about whether some individuals are risk-loving or future-biased.

## 2 Theoretical framework for hypothesis development

### 2.1 Setup

Consumers have unit demand for a good  $x$  and spend their remaining money on an untaxed composite good  $y$  (the numeraire). We assume quasilinear preferences: the utility of purchasing good  $x$  is given by  $v x - p$ , where  $x \in \{0, 1\}$ ,  $v$  is the utility from the product, and  $p$  is its total price.

The total price consists of a salient component  $p_s$  and an opaque component  $p_o$ , with  $p = p_s + p_o$ . In our empirical application,  $p_s$  represents the displayed price of the product while  $p_o$  represents the sales tax.

Consumers costlessly incorporate  $p_s$  into their decision, but may have trouble properly processing  $p_o$ . We endow  $p_o$  with the structure  $p_o = \sigma q_o$ , where  $\sigma$  is a parameter that is known to the consumers and represents the “stakes” involved, while  $q_o$  is the part that may be misprocessed. For example, a salient announcement that sales taxes will be tripled is likely to be fully noted by consumers, and corresponds to an increase in  $\sigma$ . As another example, consider  $p_o = p_s q_o$ , where  $q_o$  corresponds to the sales tax rate and  $p_o$  is the tax owed on an item sold for a posted price of  $p_s$ .

As a simple and illustrative baseline, which we generalize in the appendix, we assume that when consumers do not exert mental effort their baseline representation of  $q_o$  is given by prior beliefs that place probability  $r$  on its true value  $t$  and probability  $1 - r$  on some other value  $\hat{t}$ . This generates a heuristic, “rule-of-thumb” estimate of the opaque price  $\hat{p}_o = \sigma r t + \sigma (1 - r) \hat{t}$ .

Consumers must pay cognitive costs to better take the opaque price into account. Their choice of whether

or not to pay this cost depends on their prior. This is in contrast to “ex-post” attentional rules such as those in Chetty et al. (2007), according to which the consumer knows the ex-post benefit of paying attention before exerting any cognitive effort.<sup>4</sup> For example, consumers who are very confident in their assessment will not bother to exert mental costs. We detail the link between mental effort and improvements to the prior perception in the subsections that follow.

As an example of the prior perceptions that could be captured by our formalism, consider individuals who have a sense of how much tax they usually pay on average over all items they buy, both those subject to a tax and those that are not. A prior perception based on this loose recollection could be modeled by setting  $\hat{t} = 0$ , with  $r$  corresponding to the frequency of purchase occasions of taxable products. Cognitive costs could be expended to either improve recollection (Ratcliff, 1978) or to perform the computation directly without relying on memory samples. Alternatively, the model with  $\hat{t} = 0$  could correspond to individuals not being sure if the good is subject to the standard tax or not.

As another example,  $\hat{t} > t$  could capture individuals who without thinking would guess the sales tax to be somewhat lower than 10% of the posted price. Costly thinking could involve a series of steps to improve the approximation. For example, to compute a 7% tax, first compute 5% of the sales price as half of the 10% estimate, and then find a point that is approximately between the 5% and 10% estimates.

There are several ways to interpret our model. One is that consumers literally do not know  $p_o$ , and must search for information about it. Another, as in some of the examples above, is that consumers know what the value of  $p_o$  is, but have trouble integrating it into their decision-making. The prior over  $p_o$  might thus be interpreted as “computational uncertainty.” Our experimental data will allow us to differentiate between incorrect beliefs and computation costs as mechanisms for imperfect processing of  $p_o$ , providing more support for the latter.

We view the static costly attention models we work with to be “as if” models of this effortful thinking. Process-based models, such as sequential sampling (e.g., Fudenberg et al., 2018; Busemeyer et al. 2019), could provide more complete accounts of how the allocation of costly attention to a decision improves accuracy.<sup>5</sup>

## 2.2 Simple example with binary attention strategies

We begin with a simple example of a costly attention model, and show how it motivates the empirical tests we perform using our experiment. The simple model in this subsection is a special case of Gabaix (2014).

We suppose that computing the opaque price correctly is a binary decision: consumers can rely on their

---

<sup>4</sup>Unlike Chetty et al. (2007), we also allow arbitrary heterogeneity in prior beliefs, including overestimation. As we will show, our data strongly reject a model in which consumers either pay full attention to the tax or ignore it completely.

<sup>5</sup>Our mathematical framework could also be applied to cases in which the cost of computing the post-tax price is affective because, e.g., it is unpleasant to think about paying taxes. Feldman and Ruffle (2015) present results that could be consistent with affective costs.



initial perceptions or they can pay a cost  $\lambda$  to fully learn whether  $q_o = t$  or  $q_o = \hat{t}$ . If the consumer does not pay the cognitive cost, then he buys if and only if  $v - p_s \geq \hat{p}_o = r\sigma t + (1 - r)\sigma\hat{t}$ . If  $v - p_s > \sigma \max(t, \hat{t})$  then the consumer definitely buys, since there is no possibility that the total price exceeds the product value  $v$ . And if  $v - p_s < \sigma \min(t, \hat{t})$  then the consumer does not buy since there is no possibility that the total price is smaller than the product value  $v$ . Hence, we focus on the interesting case in which  $\sigma \min(t, \hat{t}) < v - p_s < \sigma \max(t, \hat{t})$ .

Suppose, first, that  $\hat{t} < t$ . If  $v - p_s < \hat{p}_o$  then the consumer does not buy the product if he does not pay an attention cost. If the consumer does pay an attention cost, then he learns that  $p_o = \sigma t > \hat{p}_o$ , and thus does not buy the product. Thus, if  $v - p_s < \hat{p}_o$  then the consumer does not buy the product.

If  $v - p_s \geq \hat{p}_o$  then the consumer buys if he does not pay an attention cost. The value of figuring out  $p_o$  is the value of averting a purchase if the opaque price is high:  $r(\sigma t + p_s - v)$ . Thus, the consumer pays the attention cost if  $\lambda < r(\sigma t + p_s - v)$ , or equivalently  $v - p_s < \sigma t - \lambda/r$ . Upon paying the attention cost, the consumer buys only if  $v - p_s > \sigma t$ , which cannot occur since the consumer only pays the attention cost when  $v - p_s < \sigma t - \lambda/r$ . Consequently, the consumer only buys in this case if he does not pay the attention cost.

In summary, the consumer buys if and only if both  $v - p_s \geq \hat{p}_o$  and  $v - p_s \geq \sigma t - \lambda/r$  hold. This behavior is equivalent to the behavior of a consumer who perceives  $p_o$  to be  $\tilde{p}_o = \theta p_o$ , and thus buys only if  $v - p_s \geq \theta \sigma t$ , where

$$\begin{aligned} \theta &= \frac{1}{\sigma t} \max(\hat{p}_o, \sigma t - \lambda/r) \\ &= \max\left(r + (1 - r)\frac{\hat{t}}{t}, 1 - \frac{\lambda}{\sigma t r}\right) < 1 \end{aligned} \quad (1)$$

Conversely, if  $\hat{t} > t$ , analogous reasoning implies that this behavior is equivalent to the behavior of a consumer who perceives  $p_o$  to be  $\tilde{p}_o = \theta p_o$ , for

$$\theta = \min\left(r + (1 - r)\frac{\hat{t}}{t}, 1 + \frac{\lambda}{\sigma t r}\right) > 1. \quad (2)$$

Notably, although the attention strategy depends on the transparent price  $p_s$ , we can still represent the consumer's behavior *as if* he weights  $p_o$  by some weight  $\theta$  that is independent of the price  $p_s$ . We call  $\theta$  the *revealed valuation weight*, as it is easily estimable from data. Concretely, consider a population of consumers who derive different utilities  $v$  from the product, but have the same valuation weight  $\theta$ . Let  $\Delta p$  be the decrease in the salient price  $p_s$  that generates the same change in demand as the removal of the opaque price  $p_o$ . Then by definition,  $p_s + \theta p_o - \Delta p = p_s$ , and thus  $\theta = \Delta p/p_o$ . We refer to  $\theta > 1$  as overreaction and  $\theta < 1$  as underreaction.

Importantly, the underlying model of costly attention puts structure on the relative degree of misreaction, and on its distribution in the population. First, any individual differences in  $\theta$ —generated by individual differences in priors ( $\hat{t}$  and  $r$ ) and in the cost of attention  $\lambda$ —must be persistent across stakes  $\sigma$ . In particular,  $\theta$  is increasing in  $\hat{t}$ , and  $|1 - \theta|$  is decreasing in  $r$  and increasing in  $\lambda$ .

Second,  $\theta$  is increasing in  $\sigma$  when  $\hat{t} < t$ , and is decreasing in  $\sigma$  when  $\hat{t} > t$ , with  $\lim_{\sigma \rightarrow \infty} \theta = 1$ . That is, as

stakes increase, the relative degree of misreaction decreases, and becomes arbitrarily small for sufficiently large stakes. Although intuitive, this comparative static holds only for the relative degree of misreaction  $|1 - \theta|$ . The absolute degree of misreaction,  $|p_o - \theta p_o|$ , is weakly increasing in  $\sigma$ .<sup>6</sup>

The fact that  $|1 - \theta| \rightarrow 0$  as  $\sigma \rightarrow \infty$  has several consequences. First, it implies that if  $E[\theta] < 1$  in the population, then increasing stakes should increase the average valuation weight. Second, it implies that if some individuals tend to overreact, then they should do so less when the stakes increase; that is,  $\theta$  falls with  $\sigma$  for individuals who overreact. More generally, this implies that the extent to which  $\theta$  increases with stakes  $\sigma$  is decreasing with the baseline level of  $\theta$ . Finally, if some individuals overreact, the individuals whose  $\theta$  fall the most as  $\sigma$  increases from  $\sigma_1$  to  $\sigma_2 > \sigma_1$  should on average have the highest  $\theta$  at both  $\sigma_1$  and  $\sigma_2$ .

### 2.3 Results for the Shannon model and the Gabaix (2014) Sparsity Model

While the simple example above involved a binary choice of either paying attention or not, the key results and predictions generalize to models with more continuous choice of attention. In Appendix A.1 we consider the Shannon model (Sims, 2003; Matejka and McKay, 2015; Caplin et al., 2019) of attention, in which individuals are free to choose any signal structure they wish, and pay attention costs that are linear in the expected reduction of entropy. In Appendix A.2 we consider the Gabaix (2014) model, where individuals pay an attention cost that is a monotonic function of the the difference between the posterior mean and the prior mean. We show that both models have a reduced-form representation where the consumer buys the product if its value  $v$  to the consumer exceeds  $p_s + \theta p_o$ , where  $\theta$  depends on  $p_o$  but not on  $p_s$ , and where  $\theta$  is stochastic in the Shannon model but deterministic in the Gabaix (2014). We show that all of the predictions about how the distribution of  $\theta$  varies with stakes are replicated in these two models of continuous attention.<sup>7</sup>

Finally, in Appendix B, we consider priors given by  $\hat{t} + \varepsilon$ , where  $E[\varepsilon] = 0$  and  $\hat{t}$  varies, and show that our main predictions hold under these more general assumptions as well.

### 2.4 Empirical tests of costly attention models

Following the intuition provided in the special case described in Section 2.2, and the more general results in Appendices A and B, our theoretical results provide five empirical tests. For concreteness, we focus on the case in which  $E[\theta] < 1$ , as our empirical application studies sales taxes, for which previous work has established underreaction. Consistent with our experiment, we consider a “standard stakes regime” (“standard” value of  $\sigma$ ) and a “high stakes regime” (higher value of  $\sigma$ ). All of the empirical tests are grounded in the core idea that individual differences persist across stakes, and that the revealed valuation

---

<sup>6</sup>For  $\hat{t} < t$ , for example,

$$p_o - \theta p_o = \sigma t - \max\left(\sigma t - \frac{\lambda}{r}, \sigma t r + (1-r)\sigma \hat{t}\right) = \min(\lambda/r, \sigma(1-r)(t - \hat{t})).$$

<sup>7</sup>In both models, individuals use their prior beliefs to assess the subjective expected value of incurring attention costs. This is more realistic than the assumption in Chetty et al. (2007), where before exerting any attention costs, individuals know the exact ex-post value of paying attention.

weights must approach 1 as stakes increase. The tests below correspond to different cuts of the data that can provide evidence for this idea.

**Prediction 1.** *The average revealed valuation weight,  $E[\theta]$ , is higher in the high stakes regime.*

**Prediction 2.** *There are stable individual differences that are persistent across stakes. Consumers with higher values of  $\theta$  in the standard stakes regime will also have higher values of  $\theta$  in the high stakes regime.*

**Prediction 3.** *Consumers with the highest values of  $\theta$  in the standard stakes regime will increase their  $\theta$  by the smallest amount when put in the high stakes regime.*

**Prediction 4.** *Consumers whose  $\theta$  increases the least in response to the high stakes regime have the highest values of  $\theta$  in both the standard and high stakes regimes.*

**Prediction 5.** *If some consumers have  $\theta > 1$  in the standard stakes regime then some consumers will adjust their  $\theta$  downward when put in the high stakes regime.*

Although mathematically straightforward, Prediction (4) is particularly demanding, including relative to Prediction (3). In essence, it is saying that if the distribution of  $\theta$  at two stakes levels  $\sigma_1$  and  $\sigma_2 > \sigma_1$  is given by the random variables  $X_1$  and  $X_2$ , then  $E[X_1|X_1 - X_2 = \Delta]$  and  $E[X_2|X_1 - X_2 = \Delta]$  are both increasing in  $\Delta$ . This implies a special structure on the joint distribution of  $X_1$  and  $X_2$ , as typically  $X_1 - X_2$  is “big” when  $X_1$  is “big” and  $X_2$  is “small” rather than “big.” As we discuss more in Section 2.4.1, heterogeneity in prior perceptions across stakes would cut strongly against this prediction.

Predictions 3 and 4 may fail when, e.g., all consumers have  $\theta < 1$  and consumers either have very high attention costs that make their misreaction  $1 - \theta$  inelastic to variation in stakes or they have moderate attention costs that make their misreaction moderately elastic. In this case, the consumers with moderate attention costs will both have higher  $\theta$  and increase their  $\theta$  the most—in other words, the predictions break down because the elasticity of misreaction is positively correlated with  $\theta$ . However, when some consumers significantly overreact, as we will show empirically, predictions 3 and 4 are very likely to continue to hold, because now consumers with the highest  $\theta$  will adjust downward. In Appendix D, we show formally that when there is sufficiently high variation in  $\theta$ , consistent with our empirical estimates, Predictions 3 and 4 continue to hold even when there are individual differences in the elasticity of misreaction that are correlated with  $\theta$ .

In Appendix D we show that Predictions 3 and 4 are mostly robust to the presence of heterogeneous attention costs.

#### **2.4.1 Using the tests to differentiate from other mechanisms of misreaction**

The predictions do not (collectively) hold for any other mechanism that generates misreaction to opaque prices. Thus, the predictions collectively serve to draw a sharp distinction between attention costs and other plausible mechanisms such as complete unawareness of the opaque price (e.g., Gabaix and Laibson, 2006), incorrect beliefs generated by systematic mislearning (e.g., Hanna et al., 2014) or misleading marketing (e.g., Anagol et al., 2017), forgetting (e.g., Bordalo et al., 2020), or a systemic lack of financial literacy that prevents consumers from reaching the right answer (e.g., Hastings et al., 2013).

For example, if consumers hold incorrect beliefs, or lack the financial literacy to integrate the opaque price into their final estimate, then their mistakes should be unresponsive to stakes.<sup>8</sup>

The commonly used attention models in which attention is exogenous to stakes but responds to non-pecuniary stimuli, such as those summarized in DellaVigna (2009), are ambiguous about Prediction 2, and are not consistent with the other four predictions. Beyond a qualitative assessment of these simpler attention models, our tests also serve as a critical examination of the extent to which exogenous attention might be a reasonable approximation. For example, even if consumers do exert more mental effort to consider sales taxes for big purchases like cars, the endogeneity of mental effort to taxes may be negligible for most goods that consumers purchase. Thus our measures of the *attention elasticities* are a key quantitative input for enriching the modeling of attention.<sup>9</sup>

Attention models in which consumers either pay full attention to the opaque price or ignore it completely (e.g., Gabaix and Laibson, 2006; Chetty et al., 2007), as well as costly attention models with homogeneous prior perceptions, could not be simultaneously consistent with Prediction 1 and the possibility of overreaction in Prediction 5. In such models, all consumers either systematically underreact or overreact, and all consumers either have a systematic tendency to increase their sensitivity as stakes increase or to decrease their sensitivity as stakes increase. Our tests thus also serve to identify an important, but rarely discussed hypothesis about costly attention processes—that consumers are highly heterogeneous in the rules of thumb they use when they exert little mental effort.

Predictions 2 and 4 are particularly demanding predictions when rules of thumb are heterogeneous, as another plausible hypothesis is that some rules of thumb, like particular forms of rounding heuristics, would not mechanically generate a strong correlation between misreaction at low and high stakes. For example, a consumer might round up a 7% tax to 10%, generating overreaction at low stakes, but round down triple that tax to 20%, generating underreaction at high stakes. Strong heterogeneity in prior perceptions across stakes would in fact predict the opposite of Prediction 4 for the high-stakes regime because of mean reversion. Hypotheses 2 and 4 rely on consumers having more stable prior perceptions across stakes, like leaning toward ignoring the tax, or like treating a 7% tax as 10% and triple that tax as 30%.

Finally, the predictions differentiate costly attention models from two starkly different theories about

---

<sup>8</sup>Eliciting beliefs or financial literacy is a complementary test of this possibility, and one that we conduct. However, these are one-sided tests, as evidence of incorrect beliefs or financial illiteracy among some consumers does not rule out costly attention as an important mechanism. Moreover, because “behavioral” consumers may not necessarily act on the answers they give to abstract beliefs questions (e.g., Bernheim and Taubinsky, 2018), we consider our predictions about behavior a more satisfactory test.

<sup>9</sup>A related class of models in which the salience weight on an attribute depends on choice sets (Bordalo et al., 2013; Koszegi and Szeidl, 2013; Bushong et al., 2021), but not on mental effort, could in principle play some role in our setting as well, although these models do not give special status to the “opaqueness” of an attribute. Differential reaction to  $p_o$  versus  $p_s$  in these models would only result from the fact that these two prices are of different magnitudes. Under the assumption that differences in reaction to  $p_o$  and  $p_s$  depend only on differences in magnitude, these models are for the most part either ambiguous on or inconsistent with our predictions. The homogeneity of degree zero assumption in Bordalo et al. (2013) implies that simply scaling up the importance of the attribute cannot change its salience. The Koszegi and Szeidl (2013) model would predict that all consumers are less sensitive to  $p_o$  than to  $p_s$  when  $p_o$  is of smaller magnitude, and that scaling up  $p_o$  would decrease the relative underreaction to  $p_o$  for all consumers. This is inconsistent with the heterogeneous response to stakes in Prediction 5. Moreover, in the context of sales taxes, the Koszegi and Szeidl (2013) model would predict that changes in relative underreaction depend on whether the amount of tax owed is increased through an increase in sales tax rates or through an increase in posted prices, since the latter also increases the salience of posted prices—this is inconsistent with our findings. The Bushong et al. (2021) model is inconsistent with the predictions and our findings for essentially the same reasons that the Koszegi and Szeidl (2013) model is, since in our setting the model operates just like the Koszegi and Szeidl (2013) model except with the opposite sign.

the interaction between incentives, mental effort, and decision quality. First, Kahneman (2003) and others argue that even if mental effort increases with incentives, this does not have to translate to better decisions.<sup>10</sup> Second, Ariely et al. (2009) argue for and test the hypothesis—building on a line of reasoning dating to Kahneman (1973)—that mental effort is not fully controlled because it is influenced by affective states of arousal, and thus higher incentives may lead mental effort to become misdirected, which would decrease decision quality. For example, a consumer considering a big-ticket purchase may become distracted by the pressure and gravity of the decision, and consequently omit considering sales taxes.

### 3 Experimental design and sample

**Summary:** Each consumer was randomly assigned three of nine household products utilized in the study and made purchase decisions for these three products in three different stores (nine total decision screens). Each store corresponded to a different sales tax rate. In store A, consumers made shopping decisions with a zero sales tax rate (no-tax store). In store B, consumers made shopping decisions with a standard tax rate identical to their city of residence (standard tax environment). In store C, consumers made shopping decisions with a sales tax rate equal to triple their standard tax rate (triple tax environment). The order of these nine sets of shopping decisions was randomized within-subject. Complete details of the experimental protocol are in Appendix R.

#### 3.1 Recruitment

The experiment was conducted in September 2016 through ClearVoice Research, a market research firm that maintains a large and demographically diverse panel of participants over the age of 18. ClearVoice often contracts with industry partners to ship products to consumers to elicit product ratings, but is additionally available to researchers for academic use.<sup>11</sup> Because ClearVoice maintains an infrastructure for easily shipping products to consumers, it is a particularly convenient platform for our incentive-compatible design. Moreover, ClearVoice provides samples that approximate the U.S. population on basic demographic characteristics.

We asked ClearVoice to only recruit panel members from states with 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.<sup>12</sup> 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.<sup>13</sup>

---

<sup>10</sup>See, e.g., Enke and Zimmermann (2019) for evidence in line with this conjecture.

<sup>11</sup>For other economic research using ClearVoice Research, see Benjamin et al. (2014), Taubinsky and Rees-Jones (2018), and Rees-Jones and Taubinsky (2020).

<sup>12</sup>If participants selected Alaska, Montana, Delaware, New Hampshire, or Oregon, the survey ended and participants were told they were ineligible. We drop nine participants who completed the survey and matched to a city with a zero sales tax rate.

<sup>13</sup>Local tax rate data is drawn from the September 2016 update of the “zip2tax” tax calculator.

### 3.2 Shopping decisions and environment

For each purchase decision, consumers first encountered a screen informing them of which store they were entering and for which product they were shopping. Consumers then clicked through to the next screen, which contained a product description and a picture, identical to how the product is presented on Amazon.com. On this same screen consumers also saw a price list containing ten prices. These prices were chosen such that the minimum for all products was \$4.00, and then increased by a multiplicative factor of 15% up to \$14.07.<sup>14</sup> At each price, consumers were asked whether or not they would be willing to purchase the product. It was explained that the price shown excluded any applicable sales taxes.

At any point, participants were able to click the “back” button to see the store in which they were shopping, and an “instructions” button to view the instructions. If a study participant selected yes (or no) for all available prices, he was directed to an additional screen where he was asked to report the highest (lowest) price at which he would be willing to buy the product—the statement on this last screen was not incentivized. Additionally, if a participant’s within-store decisions violated monotonicity, he was notified of that, and given the option to revise.<sup>15</sup>

The three different stores were described to consumers as follows:

When you purchase an item in Store A, you will pay no sales tax in addition to the price. Store A is like one of your local stores, with the taxes already included in the prices that you see on the tags of the items. When you purchase an item in Store B, you will have to pay an additional sales tax, just like you typically do at the register at your local stores (on non-tax-exempt items). The sales tax rate in Store B is the standard sales tax rate that applies in your city of residence, [participant’s city], [participant’s state]. When you purchase an item in Store C, the sales tax that you have to pay in addition to the price is much higher than what you would have to pay at your local stores. The sales tax rate in Store C is triple the standard sales tax rate that applies in your city of residence, [participant’s city], [participant’s state].

The nine household products were selected from the products previously used in Taubinsky and Rees-Jones (2018). None of the items were tax-exempt in any of the 45 states in which our participants reside. Appendix R contains screenshots of the instructions, as well as a list of the nine products, their Amazon.com prices, and their Amazon.com product descriptions.

Decisions in the experiment were incentive-compatible. All study participants who passed the necessary comprehension questions (described below) had a one-third chance of being selected to receive a \$16 budget.<sup>16</sup> Consumers who were selected to receive the \$16 budget had one tax environment and one product

---

<sup>14</sup>For store C only nine prices were included, and the maximum posted price was \$12.24. This was to ensure all consumers would stay within the \$16.00 budget, even after sales taxes were added.

<sup>15</sup>Specifically, the text was: “Your answers on the previous page are inconsistent. If you indicate that you are not willing to buy the product at a lower price, you cannot indicate that you are willing to buy the product at a higher price. For example, it’s inconsistent for a survey-respondent to say that he or she is willing to buy the product for \$5.29 at a particular store, but is not willing to buy the product for \$4.60 at that same store. Click the “back” button to adjust your answers.”

<sup>16</sup>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 \$3.00 and no products from the study.

randomly chosen. Outcomes were determined by randomly selecting one of the prices on the price list.<sup>17</sup> If consumers indicated that they did not want to purchase at that randomly chosen price, then they would keep their \$16 budget and would not receive the product. If consumers indicated they would like to purchase at the randomly generated price, then the product was sold to the consumer at that salient price  $p_s$ —meaning that the consumer kept  $16 - p_s(1 + \tau)$  dollars, where  $\tau$  is the experimentally induced tax rate. The product was then shipped to the person by ClearVoice.

Participants received a full explanation of the payout scheme, including that each question, product, and price was equally likely to be chosen. Additionally, we explicitly informed participants that “it is in your best interest to answer each question honestly.” Appendix R.1 contains a screenshot of the instructions shown to participants.

### 3.3 Ensuring comprehension

To ensure that study participants understood the environment and experimental tax rate, we had them answer six multiple-choice questions after showing them the instructions. Three of these questions concerned the payout, asking participants to identify their shopping budget, how many decisions will be randomly chosen to implement, and the prices at which they would be asked about purchasing the product.<sup>18</sup> The final three questions asked participants to identify the sales tax rate they would face for an item purchased in store A, store B, and store C—with the possible answers being “no sales tax,” “standard sales tax in city of residence,” and “triple the sales tax in city of residence.” If participants answered a question incorrectly or left it blank, they were prompted to select the correct answer before they could begin. When answering these questions, participants could access the instructions which described the tax environments, provided a visual of the price list, and explained the payout structure.

After making the purchase decisions, participants were again asked to state the meanings of the store labels, i.e., to identify the sales tax rate they faced in store A, store B, and store C. Participants were given one attempt to select the correct answer, and were informed that they needed to answer all three correctly to be eligible for the \$16 budget and the consequences of their shopping decisions. Participants were not given access to study instructions in this second round. 86% of participants correctly answered all three questions at the end of the experiment. In our main results we exclude those who fail to correctly state the meaning of the store labels, so as not to confound comprehension of study rules with actual attention costs.<sup>19</sup>

### 3.4 Additional questions

After completing the purchase decisions and additional comprehension checks, participants received a short set of questions eliciting demographic information including household income, marital status, and political beliefs. Appendix R.2.4 contains a list of these questions.

---

<sup>17</sup>For store C, each row in the MPL had a 1/9 chance of being selected which is a slightly higher probability than the 1/10 chance in stores A and B.

<sup>18</sup>Subjects had to identify each answer from a list of three choices. The correct answers are \$16, one purchase decision, and “the prices vary” respectively. Appendix R.2.2 contains a list of these questions.

<sup>19</sup>This exclusion was pre-registered on the AEA RCT registry. AEARCTR-0001597

Participants were also asked to identify the sales tax rate in their city of residence. We additionally asked them to identify how much sales tax they would owe on an \$8.00 item. The first question allows us to test if participants have incorrect beliefs about their sales tax rate, and the second question allows us to test if participants are able to perform the computations necessary to determine the tax on a particular posted price.

### 3.5 Sample

1846 consumers completed the experiment. For our primary analyses, we exclude 256 respondents who incorrectly answered one or more of the comprehension questions and an additional 56 respondents who had monotonicity violations within a price list. Our main results in Section 4 do not require excluding participants with these monotonicity violations, but our analyses in Sections 5 and 6 require monotonic preferences to identify a willingness to pay for each product. In Appendix M we replicate our analysis including those who failed our comprehension checks. We exclude nine additional participants with missing or zero sales tax rates in their city of residence. Our final sample includes 1534 respondents.

Experimental recruitment was targeted to generate a final sample approximating the income, age, and gender distribution of the U.S. adult population. Our sample has a median income of \$49000, an average income of \$60837, and an interquartile range of \$25000-\$80000. Our sample also has a median and mean age of 46 and an interquartile range of 32-59; all participants in the final sample are over the age of 18, and all but 56 participants are over the age of 21. The mean total sales tax rate charged in participants' city of residence is 7.24% (median 7.00%), with a standard deviation of 1.26%. For 90 percent of participants, the sales tax rate lies between 5.50% and 9.50%.

The distribution of these basic demographics is broadly similar to the U.S. population, although selection on other unmeasured characteristics cannot be ruled out.

## 4 The average impact of stakes on inattention

### 4.1 Descriptive summary of behavior

Figure 1 provides a summary of consumer behavior in our study. We begin with panel (a), where we summarize average demand in each tax environment. To construct average demand, we begin with product-specific demand curves  $D_{jk}(p)$  where  $j$  indexes products and  $k$  indexes the store type—A, B, or C. We then construct average demand curves  $D_{avg,k}(p) := \frac{1}{9} \sum_j D_{jk}(p)$ . Panel (a) shows that consumers do react to sales taxes, as their willingness to buy at a given posted price is decreasing in the size of the sales tax.

However, panels (b) and (c) show that consumers on average underreact to taxes. In these panels, we construct the demand curves that would be expected if consumers reacted to the taxes fully. Since we only observe purchase decisions at finitely many prices, we construct the counterfactual demand through linear interpolation, as described in Appendix F. Panel (b) reports the results for the standard tax environment, and panel (c) for the triple tax environment.

Comparing counterfactual to observed demand in the same store, we see evidence of underreaction. The



underreaction is particularly sizeable at low posted prices.<sup>20</sup>

## 4.2 Estimating average revealed valuation weights

Recall that the definition of the revealed valuation weight  $\theta_{ijk}$  for consumer  $i$  considering product  $j$  in store  $k \in \{A, B, C\}$  is that the consumer is  $\theta_{ijk}$  as responsive to a change in the tax as he is to a change in the salient posted price. That is, the consumer behaves as if his perceived price of the product, given a salient posted price  $p$ , is  $p + \theta_{ijk}p\tau_{ik} = p(1 + \theta_{ijk}\tau_{ik})$ . Note that the size of the opaque price  $p_o$  is given by  $p_o = p \cdot \tau_{ik}$  here. The consumer thus chooses to buy if his product valuation  $v_{ijk} := v_{ij} + \varepsilon_{ijk}$  is such that  $\ln v_{ijk} \geq \ln p + \ln(1 + \theta_{ijk}\tau_{ik})$ . The  $\varepsilon_{ijk}$  term is an idiosyncratic shock that can vary across stores (holding the item constant), and captures the potential noise/measurement error in person  $i$ 's evaluation of product  $j$  in store  $k$ , or people changing their mind about what product  $j$  is worth to them as they move from store A to store B or store C. To ease empirical estimation, we simplify this condition to be linear in logs by noting that  $\ln(1 + \theta_{ijk}\tau_{ik}) \approx \theta_{ijk} \ln(1 + \tau_{ik})$  up to negligible higher order terms. Under this approximation, the consumer buys if

$$\ln v_{ijk} \geq \ln p + \theta_{ijk} \ln(1 + \tau_{ik}) \quad (3)$$

We then utilize condition (3) to estimate the average revealed valuation weights by estimating the following heteroskedastic probit model:

$$1 - Pr(buy_{ijk}|p) = \Phi \left( \frac{\alpha_j + \ln(p) + \bar{\theta}_B \ln(1 + \tau_{ik}) \cdot I(k = B) + \bar{\theta}_C \ln(1 + \tau_{ik}) \cdot I(k = C)}{\sigma_j} \right) \quad (4)$$

where  $\Phi$  is the standard normal CDF. By allowing both  $\alpha_j$  and  $\sigma_j$  to vary by product, we allow the demand curves for the different products to differ both in the price sensitivity and in the aggregate valuation for the products.

Because we estimate a nonlinear probability model, the estimated coefficients  $\bar{\theta}_B$  and  $\bar{\theta}_C$  approximate the respective means  $E[\theta_{ijk}|k = B]$  and  $E[\theta_{ijk}|k = C]$  with some error when the distribution of  $\theta$  is heterogeneous within each store. In Appendix G we verify that this approximation error is negligible, and that it works ‘‘against’’ the results that follow about how the revealed valuation weights are increasing in the posted price and the tax rate.

## 4.3 Average revealed valuation weights increase as stakes increase

In our experiment, we observe consumer choice both across different salient posted prices  $p$  (within-store) and across different tax rates  $\tau$  (across stores). Both lead to an increase in the size of the tax, which is given by  $p_o = p\tau$ . In the language of our theoretical framework, we consider both increases in salient posted prices and increases in tax rates to be salient changes in stakes  $\sigma$ , and our models predict that the revealed valuation weights should increase in  $\sigma$ .

Figure 2a plots  $E[\theta|p \leq p^\dagger]$  against a price cutoff  $p^\dagger$ , such that all prices less than or equal to the cutoff

<sup>20</sup>10.10 percent of participants never choose to buy a product in store C, compared to 4.89 percent and 5.87 percent of participants in stores A and B, respectively.

value are included in calculating that average valuation weight. We estimate  $E[\theta]$  at different posted prices using the empirical model in equation (4), dropping observations with  $p$  above the cutoff. The leftmost point of each series includes just the posted prices less than or equal to \$4.60; i.e., \$4.00 or \$4.60. The rightmost point on each series corresponds to including all the posted prices. The point estimates and confidence intervals corresponding to Figure 2a are reported in Appendix H.

The figure establishes three important facts. First, on average consumers underreact to the size of the tax, both at standard-sized taxes and at tripled taxes. When pooling over all of the prices, the average valuation weight  $\theta$  in the standard tax store is 0.48 (95% CI [0.32, 0.63]), and the average valuation weight  $\theta$  in the triple tax store is 0.79 (95% CI [0.72, 0.86]).

Second, the average valuation weight is increasing in the tax rate  $\tau$ . As is immediately evident from figure 2a, the average valuation weight is significantly higher in the triple tax condition than in the standard tax at each cutoff. At the \$4.60 cutoff, the difference in  $E[\theta]$  between the triple tax and single tax environments is 0.18 (95% CI [0.09, 0.27]). This difference peaks at 0.38 (95% CI [0.26, 0.51]) at a price cutoff of \$8.05. When pooling over all prices, this difference is 0.31 (95% CI [0.20, 0.42]). These results are consistent with Prediction 1.

Third, the average valuation weights are increasing in the salient posted price  $p$ . In the standard tax environment,  $E[\theta]$  more than doubles as we move from a cutoff of \$4.60 to pooling all prices: it increases from 0.23 to 0.48 (95% CI for difference [0.14, 0.37]). Similarly, in the triple tax environment,  $E[\theta]$  approximately doubles as well, increasing from 0.40 to 0.79 (95% CI for difference [0.32, 0.44]) when moving from a price cutoff of \$4.60 to pooling over all prices. These results provide further evidence consistent with Prediction 1.

One potential concern in examining how the valuation weights vary by price is that the set of consumers on the margin at each price are mechanically different: the higher is the price, the higher is the product valuation of these marginal consumers. If valuation for the product is correlated with attention, this would confound our results about how average valuation weights covary with price. Although there is no clear reason for this to be the case, in principle this could occur. In Appendix I, we provide evidence against this concern, showing that consumers who are identified as having higher valuations for products in our experiments are not more attentive to taxes. Of course, the tax rate assignment is exogenous to these differences and is not subject to the same concern.

On the other hand, a concern with examining how valuation weights change in response to an increase in tax rates is that consumers see the triple tax store as a highly unusual environment, which affects their purchase decision beyond the pecuniary channel. Consumers might be significantly more responsive to higher tax rates simply because the increase triggers tax aversion, because the surprising and unusual environment simply draws more attention to itself (Bordalo et al., 2020), or because of experimenter demand effects in a within-subject design.

Although our results on individual differences in reaction to the tripling of the tax rate generate a number of additional tests of costly attention that are not easily explained by the confounding forces described above, we present one more test of aggregate behavior here. We address both concerns by examining how average valuation weights depend on the total size of the tax, and whether it seems to matter whether increases in the

tax come from increases in prices or increases in taxes. If our results are consistent across these two different ways of increasing tax owed, then that lends more credibility to our hypothesis of consumers exerting more mental effort in response to higher stakes. For example, if behavior is particularly responsive to a tripling of the tax rate because of experimenter demand effects, then we would expect to see that a tripling of the tax rate generates much larger changes in behavior than does an increase in price that leads to the same absolute change in the sales tax.

To do this, we divide the different prices in each store into five disjoint sets, for a total of  $5 \times 2 = 10$  sets.<sup>21</sup> For each pair, we estimate the average valuation weight using an extension of model (4) with a separate  $\bar{\theta}$  parameter for each pair. Specifically, we partition the prices into sets  $P_n$ ,  $n = 1, \dots, 5$  and estimate

$$1 - Pr(buy_{ijk}|p) = \Phi \left( \frac{\alpha_j + \sum_{n=1}^5 [\ln(p) + \bar{\theta}_{B,n} \ln(1 + \tau_{ik}) \cdot I(k=B) + \bar{\theta}_{C,n} \ln(1 + \tau_{ik}) \cdot I(k=C)] I(p \in P_n)}{\sigma_j} \right) \quad (5)$$

with  $\sigma_1$  normalized to 1. We plot the average  $\bar{\theta}_{k,n}$  against the average tax owed in the corresponding set:  $E[\bar{p}_n \tau_{ik}|k]$ , where  $\bar{p}$  is the average price in set  $n$ .

Figure 2b presents the results.<sup>22</sup> We see no trend break between the two series.<sup>23</sup> If anything, the deviation in the leftmost point in the store C series has the opposite sign predicted by experimenter demand effects generating an exaggerated response to a tripling of the taxes. Figure 2b thus provides additional evidence for a costly attention model in which attention is a function of the salient stakes, regardless of where those increases come from.

#### 4.4 Robustness and correlates of misreaction

In Appendix M, we replicate our analyses including the 14% of our respondents who were not able to correctly answer the comprehension questions about the tax rates charged in stores A, B and C. Re-inclusion of these participants increases the estimate of average underreaction, but does not change any of the comparative statics. Second, in Appendix O, we analyze whether purchase decisions could be influenced by the order in which the nine purchase decisions are presented to consumers, a potential concern with our within-subject experimental design. We find no evidence of order effects.

In Appendix L, we analyze two potential mechanisms for misreaction: (i) inaccurate beliefs about local tax rates, and (ii) inability of participants to compute the sales tax they would need to pay for an item. Utilizing survey questions eliciting participants' knowledge of their local tax rate and their computational ability, we replicate our main results restricting to participants with nearly-accurate beliefs and high computational ability. In both cases, we find strong evidence for Prediction 1, and the estimates are also of very similar magnitude to the full sample results. The results suggest inaccurate beliefs and poor computational ability

<sup>21</sup>Since the MPLs for store C only contained nine prices, the final set contains only the highest price.

<sup>22</sup>Confidence intervals are substantially smaller for the triple-tax condition because in the regression the triple tax constitutes a right-hand-side variable with substantially larger variation.

<sup>23</sup>In the spirit of equation (26) of Gabaix (2019) (the quadratic cost attention function), we estimate a quadratic fit through the point estimates in Figure 2b. We find  $E[\theta]$  is best approximated as  $E[\theta|t] = \min(0.87t - 0.21t^2, 1)$ , where  $t$  is the average tax owed. We find no evidence that the tax rate  $\tau$  affects  $E[\theta]$ , other than through  $t$ .

cannot explain the observed misreaction.

Appendix J shows that we do not find variation in the average underreaction by income, education, or political party affiliation, although these correlations are not well-powered. In that appendix, we also analyze how average valuation weights vary by local tax rates, and find some evidence that participants in high sales tax locations have lower revealed valuation weights than those from low sales tax location, although this is likely a confounded test of costly attention models because local tax variation could be related to a number of differences in geography, including consumers' views and preferences about tax rates.<sup>24</sup>

## 5 Reduced-form results on individual differences in attentional responses to stakes

In Section 4 we showed that the average revealed valuation weights in the population are increasing in stakes, supporting Prediction 1 of costly attention models. In this section, we begin to examine predictions about individual differences using simple reduced-form tests. Our approach is to create individual-level proxies for consumers' revealed valuation weights  $\theta$ , use these proxies to divide consumers into high and low valuation weight groups, and then use these groups to test comparative static predictions about individual differences in  $\theta$ . We use this methodology to provide evidence for Predictions 2-4.

### 5.1 Testing Predictions 2 and 3

#### 5.1.1 Approach

The main idea of our reduced-form tests is to identify consumers who seem to be more sensitive to taxes on one product, and to examine their sensitivity to standard and triple taxes on *the other two* products. To construct proxies for sensitivity to taxes, we first construct estimates  $\hat{\theta}_{ijk}$  of  $\theta_{ijk}$  for each consumer  $i$ , product  $j$ , and store  $k \in \{B, C\}$ . To do so, we first approximate the maximum pre-tax price  $p_{ijk}^*$  at which consumer  $i$  is willing to buy product  $j$  in store  $k \in \{A, B, C\}$  by  $\ln p_{ijk}^* = 0.5 (\ln p_{ijk}^0 + \ln p_{ijk}^1)$ , where  $p_{ijk}^0$  is the highest price at which consumer  $i$  buys product  $j$  in store  $k$  and  $p_{ijk}^1$  is the lowest price at which consumer  $i$  declines to purchase product  $j$  in store  $k$ . For consumers who were willing to buy at all prices (or no prices), we use their non-incentivized answers about the maximum price at which they would be willing to buy the given product, and to reduce the impact of outliers, we set  $p_{ijk}^*$  to the the median of the self-reports for product  $j$  in store  $k$ .<sup>25</sup> Using the buying condition in equation (3), we construct the estimates  $\hat{\theta}_{ijk}$  for  $k \in \{B, C\}$  as

$$\hat{\theta}_{ijk} = \frac{\ln(p_{ijA}^*) - \ln(p_{ijk}^*)}{\ln(1 + \tau_{ik})} \quad (6)$$

However, we cannot directly use the  $\hat{\theta}_{ijk}$  estimates to compute properties of the actual distribution of  $\theta$  without making the unrealistically strong assumption that all within-person differences in choices between

<sup>24</sup>For example, higher-tax rate jurisdictions tend to be more urban and in more liberal states, and the observed differences in average valuation weights may reflect sorting into urban versus rural jurisdictions.

<sup>25</sup>In Appendix N, we replicate the results adding the condition that we exclude participants who always buy or never buy a product in any store.

stores load on the  $\theta_{ijk}$  parameter. A mechanical reason these assumptions are too strong is that the set of prices in the experiment is finite and thus valuation weights at the individual level are not point-identified.<sup>26</sup> A perhaps more important reason is that changes in consumers' willingness to buy at certain prices may not only reflect their responses to the tax regime, but also changing perceptions of the product value or simply "experimental noise" such as consumers accidentally clicking on the wrong response. Consequently, differences in individual-level *estimates* do not imply actual individual differences; i.e., all of these considerations would generate differences in individual-level estimates even if consumers were perfectly homogeneous in their priors and attention strategies.<sup>27</sup>

We instead use the  $\hat{\theta}_{ijk}$  estimates to create proxies for high versus low valuation weight consumers. We use *one* product to divide consumers into two groups. The *low* group consists of those with low values of  $\hat{\theta}_{ijk}$  and the *high* group consists of those with high values of  $\hat{\theta}_{ijk}$ . We then estimate our empirical model in (4) *on the other two products* to estimate average valuation weights for the low and high groups.

Concretely, the procedure is as follows. We index each of the three products for each person by  $j \in \{1, 2, 3\}$ . First, we start with  $j = 1$  and we split the sample into two groups: those with  $\hat{\theta}_{i1B}$  in the top 25% of the population and those with  $\hat{\theta}_{i1B}$  in the bottom 75% of the population.<sup>28</sup> We define  $x_{i1}^{75} = I[F(\hat{\theta}_{i1B}) > 0.75]$  to be an indicator for the high group. We then use decisions regarding the *other two* products to estimate the average valuation weights  $E[\theta_{ijk}|k = K, x_{i1}^{75}, j \neq 1]$  using equation (4), where  $K \in \{B, C\}$ . We repeat the procedure twice using products 2 and 3 to generate  $x_{i2}^{75}$  and  $x_{i3}^{75}$ , and estimate  $E[\theta_{ijk}|k = K, x_{i2}^{75}, j \neq 2]$  and  $E[\theta_{ijk}|k = K, x_{i3}^{75}, j \neq 3]$ . Finally, we average the estimates from each of these three iterations to get an overall average estimate of  $\theta_{ijk}$  for those in the high and low groups:

$$E[\theta_{ijk}|k = K, x_i^{75}] = \frac{1}{3}(E[\theta_{ijk}|k = K, x_{i1}^{75}, j \neq 1] + E[\theta_{ijk}|k = K, x_{i2}^{75}, j \neq 2] + E[\theta_{ijk}|k = K, x_{i3}^{75}, j \neq 3])$$

We compute confidence intervals using percentile bootstrap, clustering by subject. In Appendix K we summarize results that utilize alternative criteria for splitting the sample into high and low groups.

The key statistical assumption that ensures consistency of our estimates is that the errors in individual-level estimates  $\hat{\theta}_{ijk}$  and  $\hat{\theta}_{ij'k'}$  are orthogonal conditional on the true underlying value:

**Assumption 1.** For any  $j$ ,  $(\hat{\theta}_{ijk} \perp \hat{\theta}_{ij'k'}) | \theta_{ijk}$  when  $j' \neq j$ , for  $k, k' \in \{B, C\}$

This assumption is weaker than the assumption that the measurement errors are mean zero (strongly

<sup>26</sup>Using a Becker-DeGroot-Marshak (BDM) mechanism would not resolve this problem. Taubinsky and Rees-Jones (2018) find that almost half of participants round their maximum willingness to pay to round numbers. This rounding implies that BDM data is also coarse in the same fashion that a discrete list of prices is coarse.

<sup>27</sup>As a concrete example of patterns of behavior that are likely "measurement error" in  $\hat{\theta}_{ijk}$ , 33.8% of consumers are willing to buy at a higher pre-tax price for at least one product in the standard tax environment than in the no tax environment, 20.4% are willing to buy at a higher pre-tax price for at least one product in the triple tax environment than in the no tax environment, and 25.9% are willing to buy at a higher pre-tax price for at least one product in the triple tax environment than in the single tax environment. Attributing such patterns of behavior to consumers' valuation weights  $\theta$  would imply substantially negative  $\theta$  for some consumers—i.e., that some consumers perceive the taxes to be subsidies. Instead, these patterns likely reflect other factors like changing perceptions of product value or "noise." Our finding of likely "measurement error" in individual-level point estimates is not unusual. As summarized by Gillen et al. (2019), it is prevalent in most experimental analyses of individual differences.

<sup>28</sup>We chose to split the sample at 25% and 75% in part because much of the mass for each the  $\hat{\theta}_{ijk}$  is at 0. Splitting at the 75th percentile ensures that all participants with  $\hat{\theta}_{ijk} = 0$  are classified into the low valuation weight group. Importantly, the procedure also works with other splits. In Appendix K, we replicate the results splitting at the median, 80th percentile, and 85th percentile.

classical measurement error), or that they are orthogonal to the true underlying  $\theta_{ijk}$  (weakly classical measurement error.) These stronger assumptions are hard to justify when the underlying model of choice is a nonlinear probability model and the price observations are interval-valued, and when some of the identification comes from unincentivized self-reports. Moreover, stronger assumptions about the nature of measurement error are not required for our procedure. In fact, the procedure does not even require that consumers' unincentivized reports about maximum buying prices approximate the truth in a meaningful way, since we use the self-reported data only to construct proxies for sensitivity to tax rates, but perform our actual estimation of  $E[\theta_{ijk}|k = K]$  using only incentivized decisions.<sup>29</sup>

## 5.1.2 Results

Table 1 presents results from the split-sample techniques described in Section 5.1. Rows (1) and (2) present estimates of  $E[\theta_{ijk}|k = K]$  for the high and low valuation weight groups, respectively; row (3) presents estimates of the difference. For comparison, row (4) presents estimates of average  $\theta$  for the full sample.

Consumers in the high valuation weight group have an average revealed valuation weight of 1.04 (95% CI [0.83, 1.24]) for standard taxes, while consumers in the low valuation weight group have an average revealed valuation weight of 0.25 (95% CI [0.08, 0.42]) for standard taxes. This implies strong individual differences in the revealed valuation weights  $\theta$ . Moreover, consistent with Prediction 2, these individual differences are large and persistent across stakes. In the triple tax store, consumers classified as having high valuation weights in the standard tax store have an average valuation weight of 1.20 (95% CI [1.10, 1.31]), while consumers classified as having low valuation weights in the standard tax store have an average valuation weight of 0.64 (95% CI [0.57, 0.72]).

Consistent with Prediction 3, the low valuation weight group exhibits a significantly larger increase in their valuation weights than the high valuation weight group when tax rates are tripled (0.16 vs. 0.39; 95% CI for difference  $[-0.43, -0.04]$ ). The high valuation group does adjust moderately upward by 0.39 (95% CI [0.26, 0.51]), but this is not inconsistent with theory when individual differences are taken into account.<sup>30</sup>

## 5.2 Testing Prediction 4

### 5.2.1 Approach

We also use the strategy in Section 5.1 to analyze heterogeneity in how the valuation weights respond to stakes. This provides a test of Prediction 4. Specifically, we use one product to divide consumers according to how much they adjust their valuation weight when the tax rate increases, and we then examine how they respond to taxes on the other two products.

<sup>29</sup>There is an argument to be made that “measurement error” is also a question of attention—such that the “random mistakes” generating what we refer to as “noise” are themselves potentially correlated with individuals’ cost of attention more generally. Our assumption is not inconsistent with this possibility, as it does not preclude the important possibility that the variance of the noise term  $\varepsilon_{ijk} = \hat{\theta}_{ijk} - \theta_{ijk}$  is related to  $\theta_{ijk}$ .

<sup>30</sup>As we will show, some individuals overreact significantly, and thus the high valuation group likely consists of individuals both with  $\theta < 1$  and with  $\theta > 1$ . Thus, because underreacting individuals may adjust upward by slightly more than the overreacting individuals adjust downward, the *average* response to stakes in this group is not guaranteed to be null (or negative).

Formally, we define  $\Delta_{ij} = \theta_{ijC} - \theta_{ijB}$  as the degree of adjustment in the revealed valuation weight when moving from the standard tax environment to the triple tax environment. We then classify consumers into high and low adjustment groups using  $\hat{\Delta}_{ij} := \hat{\theta}_{ijC} - \hat{\theta}_{ijB}$ . We define  $d_{ij}^{25} = I[F(\hat{\Delta}_{ij}) \leq 0.25]$  to be an indicator of being in a low adjustment group, where  $d_{ij}^{25} = 0$  indicates low adjustment and  $d_{ij}^{25} = 1$  indicates higher adjustment. We then estimate  $E[\Delta_{ij}|d_{ij}^{25}, j \neq j']$  for each  $j' \in \{1, 2, 3\}$ . We then average to estimate  $E[\Delta_{ij}|d_{ij}^{25}]$  separately for the high and low adjustment groups.

## 5.2.2 Results

Table 2 reports the results. We find that there are significant individual differences: consumers in the low adjustment group increase their valuation weights by an average of 0.01 (95% CI  $[-0.15, 0.17]$ ), and consumers in the high adjustment group increase their valuation weights by an average of 0.43 (95% CI  $[0.30, 0.55]$ ). The results imply substantial underlying heterogeneity in  $\Delta_{ij}$ .

Consistent with Prediction 4, we find that consumers in the low adjustment group have higher valuation weights in both the standard tax regime (0.85 vs. 0.34; 95% CI for difference  $[0.30, 0.75]$ ) and in the triple tax regime (0.86 vs. 0.76; 95% CI for difference  $[-0.01, 0.20]$ ). The result for the triple tax regime is significant at the 10% level ( $p$ -value = 0.067).

As we discussed in Section 2.2, Prediction 4 is a particularly demanding test. If, for example, the distribution of  $\theta_{ijB}$  and  $\theta_{ijC}$  took the form  $\theta_{ijC} = a_0 + \theta_{ijB} + \varepsilon_{ij}$  for some constant  $a_0$  and some random variable  $\varepsilon_{ij}$  independent of  $\theta_{ijB}$ , then  $E[\theta_{ijC}|\theta_{ijC} - \theta_{ijB} = \Delta]$  would be increasing in  $\Delta$ , not decreasing. Intuitively, small values of  $\Delta$  would imply a small idiosyncratic component, and thus a smaller value of  $\theta_{ijC}$ . In general, any stochasticity in  $\theta_{ijC} - \theta_{ijB}$  that is independent of the value of  $\theta_{ijB}$  would push against our empirical result. Our result is thus consistent with the special structure that costly attention models impose on revealed valuation weights.

## 5.3 Robustness

In Appendix L, we replicate Tables 1 and 2 on the subsample of participants with nearly accurate beliefs about their sales tax rate and with strong computational ability. In Appendix M we confirm that the results hold for the full sample of participants, including those failing comprehension checks. In Appendix N, we replicate results adding the condition that we exclude participants who always buy or never buy a product in any store. In all three cases, the results conform with Predictions 2-4. In Appendix K, we verify our results still support these theoretical predictions when using alternative percentile cutoffs to split the sample.

## 6 Overreaction and heterogeneous attentional responses to stakes

While the evidence in Section 5 is consistent with at least moderate individual differences, it leaves open three key questions. First, are the individual differences large enough that some consumers overreact to standard taxes? Second, if we detect overreaction, can we show that some consumers decrease their valuation weights when the stakes are increased? Third, how big is the variance of the valuation weights, which

Taubinsky and Rees-Jones (2018) and Farhi and Gabaix (2020) show is a key input into efficiency cost calculations? In this section, we develop econometric techniques for computing lower bounds on individual heterogeneity, which enable us to answer the three questions above.

A key moment that we have not exploited in the analysis in the previous section is how well correlated the binary proxies are with each other. In what follows, we will show how this moment, combined with the results in Section 5, helps generate a lower bound on the variance of  $\theta$ . Intuitively, if our classification of consumers into high versus low  $\theta$  groups is very imprecise, then the fact that average  $\theta$  is still so different for consumers in the two groups we create must imply that the individual differences in  $\theta$  are so high that even conditioning on poor proxies for  $\theta$  yields large differences in conditional means. The lower bound on the variance of  $\theta$  then provides a lower bound on the extent to which some individuals must overreact, in a manner that we describe in more detail in Section 6.1 below.

## 6.1 Methods for quantifying individual differences

We begin with a general result, and then adapt it to our setting.

**Proposition 1.** *Let  $Y$  have support  $[\underline{Y}, \bar{Y}]$ , and let  $X_1$  and  $X_2$  be binary variables that are independently and identically distributed conditional on each realization of  $Y$ . Then*

$$\text{Var}[Y] \geq \frac{\text{Cov}[Y, X_1] \cdot \text{Cov}[Y, X_2]}{\text{Cov}[X_1, X_2]} \quad (7)$$

and

$$(\bar{Y} - E[Y])(E[Y] - \underline{Y}) \geq \text{Var}[Y]. \quad (8)$$

*Both bounds are tight, and are obtained when  $Y$  is Bernoulli.*

The result in (7) formalizes the intuition above: the less well-correlated the proxies  $X_i$  are with each other, the higher must be variance, given an estimate of the covariance between  $Y$  and  $X_i$ . We prove this result through an application of the Cauchy-Schwarz inequality.

The result in (8) is the Bhatia and Davis (2000) inequality. The intuition is that the variance of a random variable  $Y$  with a given mean  $E[Y]$  and bounded support cannot be higher than the variance of a Bernoulli random variable with mean  $E[Y]$  and all mass on the two endpoints of the support.

Proposition 1 enables us to use the types of binary proxies utilized in Section 5 to compute bounds on the variance and support of  $\theta_{ij}$ . Define  $x_{ijk}^q = I[F(\hat{\theta}_{ij} > 0.01q)]$  as an indicator for  $\hat{\theta}_{ijk}$  being in the  $q$ th percentile or higher in store  $k$ . This is analogous to Section 5, where we set  $q = 75$  for store  $k = B$ .

**Corollary 1.** *Assume the distribution of  $\theta_{ijk}$  is supported on  $[0, \bar{\theta}]$ , where  $\theta_{ijk}$  is the revealed valuation weight for product  $j$  of individual  $i$  in store  $k$ . Then for  $j \neq j' \neq j''$ : the variance of  $\theta_{ijk}$  in store  $K$  is*

$$\text{Var}[\theta_{ijk}|k=K] \geq \frac{\text{Cov}[\theta_{ijk}, x_{ij'k}^q|k=K] \cdot \text{Cov}[\theta_{ijk}, x_{ij''k}^q|k=K]}{\text{Cov}[x_{ij'k}^q, x_{ij''k}^q|k=K]} \quad (9)$$



and

$$\bar{\theta} \geq E[\theta_{ijk}|k=K] + \frac{\text{Cov}[\theta_{ijk}, x_{ij'k}^q|k=K] \cdot \text{Cov}[\theta_{ijk}, x_{ij''k}^q|k=K]}{E[\theta_{ij'k}|k=K] \cdot \text{Cov}[x_{ij'k}^q, x_{ij''k}^q|k=K]} \quad (10)$$

To derive the corollary, we set  $x_{ij'k}^q$  and  $x_{ij''k}^q$  to correspond to  $X_1$  and  $X_2$  and set  $\theta_{ijk}$  to correspond to  $Y$  in Proposition 1. The assumptions of Proposition 1 are satisfied because  $x_{ij'k}^q$  and  $x_{ij''k}^q$  are identically distributed by assumption, and because Assumption 1 in Section 5.1.1 implies that  $x_{ij'k}^q$  and  $x_{ij''k}^q$  are independently distributed conditional  $\theta_{ijk}$ . In particular, note that by working with  $j \neq j' \neq j''$ , we avoid any biases that result from conflating true individual differences with measurement error.

Analogously, we use Proposition 1 to derive bounds on adjustment  $\Delta_{ij} = \theta_{ijC} - \theta_{ijB}$ . We define  $d_{ij}^q = I[F(\hat{\Delta}_{ij} < 0.01q)]$  as a binary indicator for  $\hat{\Delta}_{ij}$  being in  $q$ th decile or lower. This is analogous to Section 5, where we used  $q = 25$ .

**Corollary 2.** *Assume the distribution of  $\Delta_{ij}$  is supported on  $[\underline{\Delta}, 1]$ , where  $\Delta_{ij} = \theta_{ijC} - \theta_{ijB}$ . Then given instruments  $d_{ij'}^q$  and  $d_{ij''}^q$  computed for products  $j'$  and  $j''$  (with no two of  $j, j'$  and  $j''$  equal):*

$$\text{Var}[\Delta_{ij}] \geq \frac{\text{Cov}[\Delta_{ij}, d_{ij'}^q] \cdot \text{Cov}[\Delta_{ij}, d_{ij''}^q]}{\text{Cov}[d_{ij'}^q, d_{ij''}^q]} \quad (11)$$

and

$$\underline{\Delta} \leq E[\Delta_{ij}] + \frac{\text{Cov}[\Delta_{ij}, d_{ij'}^q] \cdot \text{Cov}[\Delta_{ij}, d_{ij''}^q]}{(E[\Delta_{ij}] - 1) \cdot \text{Cov}[d_{ij'}^q, d_{ij''}^q]} \quad (12)$$

The assumption that  $\Delta_i \leq 1$  is equivalent to assuming that when stakes increase, no consumers switch from being systematic underreactors to systematic overreactors (or that no overreacting consumers substantially *increase* their overreaction). This is consistent with the core of any costly attention model that could micro-found consumer behavior in our experiment.

While these results generate bounds on the supremum of the support, they do not quantify how many consumers overreact. To do so, we derive a bound for the fraction of overreactors,  $Pr(\theta_{ij} > 1)$ , and for the fraction of consumers who adjust their valuation weight downwards,  $Pr(\Delta_{ij} < 0)$ . These results follow from a more general result proven in Appendix C.2, which can be seen as a converse of sorts to Chebyshev's inequality.

**Proposition 2.** *Assume  $\theta_i$  has support  $[0, \bar{\theta}]$ , where  $\bar{\theta} > 1$  is the supremum of the support and can vary by store. Additionally, assume that the distribution of  $\Delta_i$  is supported on  $[\underline{\Delta}, 1]$ . Then*

$$Pr(\theta_i > 1) \geq \frac{\text{Var}[\theta_i] + E[\theta_i]^2 - E[\theta_i]}{(\bar{\theta} - 1)\bar{\theta}} \quad (13)$$

and

$$Pr(\Delta_i < 0) \geq \frac{\text{Var}[\Delta_i] + E[\Delta_i]^2 - E[\Delta_i]}{(\underline{\Delta})(\underline{\Delta} - 1)} \quad (14)$$

Both bounds are tight.

The intuition for this result is that the distribution that minimizes  $Pr(\theta_i > 1)$  subject to a variance constraint and supremum constraint is one that puts all mass on  $\theta_i \in \{0, 1, \bar{\theta}\}$ .

## 6.2 Bounds on the variance and the support

**Estimation:** In our empirical implementation of the bounds, we construct  $x_{ijk}^q$  using different values of  $q$ . This allows us to construct multiple estimates of each bound, and since the true value must be higher than all these bounds, we take the maximum over them. Formally, an immediate extension of (9) is that

$$Var[\theta_{ijk}|k = K] \geq \max_q \left\{ \frac{Cov[\theta_{ijk}, x_{ij'k}^q|k = K] \cdot Cov[\theta_{ijk}, x_{ij''k}^q|k = K]}{Cov[x_{ij'k}^q, x_{ij''k}^q|k = K]} \right\}. \quad (15)$$

In the empirical implementation, we take the maximum over  $q \in \{10, 15, \dots, 90\}$ .<sup>31</sup> We generate the bounds for  $Var[\Delta_{ij}]$  and  $\underline{\Delta}$  analogously. We calculate bootstrapped percentile-based confidence intervals from 1000 replications, clustered at the subject level, and report the 5% confidence bound.<sup>32</sup>

**Results:** Table 3 presents our estimates. We estimate a lower bound of 0.83 (5% confidence bound of 0.52) for  $Var[\theta_{iB}]$  and of 0.71 (5% confidence bound of 0.59) for  $Var[\theta_{iC}]$ . We estimate a lower bound for  $Var[\Delta_{ij}]$  of 0.86 (5% confidence bound of 0.31). These results provide evidence for significant dispersion in revealed valuation weights in both tax environments, as well as for adjustment when switching tax environments.

Table 3 also presents the supremum lower bound estimates. We estimate a lower bound of  $\bar{\theta}_B$  to be 2.21 (5% confidence bound of 1.55) and for  $\bar{\theta}_C$  to be 1.69 (5% confidence bound of 1.54). Both of these bounds are significantly above 1, indicating there are overreactors in the experimental population. Row (2) presents the estimate as a lower bound on  $-\underline{\Delta}$ . We estimate an upper bound on  $\underline{\Delta}$  to be 0.94 (5% confidence bound of 0.16). This result is consistent with Prediction 5 of our theoretical results.<sup>33</sup>

## 6.3 Bounds on propensity to overreact to taxes

By substituting the lower bound for  $Var[\theta_{ijk}|k = K]$  from equation (15) into the lower bound from equation (13), we derive a bound for the fraction of overreactors as a function of  $\bar{\theta}$ .

Figure 3a plots  $Pr(\theta_{ijk} > 1|k = K)$  as a function of  $\bar{\theta}$ , both for  $K = B$  and for  $K = C$ , along with 5% confidence bounds computed by bootstrap clustered at the subject level. When  $\bar{\theta} = 2.25$ , we estimate that

<sup>31</sup>To estimate  $Cov[\theta_{ijk}, x_{ij'k}^q|k = K]$ , for each  $q$ , we use the expansion

$$Cov[\theta_{ijk}, x_{ij'k}^q|k = K] = E[x_{ij'k}^q|k = K] \cdot (1 - E[x_{ij'k}^q|k = K]) \cdot (E[\theta_{ijk}|x_{ij'k}^q = 1, |k = K] - E[\theta_{ijk}|x_{ij'k}^q = 0, |k = K])$$

with the conditional means  $E[\theta_{ij}|x_{ij'}^q = 0, |k = K]$  estimated as in Section 5.

<sup>32</sup>Note that since we are calculating lower bounds there is no concept of a 95% confidence bound. In other words, it does not make sense to provide a statistical upper bound on a lower bound.

<sup>33</sup>To our knowledge, our finding of overreaction is new to the literature. While Taubinsky and Rees-Jones (2018) use observable covariates to measure individual differences, their methods are able to achieve a lower bound on the variance of approximately only 0.1, and therefore cannot be used to establish overreaction.

at least 20.5% (5% confidence bound of 9.5%) of the population is overreacting in store B, and at least 19.3% (5% confidence bound of 14.5%) is overreacting in store C. Both bounds are decreasing in  $\bar{\theta}$ : for  $\bar{\theta} = 4.25$ , we bound the fraction of overreactors at 4.2% (5% confidence bound of 2.0%) of the population is overreacting in store B, and at least 3.9% (5% confidence bound of 3.0%) in the triple tax environment.

Using equation (14), we derive an analogous bound for the fraction of consumers who adjust their valuation weight downward in response to higher stakes. Figure 3b plots  $Pr(\Delta_{ij} < 0)$  as a function of  $\underline{\Delta}$ . For  $\underline{\Delta} = -0.94$ , the bound computed in table 3, we estimate that at least 35.5% (5% confidence bound of 6.1%) of consumers negatively adjust their revealed valuation weights when switching from the standard tax regime to the triple tax regime. This lower bound is decreasing in the magnitude of  $\underline{\Delta}$ , and at  $\underline{\Delta} = -4.25$  we estimate a lower bound on the fraction of participants with negative adjustment to be 4.1% (5% confidence bound of 0.7%).

**Robustness** In Appendix L, we verify that the results hold on the subsample of participants with nearly accurate beliefs about their sales tax rate and with strong computational ability. In Appendix M we confirm that the results hold for the full sample of participants, including those failing comprehension checks.

## 7 Concluding remarks

In this paper, we provide tests of costly attention models in a concrete and policy-relevant setting. A better understanding of the mechanisms can better inform both positive and normative analysis.

Evidence of costly attention implies that shrouding taxes can generate deadweight loss by imposing cognitive costs on consumers. Evidence of significant heterogeneity in attention, generating both under- and overreaction to opaque prices, implies that there may be significant deadweight loss from misallocation of products to consumers, which we show in Appendix Q by utilizing the deadweight loss formula from Taubinsky and Rees-Jones (2018).

Both the heterogeneity and the elasticity with respect to stakes can also have important implications for how firms design “shrouded prices” in their contracts. For example, Gabaix and Laibson (2006), Heidhues et al. (2017), and others derive a number of interesting implications about market structure under the assumption that consumers either perceive shrouded fees correctly or ignore them completely. Our results on sales taxes suggest that consumer attention to shrouded prices might be significantly more nuanced than what is assumed in these models. Working out the behavioral IO implications of the richer models of inattention that our data supports could be an interesting avenue for further research.

Our evidence of costly attention does not imply that there shouldn’t be misreaction in significantly higher-stakes environments, as in, e.g., Bradley and Feldman’s (2020) study of flight ticket taxes, or mistakes in high-stakes financial domains such as retirement savings, mortgage contracts, and so forth. Higher-stakes decisions are often also more complex, which reduces available bandwidth for processing of any particular opaque attribute. Extrapolation to settings that differ from ours requires a study of the consequences of domain complexity.

Yet we still think that providing evidence for costly attention in our setting can update researchers’ priors

about the likelihood that similar costly attention mechanisms have a potentially important qualitative role in other settings, including those with higher stakes. Empirical settings that in principle can be analyzed using quasi-experimental analogues of our methods would involve panel datasets where it is possible to observe individuals' behavior across multiple changes in the opaque attributes. An example of an ostensibly opaque price that changes at high-frequency within individual is the utilization cost of electricity-using appliances.

There may also be important interactions between costly attention models and other mechanisms. A key open question is where consumers' highly heterogeneous rules of thumb come from. Theories of memory and "experience effects" may serve as foundations (e.g., Mullainathan, 2002; Malmendier and Nagel, 2016; Bordalo et al., 2020). Moreover, while we find that the majority of consumers are capable of computing post-tax prices, in other domains consumers may reach systematically wrong answers regardless of effort if they have misspecified models of the world (Schwartzstein, 2014; Gagnon-Bartsch et al., 2020).

Despite the possibility of other important sources of mistakes, our study points to attention costs as a plausible and important source of misreaction to opaque prices. The theoretical and empirical framework that we have developed could be fruitfully extended to quantify the importance of costly attention mechanisms in a variety of other economically important settings.

## References

- ABALUCK, J. AND A. ADAMS (forthcoming): "What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses," *Quarterly Journal of Economics*.
- ALLCOTT, H. AND D. TAUBINSKY (2015): "Evaluating Behaviorally-Motivated Policy: Experimental Evidence from the Lightbulb Market," *American Economic Review*, 105, 2501–2538.
- ALLCOTT, H. AND N. WOZNY (2014): "Gasoline Prices, Fuel Economy, and the Energy Paradox," *Review of Economics and Statistics*, 96, 779–795.
- AMBUEHL, S., A. OCKENFELS, AND C. STEWART (2018): "Who Opts In," *Rotman School of Management Working Paper*.
- ANAGOL, S., S. COLE, AND S. SARKAR (2017): "Understanding the Advice of Commissions-Motivated Agents: Evidence from the Indian Life Insurance Market," *The Review of Economics and Statistics*, 99, 1–15.
- ARIELY, D., U. GNEEZY, G. LOEWENSTEIN, AND N. MAZAR (2009): "Large Stakes and Big Mistakes," *Review of Economic Studies*, 76, 451–469.
- BARTOS, V., M. BAUER, J. CHYTILOVÁ, AND F. MATĚJKA (2016): "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *American Economic Review*, 106, 1437–75.
- BENJAMIN, D. J., O. HEFFETZ, M. S. KIMBALL, AND N. SZEMBROT (2014): "Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference," *American Economic Review*, 104, 2698–2735.

- BERNHEIM, B. D. AND D. TAUBINSKY (2018): “Behavioral Public Economics,” in *The Handbook of Behavioral Economics*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, New York: Elsevier, vol. 1, 381–516.
- BHARGAVA, S., G. LOEWENSTEIN, AND J. SYDNOR (2017): “Choose to Lose: Health Plan Choices from a Menu with Dominated Options,” *Quarterly Journal of Economics*, 132, 1319–1372.
- BHATIA, R. AND C. DAVIS (2000): “A Better Bound on the Variance,” *American Mathematical Monthly*, 107, 353–357.
- BORDALO, P., N. GENNAIOLI, AND A. SHLEIFER (2013): “Salience and Consumer Choice,” *Journal of Political Economy*, 121, 803–843.
- (2020): “Memory, Attention, and Choice,” *Quarterly Journal of Economics*, 135, 1399–1442.
- BRADLEY, S. AND N. E. FELDMAN (2020): “Hidden Baggage: Behavioral Responses to Changes in Airline Ticket Tax Disclosure,” *American Economic Journal: Economic Policy*, 12, 58–87.
- BUSHONG, B., J. SCHWARTZSTEIN, AND M. RABIN (2021): “A Model of Relative Thinking,” *The Review of Economic Studies*, 88, 162–191.
- CAPLIN, A. (2016): “Measuring and Modeling Attention,” *Annual Review of Economics*, 8, 379–403.
- CAPLIN, A., D. CSABA, J. LEAHY, AND O. NOV (2020): “Rational Inattention, Competitive Supply, and Psychometrics,” *Quarterly Journal of Economics*, 135, 1681–1724.
- CAPLIN, A., M. DEAN, AND J. LEAHY (2019): “Rational Inattention, Optimal Consideration Sets, and Stochastic Choice,” *Review of Economic Studies*, 86, 1061–1094.
- CARVALHO, L. AND D. SILVERMAN (2019): “Complexity and Sophistication,” Working Paper 26036, National Bureau of Economic Research.
- CHETTY, R., A. LOONEY, AND K. KROFT (2007): “Salience and Taxation: Theory and Evidence,” Working Paper 13330, National Bureau of Economic Research.
- (2009): “Salience and Taxation: Theory and Evidence,” *American Economic Review*, 99, 1145–1177.
- DEAN, M. AND N. NELIGH (2019): “Experimental Tests of Rational Inattention,” Working paper.
- DELLAVIGNA, S. (2009): “Psychology and Economics: Evidence from the Field,” *Journal of Economic Literature*, 47, 315–372.
- ENKE, B. AND F. ZIMMERMANN (2019): “Correlation Neglect in Belief Formation,” *Review of Economic Studies*, 86, 313–332.
- FARHI, E. AND X. GABAIX (2020): “Optimal Taxation with Behavioral Agents,” *American Economic Review*, 110, 298–336.
- FELDMAN, N., J. GOLDIN, AND T. HOMONOFF (2018): “Raising the Stakes: Experimental Evidence on the Endogeneity of Taxpayer Mistakes,” *National Tax Journal*, 71, 201–230.
- FELDMAN, N. E. AND B. J. RUFFLE (2015): “The Impact of Including, Adding, and Subtracting a Tax on Demand,” *American Economic Journal: Economic Policy*, 7, 95–118.

- FUDENBERG, D., P. STRACK, AND T. STRZALECKI (2018): “Speed, Accuracy, and the Optimal Timing of Choices,” *American Economic Review*, 108, 3651–3684.
- GABAIX, X. (2014): “A Sparsity Based Model of Bounded Rationality,” *Quarterly Journal of Economics*, 129, 1661–1710.
- (2019): “Behavioral Inattention,” in *Handbook of Behavioral Economics*, ed. by D. Bernheim, S. DellaVigna, and D. Laibson, Elsevier, vol. 2, 261–343.
- GABAIX, X. AND D. LAIBSON (2006): “Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets,” *Quarterly Journal of Economics*, 121, 505–540.
- GABAIX, X., D. LAIBSON, G. MOLOCHE, AND S. WEINBERG (2006): “Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model,” *American Economic Review*, 96, 1043–1068.
- GAGNON-BARTSCH, T., M. RABIN, AND J. SCHWARTZSTEIN (2020): “Channeled Attention and Stable Errors,” Working paper.
- GILLEN, B., E. SNOWBERG, AND L. YARIV (2019): “Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study,” *Journal of Political Economy*, 127, 1826–1863.
- GOLDIN, J. AND T. HOMONOFF (2013): “Smoke Gets in Your Eyes: Cigarette Tax Salience and Regressivity,” *American Economic Journal: Economic Policy*, 5, 302–336.
- HANDEL, B. R. AND J. T. KOLSTAD (2015): “Health Insurance for “Humans”: Information Frictions, Plan Choice, and Consumer Welfare,” *American Economic Review*, 105, 2449–2500.
- HANNA, R., S. MULLAINATHAN, AND J. SCHWARTZSTEIN (2014): “Learning Through Noticing: Theory and Evidence from a Field Experiment,” *Quarterly Journal of Economics*, 129, 1311–1353.
- HASTINGS, J. S., B. C. MADRIAN, AND W. L. SKIMMYHORN (2013): “Financial Literacy, Financial Education, and Economic Outcomes,” *Annual Review of Economics*, 5, 347–373.
- HAUSMAN, J. (2001): “Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left,” *Journal of Economic Perspectives*, 15, 57–67.
- HEIDHUES, P., B. KŐSZEGI, AND T. MUROOKA (2017): “Inferior Products and Profitable Deception,” *Review of Economic Studies*, 84, 323–356.
- HOSSAIN, T. AND J. MORGAN (2006): “...Plus Shipping and Handling: Revenue (Non)Equivalence in Field Experiments on eBay,” *The B.E. Journal of Economic Analysis & Policy*, 5, 1–30.
- KAHNEMAN, D. (1973): *Attention and Effort*, Prentice-Hall.
- (2003): “Maps of Bounded Rationality: Psychology for Behavioral Economics,” *American Economic Review*, 93, 1449–1475.
- KOSZEGI, B. AND A. SZEIDL (2013): “A Model of Focusing in Economic Choice,” *Quarterly Journal of Economics*, 128, 53–104.
- KROFT, K., J.-W. P. LALIBERTE, R. LEAL-VIZCAINO, AND M. J. NOTOWIDIGDO (2020): “Salience and Taxation with Imperfect Competition,” Working Paper 27409, National Bureau of Economic Research.
- LACETERA, N., D. G. POPE, AND J. R. SYDNOR (2012): “Heuristic Thinking and Limited Attention in the Car Market,” *American Economic Review*, 102, 2206–36.

- LIEDER, F. AND T. L. GRIFFITHS (2019): “Resource-Rational Analysis: Understanding Human Cognition as the Optimal Use of Limited Computational Resources,” *Behavioral and Brain Sciences*, 1–85.
- MAĆKOWIAK, B., F. MATÊJKA, AND M. WIEDERHOLT (2018): “Rational Inattention: A Disciplined Behavioral Model,” Working paper.
- MALMENDIER, U. AND S. NAGEL (2016): “Learning from Inflation Experiences,” *Quarterly Journal of Economics*, 131, 53–87.
- MARTIN, D. (2016): “Rational Inattention in Games: Experimental Evidence,” Working paper.
- MATEJKA, F. AND A. MCKAY (2015): “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model,” *American Economic Review*, 105, 272–298.
- MUELLER, A. I., J. SPINNEWIJN, AND G. TOPA (2021): “Job Seekers’ Perceptions and Employment Prospects: Heterogeneity, Duration Dependence, and Bias,” *American Economic Review*, 111, 324–63.
- MULLAINATHAN, S. (2002): “A Memory-Based Model of Bounded Rationality,” *Quarterly Journal of Economics*, 117, 735–774.
- RATCLIFF, R. (1978): “A Theory of Memory Retrieval,” *Psychological Review*, 85, 59–108.
- REES-JONES, A. AND D. TAUBINSKY (2020): “Measuring Scheduling,” *Review of Economic Studies*, 87, 2399–2438.
- SCHWARTZSTEIN, J. (2014): “Selective Attention and Learning,” *Journal of the European Economic Association*, 12, 1423–1452.
- SHENHAV, A., S. MUSSLICK, F. LIEDER, W. KOOL, T. L. GRIFFITHS, J. D. COHEN, AND M. M. BOTVINICK (2017): “Toward a Rational and Mechanistic Account of Mental Effort,” *Annual Review of Neuroscience*, 40, 99–124, PMID: 28375769.
- SHLAIN, A. (2019): “More than a Penny’s Worth: Left-Digit Bias and Firm Pricing,” Working paper.
- SIMS, C. A. (2003): “Implications of Rational Inattention,” *Journal of Monetary Economics*, 50, 665 – 690.
- TAUBINSKY, D. AND A. REES-JONES (2018): “Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment,” *The Review of Economic Studies*, 85, 2462–2496.

## Tables and Figures

Table 1: Average revealed valuation weights by group

	Standard	Triple	Triple – Standard
(1): High valuation wgt.	1.04 [0.83, 1.24]	1.20 [1.10, 1.31]	0.16 [−0.01, 0.33]
(2): Low valuation wgt.	0.25 [0.08, 0.42]	0.64 [0.57, 0.72]	0.39 [0.26, 0.51]
(3): (1) – (2)	0.79 [0.53, 1.04]	0.56 [0.44, 0.67]	−0.23 [−0.43, −0.04]
(4): Full sample	0.48 [0.32, 0.63]	0.79 [0.72, 0.86]	0.31 [0.20, 0.42]

Rows (1) and (2) of this table present estimates for the high and low valuation weight groups, whose construction is described in Section 5.1. Row (3) presents the difference of the estimates in rows (1) and (2), for each column. Row (4) presents estimates using the full sample. The “Standard” column contains estimates of store  $B$  valuation weights in each of the two groups, as well as the differences between these groups. The “Triple” column contains estimates of store  $C$  valuation weights in each of the two groups, as well as the differences between these groups. The “Triple – Standard” column presents estimates of  $E[\theta_{ijC}] - E[\theta_{ijB}]$  for each of the two groups in rows (1) and (2), and contains the differences in differences in row (3). Bootstrapped confidence intervals from 1000 replications, clustered at the subject level, are reported in brackets.

Table 2: Average revealed valuation weights by adjustment group

	Standard	Triple	Triple – Standard
(1): Low Adj.	0.85 [0.64, 1.07]	0.86 [0.77, 0.96]	0.01 [−0.15, 0.17]
(2): High Adj.	0.34 [0.17, 0.51]	0.76 [0.68, 0.85]	0.43 [0.30, 0.55]
(3): (1) – (2)	0.52 [0.30, 0.75]	0.10 [−0.01, 0.20]	−0.42 [−0.60, −0.24]
(4): Full sample	0.48 [0.32, 0.63]	0.79 [0.72, 0.86]	0.31 [0.20, 0.42]

Rows (1) and (2) of this table present estimates for the low and high adjustment groups, whose construction is described in Section 5.2. Row (3) presents the difference of the estimates in rows (1) and (2), for each column. Row (4) presents estimates using the full sample. The “Standard” column contains estimates of store  $B$  valuation weights in each of the two groups, as well as the differences between these groups. The “Triple” column contains estimates of store  $C$  valuation weights in each of the two groups, as well as the differences between these groups. The “Triple – Standard” column presents estimates of  $E[\theta_{ijC}] - E[\theta_{ijB}]$  for each of the two groups in rows (1) and (2), and contains the differences in differences in row (3). Bootstrapped confidence intervals from 1000 replications, clustered at the subject level, are reported in brackets.



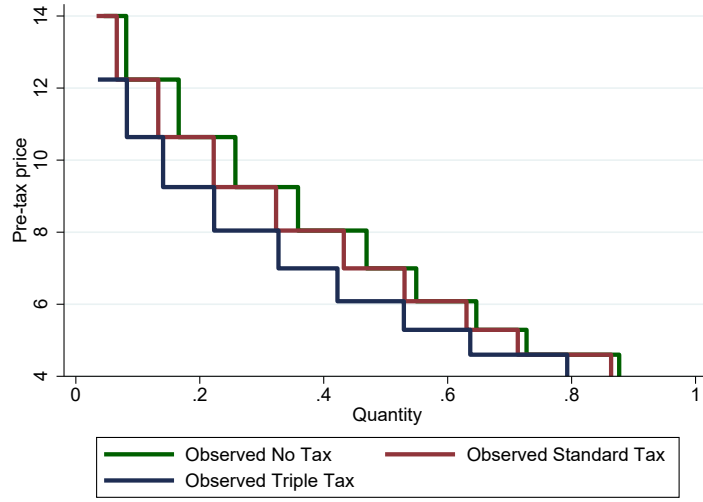
Table 3: Bounds on the dispersion of revealed valuation weights

	Standard ( $\theta_B$ )	Triple ( $\theta_C$ )	$\theta_B - \theta_C$
Variance (Lower Bound)	0.83 [0.52]	0.71 [0.59]	0.86 [0.31]
Supremum (Lower Bound)	2.21 [1.55]	1.69 [1.54]	0.94 [0.16]

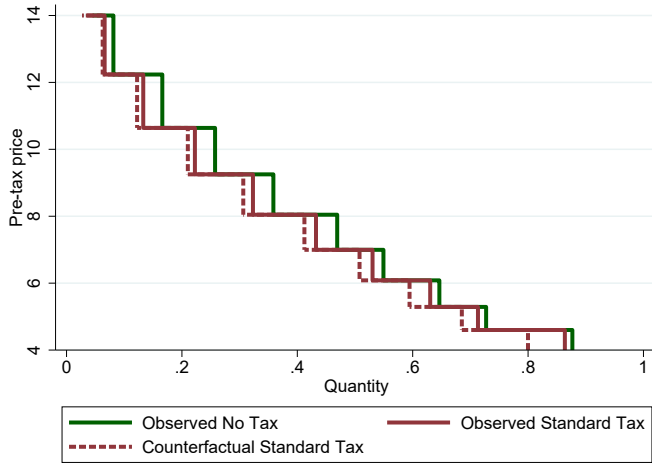
Columns (1) and (2) of this table present store-specific estimates of the lower bound on  $Var[\theta_{ijB}]$  and  $Var[\theta_{ijC}]$ , and on the supremum  $\bar{\theta}$ . Column (3) presents estimates of the lower bound of  $Var[\theta_{ijB} - \theta_{ijC}]$  and the supremum of  $\theta_{ijB} - \theta_{ijC}$ . The methodology is described in Section 6.1 and the estimating equations are described in Section 6.2. Fifth percentile results from 1000 bootstrap replications, clustered by subject, are reported in brackets.

Figure 1: Demand curves

(a) Observed demand curves



(b) Observed vs. counterfactual demand: standard taxes



(c) Observed vs. counterfactual demand: triple taxes

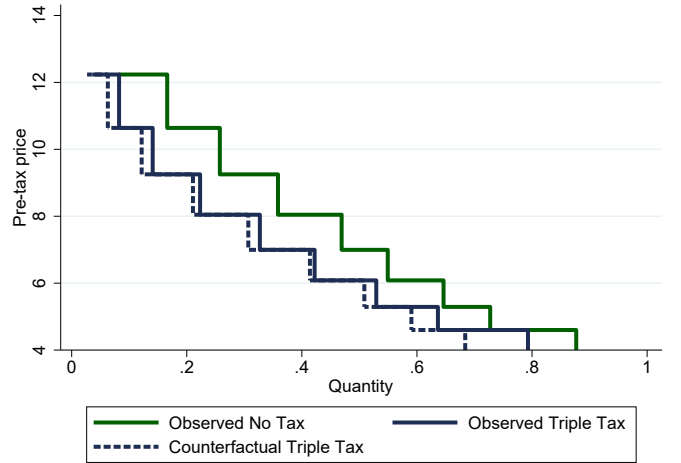
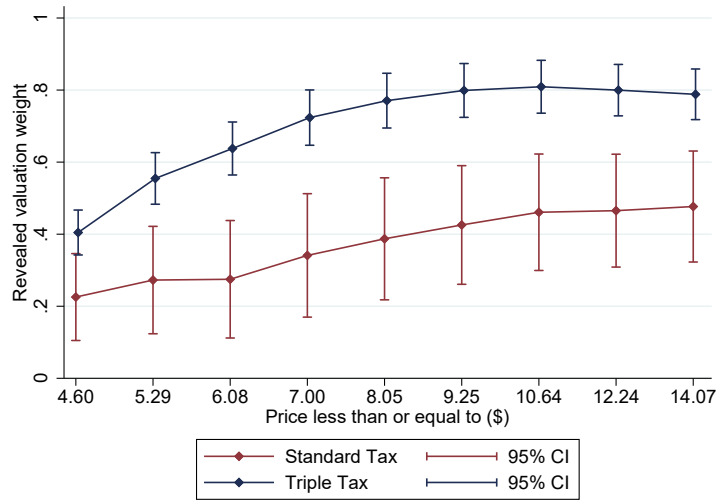


Figure 1 presents demand curves, averaging across all nine products. We start by constructing demand curves  $D_{jk}(p)$  where  $j$  indexes products and  $k$  indexes the store type, A (no tax), B (standard tax), or C (triple tax). These correspond to the fraction of consumers willing to buy product  $j$  at each price on the price list. For stores A and B, choices at ten prices from \$4.00 to \$14.07 were observed, and for store C choices at nine prices from \$4.00 to \$12.24 were observed. Panel (a) presents the average demand curves  $D_{avg,k}(p) := \frac{1}{9} \sum_j D_{jk}(p)$  for each tax-environment using observed choices. For panel (b), we construct the demand curves that would be expected in store B if consumers reacted to the taxes fully. We then compare this to the observed demand in stores A and B. For panel (c), we construct the demand curves that would be expected in store C if consumers reacted to the taxes fully. We then compare this to the observed demand in stores A and C. We construct the counterfactual demand through linear interpolation, as described in Section 4.1.

Figure 2: Average revealed valuation weights by stakes

(a) Average revealed valuation weight for posted prices at or below a cutoff



(b) Average revealed valuation weight by average tax owed

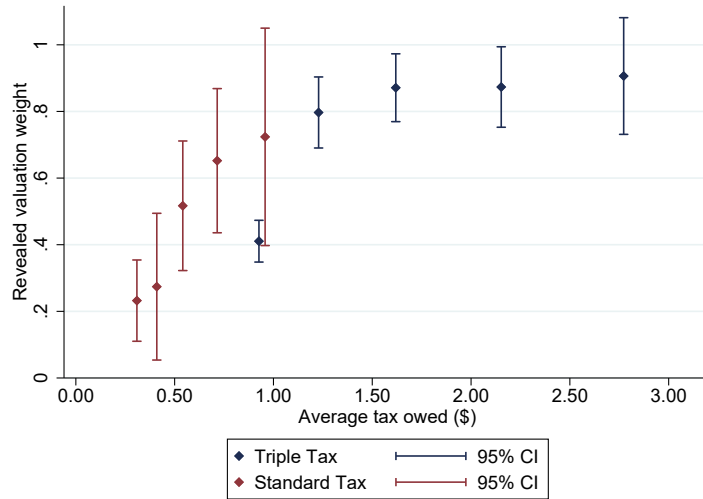
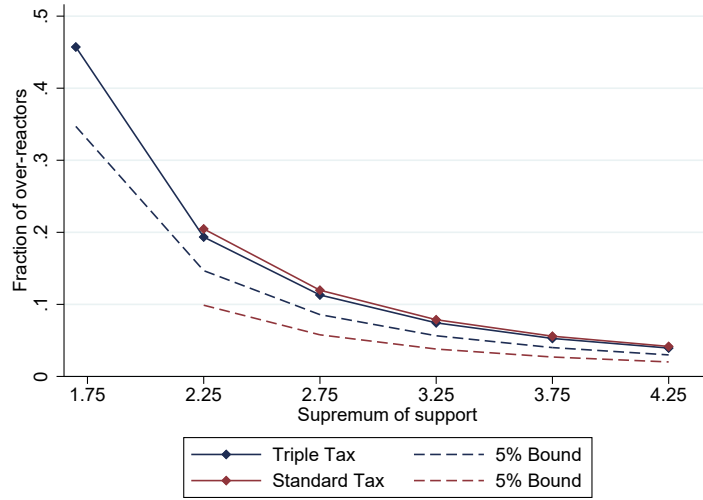


Figure 2a presents estimates of store-specific estimates of average  $\theta$  for prices less than or equal to the cutoff specified on the x-axis.  $\theta$  is defined as the revealed valuation weight that consumers place on the sales tax, with  $\theta = 0$  corresponding to complete neglect of the tax and  $\theta = 1$  corresponding to putting the same weight on the tax as on the salient posted price. Each value on the x-axis corresponds to a different posted price on the price list presented to consumers. The results are estimated using equation (4) for prices below the cutoff.

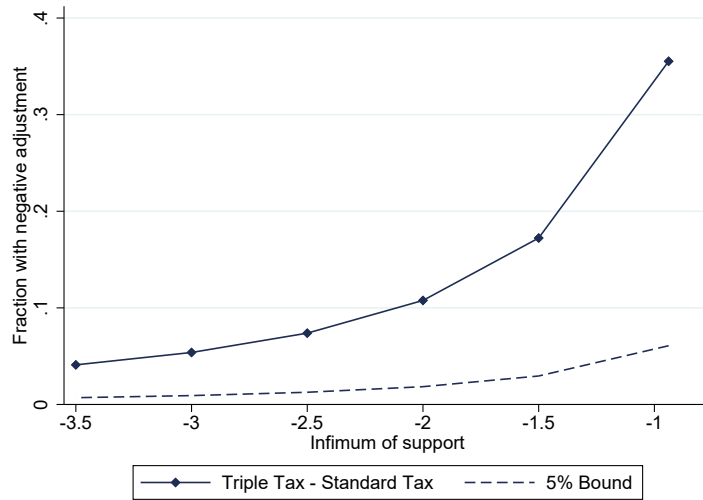
Figure 2b presents store-specific estimates of average  $\theta$  by the average tax owed within each bin. For each tax environment—store B and store C—each bin is constructed by dividing the prices in the experiment into 5 ordered pairs. The average tax owed is constructed by taking the average of the two prices in each bin, and multiplying it by the average tax rate in stores B and C, respectively. The estimating equation is an extension of equation (4), described in equation (5). Standard errors are clustered at the subject level.

Figure 3: Lower bound on the fraction of individuals who overreact and whose valuation weight  $\theta$  decreases in response to higher stakes

(a) Lower bound on the fraction of individuals who overreact to the sales tax



(b) Lower bound on the fraction of individuals whose valuation weight  $\theta$  decreases in response to higher stakes



Panel (a) of this figure presents store-specific estimates for the lower bound on the fraction of consumers with  $\theta > 1$ , as a function of the supremum of the support of  $\theta$ , while panel (b) presents estimates for the lower bound on  $Pr(\theta_{iC} - \theta_{iB}) < 0$  as a function of the infimum of the support.  $\theta$  is defined as the revealed valuation weight that consumers place on the sales tax, with  $\theta = 0$  corresponding to complete neglect of the tax and  $\theta = 1$  corresponding to putting the same weight on the tax as on the salient posted price. The lower bound on  $Pr(\theta_{ik}) > 1, k \in B, C$  is estimated from equation (13). The lower bound for the fraction of consumers who adjust their valuation weight downward in response to higher stakes is estimated using equation (14). The dashed lines present the 5% lower bound computed from a percentile-based bootstrap procedure (1000 replications, clustered at the subject level).

# Online Appendix

## Rules of Thumb and Attention Elasticities Evidence from Under- and Overreaction to Taxes

William Morrison and Dmitry Taubinsky

### Table of Contents

---

<b>A Results for the Shannon model and the Gabaix (2014) Sparsity Model</b>	<b>36</b>
A.1 Shannon model with heterogeneous priors . . . . .	36
A.2 The Gabaix (2014) model . . . . .	38
<b>B General results about costly attention models and proofs for Appendix A</b>	<b>38</b>
B.1 Lemma for revealed valuation weight representation . . . . .	38
B.2 Models in the spirit of rational inattention . . . . .	39
B.3 Gabaix (2014) model of attention adjustment . . . . .	42
B.4 Proof of Proposition A.1 . . . . .	44
B.5 Proof of Proposition A.2 . . . . .	46
<b>C Proofs of Propositions in the body of the paper</b>	<b>47</b>
C.1 Proof of Proposition 1 . . . . .	47
C.2 Proof of Proposition 2 . . . . .	48
<b>D Predictions 3 and 4 with heterogeneity in attention costs</b>	<b>50</b>
D.1 Simple example with binary attention costs . . . . .	51
<b>E Relation to sales tax literature</b>	<b>52</b>
E.1 Detailed comparisons to TRJ . . . . .	53
<b>F Counterfactual demand curve construction</b>	<b>54</b>
<b>G Interpreting coefficients in the probit regression</b>	<b>55</b>
<b>H Point estimates and confidence intervals for Figures 2a and 2b</b>	<b>56</b>
<b>I Relationship between average revealed valuation weights and marginal utility of money</b>	<b>57</b>
<b>J Covariates of attention</b>	<b>59</b>
J.1 Local tax rate variation . . . . .	59

J.2	Demographics . . . . .	61
<b>K</b>	<b>Alternative construction of proxies for valuation weights</b>	<b>64</b>
<b>L</b>	<b>Replication of results restricting to participants with nearly-accurate beliefs and high computational ability</b>	<b>66</b>
<b>M</b>	<b>Replication of main results without excluding study participants failing comprehension questions or violating monotonicity</b>	<b>70</b>
<b>N</b>	<b>Replication of main results excluding participants who always or never buy a product in at least one store</b>	<b>73</b>
<b>O</b>	<b>Order Effects</b>	<b>75</b>
<b>P</b>	<b>Comparison of demand curves to Amazon.com prices</b>	<b>77</b>
<b>Q</b>	<b>Welfare implications of overreaction</b>	<b>78</b>
<b>R</b>	<b>Additional details of the experiment</b>	<b>79</b>
R.1	Instructions . . . . .	81
R.2	Text of questions . . . . .	86
R.3	Items used in the study . . . . .	90

## A Results for the Shannon model and the Gabaix (2014) Sparsity Model

Here show that the main results and predictions of Section 2.2 generalize to prominent models of inattention with continuous costs. We present proofs of these results in Appendix B after establishing additional results.

### A.1 Shannon model with heterogeneous priors

In contrast to our simple example in the text, the Shannon model allows for a range of cognitive effort. In our setting, the Shannon model posits that consumers pay some cost to adjust their initial weight  $r$  closer to the truth. Higher attention costs move perceptions closer to the truth in expectation, but this link is stochastic.

Formally, the Shannon model is as follows in our setting:

1. Consumers choose a joint distribution  $\pi$  over signals and  $q_o$ . Without loss of generality, we associate each signal with a posterior belief  $\rho$  of the probability that  $q_o = t$ . The revision  $\rho - r$  can be thought of as the extent to which the consumer adjusts his estimate closer to  $t$  after thinking more. The distribution  $\pi$  must satisfy the Bayesian consistency requirement  $\rho = \frac{r\pi(\rho|t)}{r\pi(\rho|t) + (1-r)\pi(\rho|\hat{t})}$ .
2. The cost of the information structure  $\pi$  is  $c(\pi) = \lambda (H(r) - E_\pi[H(\rho)])$ , where  $H(x) = -x \log x - (1-x) \log(1-x)$  is the entropy of a probability distribution that places probability  $x$  on  $q_o = t$  and probability  $1-x$  on  $q_o = \hat{t}$ .

3. Consumers choose to buy at a posterior  $\rho$  if and only if  $v - p_s - \sigma(\rho t + (1 - \rho)\hat{t}) > 0$ . We will use  $b(\rho) \in \{0, 1\}$  to note whether it is optimal for a consumer to buy given  $\rho$ .
4. Consumers thus choose  $\pi$  to maximize  $E[(v - p_s - \sigma q_o)b(\rho)] - c(\pi)$ .

As with the binary attention model, we show that the Shannon model has a simple reduced-form representation. We derive this result using the necessary and sufficient conditions of the posterior-based approach provided in Caplin et al. (2019).

**Proposition A.1.** *For each triplet  $\Xi = (\lambda, r, \hat{t})$  and stakes  $\sigma$  in the Shannon model, there exists a distribution  $F_{\Xi, \sigma}$  such that the behavior of all consumers with parameters  $\Xi$  can be represented by a revealed valuation weight model in which consumers choose to buy if and only if  $v \geq p_s + \theta p_o$ , where  $\theta \sim F_{\Xi, \sigma}$ . The weights satisfy:*

1.  $\lim_{\sigma \rightarrow \infty} F_{\Xi, \sigma} \xrightarrow{d} 1$ . That is, relative misreaction converges (in distribution) to zero as the stakes become large.
2. The mean valuation weight  $\bar{\theta}_{\Xi, \sigma} = \int \theta dF_{\Xi, \sigma}(\theta)$  is increasing in  $\hat{t}$ , with  $\bar{\theta}_{\Xi, \sigma} = 1$  when  $\hat{t} = t$ . The relative average misreaction,  $|1 - \bar{\theta}_{\Xi, \sigma}|$ , is decreasing in  $r$ .

Proposition A.1 shows that behavior in the Shannon model can be represented using a reduced form similar to the one we derived in the binary attention strategy example. This reduced form follows the same comparative statics. The main difference is that because the consequences from exerting mental effort are stochastic in the Shannon model, a consumer's valuation weight is represented by  $\theta = \bar{\theta} + v$ , where  $v$  is a mean-zero error term that varies from decision to decision, and  $\bar{\theta}$  is the stable component across decisions. Our experiment, which focuses on stable individual differences, will focus on characterizing the distribution of  $\bar{\theta}$ , but will not be informative about the idiosyncratic component  $v$ . When we study individual differences in consumers' valuation weights in the empirical analysis in Sections 5 and 6, what we mean with respect to the Shannon model is differences in  $\bar{\theta}$ , not  $\bar{\theta} + v$  (a slight abuse of terminology).

Intuitively, the representation continues to hold because like the reduced-form revealed valuation weight model, the Shannon model also predicts that the probability of choosing to buy the item should depend only on the transparent surplus  $v - p_s$ , and not on  $v$  and  $p_s$  separately. Although this property holds for many costly attention models, it does not hold for all models that generate misreaction. For example, salience and focusing models such as those of Bordalo et al. (2013) and Koszegi and Szeidl (2013) do not always have this property.

Proposition A.1 shows that when stakes are large, the stable components  $\bar{\theta}$  converge to 1 and the stochasticity vanishes (i.e.,  $v$  converges to 0 in distribution). The last part of Proposition A.1 also shows that there will be stable individual differences in  $\theta$  that are shaped by consumers' initial perceptions of  $q_o$ . For example, consumers who initially overestimate  $q_o$  have  $\theta > 1$ , while consumers who initially underestimate  $q_o$  have  $\theta < 1$ . Consequently, in the presence of both over- and underestimation, the Shannon model predicts that a large increase in stakes lowers  $\theta$  for some consumers and increases  $\theta$  for other consumers, in line with our binary attention example.

## A.2 The Gabaix (2014) model

A second model that allows for a continuous range of cognitive effort is the Gabaix (2014) model. We utilize the binary action extension of the model.<sup>34</sup> In this model, the consumer chooses a weight  $m \in [0, 1]$  to form an estimate  $\hat{q}_o(m) = mq_o + (1 - m)\bar{q}_o$ , where  $\bar{q}_o = rt + (1 - r)\hat{t}$  is the default perception. The cost of choosing  $m > 0$  is given by  $\lambda m^\alpha$ , for  $\alpha \geq 0$ .

The consumer approximates the benefits of choosing  $m > 0$  as follows. First, the consumer computes the benefits of choosing the full attention strategy  $m = 1$ , which we denote by  $B$ . As we have shown in Section 2.2, the benefit of acquiring information is given by  $B = \min((1 - r)(v - p_s - t), r(p_s + t - v))$ . Consumers then approximate the benefit of choosing  $m \in [0, 1]$  by the quadratic approximation  $B - (1 - m)^2 B$ .

The special case  $\alpha = 0$  corresponds to our binary attention example in Section 2.2. However, for  $\alpha > 0$  this model allows for partial attention, like the Shannon model. For example, when  $\alpha = 1$ , the consumer chooses  $m^* = \max(1 - \lambda/(2B), 0)$ . When  $\alpha = 2$  the consumer chooses  $m^* = \frac{B}{\lambda + B}$ .

**Proposition A.2.** *For each triplet  $\Xi = (\lambda, r, \hat{t})$  and stakes  $\sigma$  in the Gabaix (2014) model, there exists a  $\theta_{\Xi, \sigma} \in \mathbb{R}$  such that consumers with parameters  $\Xi$  can be represented by a revealed valuation weight model in which consumers choose to buy if and only if  $v \geq p_s + \theta_{\Xi, \sigma} p_o$ . The valuation weights satisfy:*

1.  $|1 - \theta_{\Xi, \sigma}|$  is decreasing in  $\sigma$  and converges to zero as  $\sigma \rightarrow \infty$ . That is, relative misreaction is decreasing in stakes and converges to zero.
2. The valuation weight  $\theta_{\Xi, \sigma}$  is increasing in  $\hat{t}$ , with  $\theta_{\Xi, \sigma} = 1$  when  $\hat{t} = t$ . Moreover,  $|1 - \theta_{\Xi, \sigma}|$  is decreasing in  $r$ .

Like Proposition A.1, Proposition A.2 shows that the Gabaix (2014) model has a simple revealed valuation weight representation that features all of the properties of the binary attention example. As with the binary attention example, it is possible to obtain closed-form solutions for  $\theta$  in terms of the model primitives when closed-form solutions exist for the choice  $m^*$ , as in the simple examples for  $\alpha = 1, 2$  that we summarized above.

Unlike the Shannon model, the revealed valuation weights in this model are deterministic rather than stochastic. Whether within-person stochasticity of attention is an empirically large phenomenon remains an open question; our experiment will focus only on stable individual differences.

## B General results about costly attention models and proofs for Appendix A

### B.1 Lemma for revealed valuation weight representation

We begin by establishing the following set of results, which we will use repeatedly throughout the proofs.

**Lemma B.1.** *Suppose that the probability that consumer  $i$  chooses to buy the product given a valuation  $v$ , salient price  $p_s$ , and opaque price  $p_o$  is given by  $G(v - p_s, p_o)$ , with  $G$  increasing in the first argument.*

<sup>34</sup>This model is developed in Appendix XV.F of Gabaix (2014). We thank Xavier Gabaix for kindly distilling this model for us in personal communication, and writing out the special case that is applicable to our economic environment. Our formulation follows the sketch provided to us by Xavier Gabaix.



Then the consumer's decision process can be represented as if the consumer chooses to buy if and only if  $v - p_s - \theta p_o \geq 0$ , where  $\theta$  is a random variable whose distribution is independent of  $v$  and  $p_s$ . Moreover:

1. If  $G \in \{0, 1\}$  for all  $v, p_s, p_o$ , then the distribution of  $\theta$  is degenerate (i.e., it is a scalar).
2. If  $G_2(u, p_o) \geq G_1(u, p_o)$  for all  $u$ , then the distribution of  $\theta$  corresponding to  $G_2$  first order stochastically dominates the distribution of  $\theta$  corresponding to  $G_1$ .
3. If  $G(p_o - \delta, p_o) = 1 - G(p_o + \delta, p_o)$  for all  $\delta$ , then the distribution of  $\theta$  is symmetric about 1 and satisfies  $E[\theta] = 1$ .

*Proof.* Fix  $p_o$ , and let  $F(\theta|p_o)$  be the distribution of  $\theta$  in the reduced-form representation given by  $F(\theta|p_o) = G(p_o\theta, p_o)$ . In the reduced-form model, the probability that a consumer buys is given by  $Pr\left(\theta \leq \frac{v-p_s}{p_o}\right) = F\left(\frac{v-p_s}{p_o}|p_o\right) = G(v-p_s, p_o)$ .

If  $G \in \{0, 1\}$ , as in condition (1), then there exists a value  $u^\dagger = v - p_s$  such that the consumer buys if and only if  $v - p_s \geq u^\dagger$ . Equivalently, the consumer buys if  $v - p_s \geq \theta p_o$ , where  $\theta = u^\dagger / p_o$ .

To prove condition (2), note that the assumption implies that  $F_2(\theta|p_o) = G_2(p_o\theta, p_o) \geq G_1(p_o\theta, p_o) = F_1(\theta|p_o)$ .

To prove (3), note that the assumptions imply that  $F(\theta|p_o) = G(p_o\theta, p_o) = G(p_o(1-\theta), p_o) = F(1-\theta, p_o)$ . This implies that the density function corresponding to  $F$  is symmetric around a mode of 1. Therefore,  $E[\theta] = 1$ .  $\square$

Lemma B.1 implies that any attention model that predicts that consumers are more likely to buy when the transparent surplus  $v - p_s$  is higher can be represented using the reduced-form attention weight model. The additional statements in the Lemma help provide further structure on the attention weights. For example, when the buying decision is not stochastic, as in the Gabaix (2014) model, the reduced-form valuation weight will not be stochastic either.

## B.2 Models in the spirit of rational inattention

We consider a model in which  $p_o = \sigma\omega$ , where  $\sigma$  are the salient stakes, and  $\omega \in \Omega$  is the initially unknown state. The set  $\Omega$  includes the true value  $q_o$ . A consumer has a prior  $\mu$  about  $\omega$ . The consumer selects a probability distribution over signals, with each signal identified with a posterior  $\gamma \in \Gamma = \Delta(\Omega)$ . Formally, the consumer selects a mapping  $\pi : \Omega \rightarrow \Delta(\Gamma)$ . The posterior  $\gamma$  must satisfy

$$\gamma(\omega) = Pr(\omega|\gamma) = \frac{\mu(\omega)\pi(\gamma|\omega)}{\int_{\omega'} \mu(\omega')\pi(\gamma|\omega')d\omega'}$$

where  $\pi(\gamma|\omega)$  is the probability of signal  $\gamma$  given state  $\omega$ . The cost of selecting  $\pi$  is  $K(\pi) \in \mathbb{R}^+$ , where  $\mathbb{R}^+$  denotes the non-negative reals. Given a posterior  $\gamma$ , the consumer chooses to buy if and only if  $v - p_s - \sigma \int \omega\gamma(\omega)d\omega \geq 0$ .

The net utility of choosing  $\pi$  is given by  $V(\pi) = Q(\pi)(v - p_o) - R(\pi)$ , where  $Q(\pi) = \int \gamma(\omega)\pi(\gamma)d\omega d\gamma$  is the expected probability of buying, and

$$R(\pi) = - \int \omega\gamma(\omega)\pi(\gamma)d\omega d\gamma - K(\pi)$$

is the next expected cost, inclusive of both the attention cost and expected size of the opaque price.

**Lemma B.2.** *Let  $\pi$  be the information structure chosen for  $u = v - p_s$  and let  $\pi'$  be the information structure chosen for  $u' = v' - p'_s$ , with  $u' < u$ . Then  $Q(\pi) \geq Q(\pi')$*

*Proof.* Suppose the contrary:  $Q(\pi) < Q(\pi')$ . Then  $uQ(\pi) - R(\pi) \geq uQ(\pi') - R(\pi')$ , which implies  $u(Q(\pi) - Q(\pi')) \geq R(\pi) - R(\pi')$ . Similarly, if  $\pi$  is optimal at  $u'$ , then  $u'(Q(\pi') - Q(\pi)) \geq R(\pi') - R(\pi)$ , or  $u'(Q(\pi) - Q(\pi')) \leq R(\pi) - R(\pi')$ . This implies that  $u(Q(\pi) - Q(\pi')) \geq u'(Q(\pi) - Q(\pi'))$ , which is impossible if  $u > u'$ .  $\square$

Lemma B.2 implies that the ex-ante expected likelihood of buying is increasing in  $u = v - p_s$ . However, it does not by itself imply that the ex-ante expected likelihood of buying is increasing in every state, and  $\omega = q_o$  in particular. If as  $u$  increases, the relative likelihood of buying in  $\omega = q_o$  decreases sufficiently quickly, then the likelihood of buying in that state would not decrease. Although we have not confirmed this exhaustively, this seems like an unlikely result. Below, we confirm that the likelihood of buying in state  $\omega = q_o$  is increasing in  $u$  when attention costs are proportional to (Shannon) entropy reduction.

**Lemma B.3.** *Let the cost function be given by  $\lambda(H(\mu) - E[H(\gamma)])$ , where  $H$  denotes entropy. Then the probability of buying in the particular state  $\omega = q_o$  is increasing in  $u = v - p_s$ , and does not depend on  $v$  and  $p_s$  separately.*

*Proof.* If the probability of buying is in the interior  $(0, 1)$ , then Theorem 1 in Matejka and McKay (2015) implies that

$$Pr(buy|\omega = q_o) = \frac{Qe^{\frac{u - \sigma q_o}{\lambda}}}{(1 - Q) + Qe^{\frac{u - \sigma q_o}{\lambda}}} \quad (16)$$

Since the right-hand side of (16) is increasing in both  $Q$  and  $u$ , Lemma B.2 implies that  $Pr(buy|\omega = q_o)$  is increasing in  $u$ .  $\square$

The last result immediately leads to the following:

**Proposition B.1.** *If attention costs are proportional to entropy reduction, then the consumer's behavior can be represented by the reduced-form valuation weight model.*

*Proof.* Lemma B.3 implies that the probability that the consumer buys the product can be written as  $G(v - p_s, p_o)$ , where  $G$  is increasing in the first argument. Lemma B.1 leads to the result.  $\square$

A key general comparative static is that systematic misreaction ( $E[\theta] \neq 1$ ) cannot occur if the consumer has an unbiased prior. Thus, systematic misreaction can only come about from biased initial perceptions.

**Proposition B.2.** *Suppose that the prior  $\mu$  is symmetric around  $q_o$ ; i.e.,  $\mu(\omega) = \mu(\omega')$  if  $|\omega - q_o| = |\omega' - q_o|$ . If attention costs are proportional to entropy reduction, then the consumer's behavior is represented with a reduced-form valuation weight model in which  $E[\theta] = 1$ .*

*Proof.* For  $u = \sigma q_o + \delta$ , let  $\alpha_\delta(\omega)$  be the probability of choosing to buy in state  $\omega$  at the optimal attention strategy. Now when  $u = \sigma q_o - \delta$ , the relative gains from not buying at  $\omega' = q_o - (\omega - q_o)$  are equal to the relative gains from buying at  $\omega$  when  $u = \sigma q_o + \delta$ . By symmetry, this implies that at the optimal attention strategy, the probability of buying when  $u = \sigma q_o - \delta$ , denoted  $\alpha_{-\delta}(\omega)$ , must satisfy  $1 - \alpha_{-\delta}(2q_o - \omega) = \alpha_\delta(\omega)$ . In particular, this implies that  $\alpha_\delta(q_o) + \alpha_{-\delta}(q_o) = 1$ . Point 3 of Lemma B.1 then implies the result.  $\square$

**Generalization to other cost functions:** In general, Proposition B.2 will hold whenever i) there exists a reduced-form valuation weight representation and ii) the attention cost function satisfies a basic “anonymity” assumption such that the “labels” of the states do not matter, only the probabilities of the states and their contingent payoffs.

Finally, we establish a general result about stakes and attention costs.

**Proposition B.3.** *Let the cost function be given by  $\lambda(H(\mu) - E[H(\gamma)])$ , where  $H$  denotes entropy. As  $\lambda \rightarrow 0$  or as  $\sigma \rightarrow \infty$ , the distribution of  $\theta$  approaches, in probability, a distribution that places unit mass on 1.*

*Proof.* We first show that as  $\lambda \rightarrow 0$ ,  $Pr(\text{buy} | v - p_s - p_o \geq 0) \rightarrow 1$  and  $Pr(\text{buy} | v - p_s - p_o < 0) \rightarrow 0$ . Let  $Q$  denote the ex-ante expected probability of buying. By Proposition 1 of Caplin et al. (2019),  $Q$  must satisfy

$$\sum_{\omega} \frac{\exp\left(\frac{v - p_s - \sigma \omega}{\lambda}\right) \mu(\omega)}{Q \exp\left(\frac{v - p_s - \sigma \omega}{\lambda}\right) + (1 - Q)} \leq 1,$$

with equality if  $Q > 0$ . Now

$$\begin{aligned} \lim_{\lambda \rightarrow 0} \sum_{\omega} \frac{\exp\left(\frac{v - p_s - \sigma \omega}{\lambda}\right) \mu(\omega)}{Q \exp\left(\frac{v - p_s - \sigma \omega}{\lambda}\right) + (1 - Q)} &= \lim_{\lambda \rightarrow 0} \sum_{\omega \leq v - p_s} \frac{\exp\left(\frac{v - p_s - \sigma \omega}{\lambda}\right) \mu(\omega)}{Q \exp\left(\frac{v - p_s - \sigma \omega}{\lambda}\right) + (1 - Q)} \\ &= \sum_{\omega \leq v - p_s} \frac{\mu(\omega)}{\lim_{\lambda \rightarrow 0} Q} \\ &= \frac{Pr(\omega \leq v - p_s)}{\lim_{\lambda \rightarrow 0} Q} \end{aligned}$$

Thus  $Q_0 := \lim_{\lambda \rightarrow 0} Q \geq Pr(\omega \leq v - p_s)$ . If  $Pr(\omega \leq v - p_s) = 1$  then we are done since in that case the consumer buys with probability 1, just like the fully attentive consumer (recall the assumption that  $q_o \in \Omega$ ). If  $Pr(\omega \leq v - p_s) = 0$  then  $Q_0 = 0$  by rational expectations, so we are again done.

Consider now the case in which  $Pr(\omega \leq v - p_s) \in (0, 1)$ , which implies that  $Q_0 = Pr(\omega \leq v - p_s) \in (0, 1)$ . In this case, Theorem 1 of Matejka and McKay (2015) implies that

$$\begin{aligned} \lim_{\lambda \rightarrow 0} Pr(\text{buy} | \omega = q_o) &= \lim_{\lambda \rightarrow 0} \sum_{\omega \leq v - p_s} \frac{Q \exp\left(\frac{v - p_s - \sigma q_o}{\lambda}\right)}{(1 - Q) + Q \exp\left(\frac{v - p_s - \sigma q_o}{\lambda}\right)} \\ &= \begin{cases} 0 & \text{if } v - p_s - \sigma q_o < 0 \\ 1 & \text{if } v - p_s - \sigma q_o > 0 \end{cases} \end{aligned}$$

Consider now the impact of increasing  $\sigma$ . It is sufficient to show that as  $l \rightarrow \infty$ ,  $Pr(buy|lv, lp_s, lp_o) \rightarrow 0$  if  $v - p_s - p_o < 0$  and  $Pr(buy|lv, lp_s, lp_o) \rightarrow 1$  if  $v - p_s - p_o > 0$ . This is because  $Pr(\theta > x) = Pr(buy|v - p_s = xp_o)$ . Thus if  $x > 1$  and  $Pr(buy|lv, lp_s, lp_o) \rightarrow 0$  if  $v - p_s - p_o < 0$ , then  $Pr(buy|v - p_s = xp_o) \rightarrow 0$  as  $\sigma \rightarrow \infty$ . Conversely,  $Pr(\theta < x) = 1 - Pr(buy|v - p_s = xp_o)$ . Thus if  $x < 1$  and  $Pr(buy|lv, lp_s, lp_o) \rightarrow 1$  if  $v - p_s - p_o > 0$ , then  $Pr(buy|v - p_s = xp_o) \rightarrow 1$  as  $\sigma \rightarrow \infty$ .

To that end, note that the impact on attention strategies of scaling up payoffs by  $l$  is equivalent to scaling down the attention costs to  $\lambda/l$ . But since behavior approaches the full attentive benchmark when  $\lambda \rightarrow 0$ , the conclusion follows.  $\square$

**Generalization to other cost functions:** The result about stakes follows more generally. If attention costs are given  $K = \lambda K_o$ , then scaling up stakes by  $k$  has the same impact on attention strategies as scaling down attention costs to  $\lambda/k$ . Then the reasoning in the proof of Proposition B.3 implies that any cost function that generates behavior that is continuous in  $\lambda$  at 0 will also generate the prediction that the distribution of  $\theta$  approaches 1.

### B.3 Gabaix (2014) model of attention adjustment

Again, we consider a model in which  $p_o = \sigma\omega$ , where  $\sigma$  are the salient stakes, and  $\omega \in \Omega$  is the initially unknown state. The set  $\Omega$  includes the true value  $q_o$ . The consumer has a prior  $\mu$  about  $\omega$ . We set  $\bar{q}_o = \int \omega \mu(\omega)$ .

Consumers form an estimate of  $q_o$  given by  $q_o^*(m) = mq_o + (1 - m)\bar{q}_o$ . The case  $m = 0$  corresponds to no adjustment and the case  $m = 1$  corresponds to full adjustment. The attention cost of choosing  $m \geq 0$  is  $\lambda m^\alpha$ , where  $\alpha \geq 0$ . Consumers approximate the perceived benefit of choosing  $m \geq 0$  with the quadratic approximation  $B - (1 - m)^2 B$ , where  $B$  is the benefit of full information. Formally,

$$B = \int_{\sigma\omega \leq v - p_s} (v - p_s - \omega) \mu(\omega) \text{ if } v - p_s - \sigma\bar{q}_o < 0$$

$$B = \int_{\sigma\omega \geq v - p_s} (p_s + \omega - v) \mu(\omega) \text{ if } v - p_s - \sigma\bar{q}_o \geq 0$$

**Lemma B.4.** *The consumer's propensity to buy is monotonically increasing in  $u = v - p_s$ .*

*Proof.* To establish monotonicity in  $u = v - p_s$ , we need to show that as  $u$  increases, the consumer cannot go from buying to not buying. Suppose first that  $u - \sigma\bar{q}_o < 0$ , so that the consumer does not buy when  $m = 0$ . If the consumer buys at the optimal  $m$  at that  $u$ , then  $u - \sigma(mq_o + (1 - m)\bar{q}_o) \geq 0$  by definition, which is possible only if  $u - \sigma q_o > 0$ . Now since  $B(u) = \int_{\sigma\omega \leq u} (u - \sigma\omega) \mu(\omega)$  for  $u - \sigma\bar{q}_o < 0$ ,  $B$  is an increasing function of  $u$  when  $u - \sigma\bar{q}_o < 0$ . And since  $m$  is increasing in  $B$ , this means that  $m$  is increasing in  $u$  when  $u < \sigma q_o$ .

Let  $m'$  be the chosen attention weight for some  $u' \in (u, \sigma q_o)$ . Since  $m' > m$ , and  $u' > u > \sigma q_o$ , it follows that  $u' - \sigma(m'q_o + (1 - m')\bar{q}_o) \geq 0$  if  $u - \sigma(mq_o + (1 - m)\bar{q}_o) \geq 0$ , and thus the consumer buys when  $v - p_s = u'$ . Finally, note that if  $u' \geq \sigma\bar{q}_o$  and  $u' > \sigma q_o$ , then the consumer buys when  $v - p_s = u'$ .

Thus, if  $u - \sigma \bar{q}_o < 0$  but the consumer buys when  $v - p_s = u$ , then the consumer buys for all other  $v, p_s$  such that  $v - p_s > u$ .

Second, suppose that  $u - \sigma \bar{q}_o \geq 0$  and the consumer buys for this value of  $v - p_s = u$ . Then for the optimal  $m$  at that  $u$ ,  $u - \sigma(mq_o + (1 - m)\bar{q}_o) \geq 0$ . Now if  $u \geq \sigma q_o$ , then plainly the consumer buys for any  $u' > u$ . Suppose instead that  $u < \sigma q_o$ . The value of full information is  $B = \int_{\sigma\omega \geq u} (\sigma\omega - u)\mu(\omega)$ , which is decreasing in  $u$ . Consequently,  $m$  is decreasing in  $u$  for  $u \geq \sigma \bar{q}_o$ . This means that the optimal attention weight  $m'$  at  $u'$  is  $m' \leq m$ . Then since  $m' \leq m$ , it holds that  $u' - \sigma(m'q_o + (1 - m')\bar{q}_o) \geq 0$  if  $u - \sigma(mq_o + (1 - m)\bar{q}_o) \geq 0$ .  $\square$

Since the propensity to buy is deterministic and is increasing in  $u = v - p_s$ , Lemma B.1 then implies:

**Proposition B.4.** *Consumer behavior in the Gabaix (2014) model of attention adjustment can be represented by a revealed valuation weight model in which the consumer chooses to buy if and only if  $v - p_s - \theta p_o \geq 0$  for  $\theta \in \mathbb{R}$ .*

We next consider comparative statics on  $\lambda$  and  $\sigma$ .

**Proposition B.5.** *In the revealed valuation weight representation,  $\theta = 1$  if  $\bar{q}_o = q_o$ . The relative misreaction  $|1 - \theta|$  is increasing in  $\lambda$  and is decreasing in  $\sigma$ , with  $\lim_{\lambda \rightarrow 0} |1 - \theta| = 0$  and  $\lim_{\sigma \rightarrow \infty} |1 - \theta| = 0$ .*

*Proof.* The case  $\bar{q}_o = q_o$  is immediate, since in this case  $q_o^* = q_o$  for all  $m$ .

Let  $m(u)$  be the optimal  $m$  chosen when  $v - p_s = u$ . Note that since  $B$  is continuous in  $u$ ,  $m(u)$  is continuous in  $u$  as well. Define  $u^\dagger$  to be the smallest  $u$  such that  $u \geq m(u)\sigma q_o + (1 - m(u))\sigma \bar{q}_o$ . Continuity implies that  $u^\dagger$  must satisfy

$$u^\dagger = m(u^\dagger)\sigma q_o + (1 - m(u^\dagger))\sigma \bar{q}_o \quad (17)$$

Recall that Lemma B.4 implies that there is a unique  $u^\dagger$  satisfying this equation.

Then

$$\theta = \frac{\sigma q_o}{u^\dagger} = \frac{q_o}{m(u^\dagger)q_o + (1 - m(u^\dagger))\bar{q}_o} \quad (18)$$

Note that  $\theta$  is a function of  $m$  and  $u^\dagger$  only, and does not directly depend on stakes. The combination of (17) and (18) imply that  $m(u)$  is decreasing in  $\lambda$  and increasing in  $\sigma$  for all  $u$ .

Consider first the case in which  $\bar{q}_o < q_o$ . The case  $\bar{q}_o > q_o$  follows analogously. Since  $m$  is decreasing in  $\lambda$  for all  $u$ , the assumption  $q_o > \bar{q}_o$  implies that  $q^* = mq_o + (1 - m)\bar{q}_o$  is decreasing in  $\lambda$  for all values of  $u$ . Consequently, the solution  $u^\dagger$  to equation (17) decreases in  $\lambda$ , and thus  $\theta$  must be increasing in  $\lambda$  as well. Finally, since  $m \rightarrow 1$  as  $\lambda \rightarrow 0$ , it follows that  $\lim_{\lambda \rightarrow 0} \theta = 1$ .

Next, consider the impact of increasing stakes from  $\sigma$  to  $\sigma' > \sigma$ . Let  $B(\sigma, u)$  denote the value of acquiring full information at stakes  $\sigma$  and transparent surplus  $v - p_s = u$ . Now for  $u' = (\sigma'/\sigma)u$ , and  $u > \sigma \bar{q}_o$

$$B(\sigma', u') = \int_{\omega \geq u'/\sigma'} (\sigma'\omega - u')\mu(\omega) = \frac{\sigma'}{\sigma} \int_{\omega} (\sigma\omega - u)\mu(\omega) = \frac{\sigma'}{\sigma} B(\sigma, u) \quad (19)$$

Since the perceived benefit of increasing  $m$  is linear in  $B$ , equation (19) above implies that increasing stakes to  $\sigma'$  has the same impact on  $m$  as decreasing attention costs from  $\lambda m^\alpha$  to  $\frac{\sigma'}{\sigma} \lambda m^\alpha$ . Thus, since  $m$  is decreasing in  $\lambda$ , it must be increasing in stakes  $\sigma$ .  $\square$

Finally, we work out a simple comparative static on prior beliefs that complements the comparative static in the body of the paper about how prior perceptions affect the revealed valuation weights  $\theta$ . We show that for a family of distributions of prior beliefs indexed by the mean and the variance, the revealed valuation weight  $\theta$  will be increasing in the mean and in the variance.

**Proposition B.6.** *Suppose that prior beliefs are given by the random variable  $d + l\varepsilon$ , where  $\varepsilon$  is a mean-zero random variable. Then the revealed valuation weight  $\theta$  is decreasing in  $d$ , and the relative misreaction  $|1 - \theta|$  is decreasing in  $l$ .*

*Proof.* Note that  $\bar{q}_o$  is constant in  $l$ , and thus increasing  $l$  cannot change behavior when  $m = 0$ . Consequently,  $B$  is proportional to  $l$ , and thus  $m$  is increasing in  $l$  as well. By the reasoning in the proof of Proposition B.5, this implies that  $|1 - \theta|$  is decreasing in  $l$ .

Next, we show that if a consumer with prior  $(d, l)$  buys when  $v - p_s = u$ , then a consumer with prior  $(d - \delta, l)$  will also buy when  $v - p_s = u$ . This will establish that  $\theta(d - \delta, l) \geq \theta(d, l)$  by Lemma B.1.

Consider first the case in which  $u - \sigma\bar{q}_o(d, l) < 0$ , so that the consumer does not buy when  $m = 0$ , but buys at the optimal  $m$  because  $u > \sigma q_o$ . Now for  $\delta$  such that  $u - \sigma\bar{q}_o(d, l) + \delta < 0$ , the consumer with prior  $(d - \delta, l)$  will also not buy when  $m = 0$ . But because  $B(u, d + \delta, l) > B(u, d, l)$  by the same reasoning as in the proof of Lemma B.4, the consumer with prior  $(d - \delta, l)$  will choose a higher  $m$ . Now since  $\sigma q_o < u < \sigma\bar{q}_o(d, l)$ , it follows that  $q_o < \bar{q}_o(d, l)$  and thus

$$\begin{aligned} m(d, l)q_o + (1 - m(\delta, l))\bar{q}_o(d, l) &\geq m(d - \delta, l)q_o + (1 - m(d - \delta, l))\bar{q}_o(d, l) \\ &> m(d - \delta, l)q_o + (1 - m(d - \delta, l))\bar{q}_o(d - \delta, l) \end{aligned}$$

Consequently, the consumer with prior  $(d - \delta, l)$  also buys.

Next, consider the case in which  $u - \sigma\bar{q}_o(d, l) > 0$  and the consumer buys for this value of  $v - p_s = u$ . Then for the optimal  $m$  at that  $u$ ,  $u - \sigma(m(d, l)q_o + (1 - m(d, l))\bar{q}_o(d, l)) \geq 0$ . Now if  $u \geq q_o$ , then plainly the consumer buys at prior  $(d - \delta, l)$  since  $\bar{q}_o(d - \delta, l) = \bar{q}_o(d, l) - \delta$ . Suppose instead that  $u < \sigma q_o$ , which also implies that  $q_o > \bar{q}_o$ . Then  $B(u, d + \delta, l) > B(u, d, l)$  by the same reasoning as in the proof of Lemma B.4. Consequently,  $m(d - \delta, l) \leq m(d, l)$ . Thus

$$\begin{aligned} m(d, l)q_o + (1 - m(\delta, l))\bar{q}_o(d, l) &\geq m(d - \delta, l)q_o + (1 - m(d - \delta, l))\bar{q}_o(d, l) \\ &> m(d - \delta, l)q_o + (1 - m(d - \delta, l))\bar{q}_o(d - \delta, l) \end{aligned}$$

which implies that the consumer with prior  $(d - \delta, l)$  also buys. □

#### B.4 Proof of Proposition A.1

Proposition B.1 establishes that the model has a revealed valuation weight representation. Proposition B.3 establishes that  $\lim_{\lambda \rightarrow 0} |1 - \theta| = 0$  and  $\lim_{\sigma \rightarrow \infty} |1 - \theta| = 0$ . This proves the first part of the proposition.

We now move on to the second statement. Set  $u = v - p_s$ . To characterize the model, begin by noting that Lemma 1 of Matejka and McKay (2015) implies that it is optimal for consumers to only choose at most

two different posteriors,  $\rho_0$  and  $\rho_1$ , such that  $b(\rho_0) = 0$  and  $b(\rho_1) = 1$ . Now Proposition 2 of Caplin et al. (2019) implies that the distribution  $\pi$  is optimal if and only if (i)  $Q\rho_1 + (1 - Q)\rho_0 = r$ , where  $Q$  is the ex-ante expected probability of buying, and (ii)

$$\begin{aligned}\frac{\rho_1}{\rho_0} &\leq e^{\frac{u-t}{\lambda}} \\ \frac{1-\rho_1}{1-\rho_0} &\leq e^{\frac{u-\hat{t}}{\lambda}}\end{aligned}$$

with equality in both equations when buying and not buying are ex-ante expected with positive probability. The constraint  $Q\rho_1 + (1 - Q)\rho_0 = r$  implies the constraints  $\rho_1 \leq r$  and  $\rho_0 \geq r$ .

When the equalities hold, we have a system of two equations and two unknowns given by

$$\begin{aligned}\rho_1 &= \rho_0 e^{\frac{u-t}{\lambda}} \\ 1 - \rho_1 &= (1 - \rho_0) e^{\frac{u-\hat{t}}{\lambda}}\end{aligned}$$

Plugging the first into the second gives  $(1 - \rho_0 e^{\frac{u-t}{\lambda}}) = (1 - \rho_0) e^{\frac{u-\hat{t}}{\lambda}}$ , or  $\rho_0 \left( e^{\frac{u-\hat{t}}{\lambda}} - e^{\frac{u-t}{\lambda}} \right) = e^{\frac{u-\hat{t}}{\lambda}} - 1$ , from which it follows that

$$\rho_0 = \frac{e^{\frac{u-\hat{t}}{\lambda}} - 1}{e^{\frac{u-\hat{t}}{\lambda}} - e^{\frac{u-t}{\lambda}}} \quad (20)$$

We then have

$$\begin{aligned}Q &= \frac{\rho_0 - r}{\rho_0 - \rho_1} \\ &= \frac{\rho_0 - r}{\rho_0(1 - e^{\frac{u-t}{\lambda}})} \\ &= \frac{1}{1 - e^{\frac{u-t}{\lambda}}} - \frac{r}{\rho_0(1 - e^{\frac{u-t}{\lambda}})} \\ &= \frac{1}{1 - e^{\frac{u-t}{\lambda}}} (1 - r/\rho_0)\end{aligned}$$

Now the ex-post probability of buying, by Bayes' rule, is

$$\begin{aligned}Pr(buy|q = t) &= \frac{Pr(q|\rho = \rho_1)Pr(buy)}{Pr(q)} \\ &= \frac{\rho_1 Q}{r} \\ &= \frac{\rho_0}{r} \frac{e^{\frac{u-t}{\lambda}}}{1 - e^{\frac{u-t}{\lambda}}} - \frac{e^{\frac{u-t}{\lambda}}}{1 - e^{\frac{u-t}{\lambda}}} \\ &= \frac{1}{e^{\frac{t-u}{\lambda}} - 1} \left( \frac{\rho_0}{r} - 1 \right)\end{aligned}$$

To consider comparative statics, first consider comparative statics on  $\rho_0$ . An alternative formulation is

$$\rho_0 = \frac{1 - e^{-\frac{\hat{t}-u}{\lambda}}}{1 - e^{-\frac{\hat{t}-t}{\lambda}}} \quad (21)$$

Now clearly  $\rho_0$  is increasing in  $u$ ; in general, the numerator goes from 0 for  $u = \hat{t}$  to 1 for  $u = \infty$ . Since the denominator is constant in  $u$ ,  $\rho_0$  is increasing in  $u$ . Next, we see that  $\rho_0$  is decreasing in  $\hat{t}$  from the formulation in equation (20), since  $e^{-\frac{u-t}{\lambda}} < 1$  and  $e^{-\frac{u-\hat{t}}{\lambda}} > 1$  but is decreasing in  $\hat{t}$ .<sup>35</sup> Finally,  $\rho$  is constant in  $r$ .

Now for comparative statics on  $Pr(buy|q = t)$ , note that  $\frac{1}{e^{-\frac{t-s}{\lambda}} - 1} > 0$  is increasing in  $u$ , and thus the probability is increasing in  $u$ . Next, since  $\rho_0$  is constant in  $r$ , the probability of buying is decreasing in  $r$ . And since  $\rho_0$  is decreasing in  $\hat{t}$ , we also see that the ex-post probability of buying is decreasing in  $\hat{t}$ .

The boundary conditions must be such that in general  $Q = \min\left(\max\left(\frac{\rho_0 - r}{\rho_0(1 - e^{-\frac{u-t}{\lambda}})}, 0\right), 1\right)$ , with  $\rho_0 = \rho_1$  if  $Q$  is not interior. It can be shown that there exist  $\underline{u}$  and  $\bar{u}$  such that  $Q = 0$  if  $u < \underline{u}$  and  $Q = 1$  if  $u > \bar{u}$ . We can show that the same comparative statics for  $r$  and  $\hat{t}$  apply to  $\underline{u}$  and  $\bar{u}$ . Intuitively, the higher are  $r$  and  $\hat{t}$ , the higher are  $\underline{u}$  and  $\bar{u}$ , since buying the good becomes only less advantageous. Formally, this is because  $Q$  is increasing in  $u$  but is decreasing in  $r$  and  $\hat{t}$ . Thus if  $r$  and  $\hat{t}$  get bigger, and  $Q$  is fixed at either 0 or 1, then the values of  $u$  have to be bigger to compensate.

## B.5 Proof of Proposition A.2

*Proof.* Proposition B.4 establishes that the model has a revealed valuation weight representation. Proposition B.5 establishes that the relative misreaction  $|1 - \theta|$  is increasing in  $\lambda$  and is decreasing in  $\sigma$ , with  $\lim_{\lambda \rightarrow 0} |1 - \theta| = 0$  and  $\lim_{\sigma \rightarrow \infty} |1 - \theta| = 0$ .

We now need to show decreasing  $\bar{q}_o$  through changes in  $r$  or  $\hat{t}$  cannot lead a consumer to go from buying to not buying. That is, the likelihood of buying is decreasing in  $\hat{q}_o$ . Combined with Lemma B.1, and the fact that the revealed valuation weight representation has  $\theta = 1$  when  $\hat{t} = t$ , this will imply the remaining statement of the proposition.

Case 1:  $t < u < \hat{t}$  and  $u - \sigma\bar{q}_o > 0$ . In this case  $u - \sigma(mt + (1 - m)\bar{q}_o) \geq 0$  for all  $m \in [0, 1]$ . Decreasing  $\hat{q}_o$  by either decreasing  $r$  or  $\hat{t}$  does not change that inequality.

Case 2:  $t < u < \hat{t}$  and  $u - \sigma\bar{q}_o < 0$ . The value of full information in this case is  $B = r(u - \sigma t)$ . If the consumer buys at the optimal  $m$  at these parameters, then  $u - \sigma(mt + (1 - m)\bar{q}_o) \geq 0$  by definition, which is possible only if  $u - \sigma t > 0$ . In this case, increasing  $r$  increases  $B$  and consequently the chosen  $m$ , and it decreases  $\bar{q}_o$ . Thus the propensity to buy increases in  $r$  when  $t < u < \hat{t}$ . Moreover, since  $B$  is not a function of  $\hat{t}$  when  $u - \sigma\bar{q}_o < 0$ , increasing  $\hat{t}$  has no impact on the consumer's propensity to buy in this region.

Case 3:  $\hat{t} < u < t$  and  $u - \sigma\bar{q}_o < 0$ . In this case  $u - \sigma(mt + (1 - m)\bar{q}_o) < 0$  for all  $m \in [0, 1]$ . The consumer does not buy for all parameters  $r$  and  $\hat{t}$  satisfying these conditions.

Case 4:  $\hat{t} < u < t$  and  $u - \sigma\bar{q}_o > 0$ . In this case,  $B = r(\sigma t - u)$ . If the consumer buys at the optimal  $m$  at these parameters, then  $u - \sigma(mt + (1 - m)\bar{q}_o) \geq 0$  by definition. Decreasing  $\bar{q}_o$  by decreasing  $r$  decreases  $B$  and thus decreases the optimal  $m$ . Since  $t > \bar{q}_o$ , decreasing  $r$  thus decreases  $mt + (1 - m)\bar{q}_o$ , and thus increases the propensity to buy. And since  $B$  is constant in  $\hat{t}$ , it is then mechanical that decreasing  $\hat{t}$  decreases

<sup>35</sup>For a function  $f(x) = \frac{x-1}{x-a}$  for  $a < 1$ , the derivative in  $x$  is  $f'(x) = \frac{(x-a) - (x-1)}{(x-a)^2} > 0$ . Thus  $\rho_0$  is monotone in  $e^{-\frac{u-t}{\lambda}}$ .



$mt + (1 - m)\bar{q}_o$ , and thus increases the propensity to buy.  $\square$

## C Proofs of Propositions in the body of the paper

### C.1 Proof of Proposition 1

*Proof.* Let  $E[X_i|Y] = \alpha(Y)$ . By the law of iterated expectations, and the conditional independence assumption that  $E[X_1X_2|Y] = E[X_1|Y]E[X_2|Y]$ ,

$$\begin{aligned} Cov[X_1, X_2] &= E[X_1X_2] - E[X_1]E[X_2] \\ &= E[E[X_1X_2|Y]] - E[E[X_1|Y]]E[E[X_2|Y]] \\ &= E[\alpha(Y)^2] - E[\alpha(Y)]^2 \\ &= Var[\alpha(Y)] \end{aligned}$$

Again by the law of iterated expectations,

$$\begin{aligned} Cov[Y, X_i] &= E[YX_i] - E[Y]E[X_i] \\ &= E[E[YX_i|Y]] - E[Y]E[E[X_i|Y]] \\ &= E[Y\alpha(Y)] - E[Y]E[\alpha(Y)] \\ &= Cov[Y, \alpha(Y)] \end{aligned}$$

The first statement of the proposition is therefore equivalent to

$$Var[Y]Var[\alpha(Y)] \geq Cov[Y, \alpha(Y)]^2,$$

which holds by the Cauchy-Schwarz inequality. More generally, if  $E[X_1X_2|Y] \geq E[X_1|Y]E[X_2|Y]$ , meaning that the two proxies are correlated conditional on  $Y$ , then  $Cov[X_1, X_2] \geq Var[\alpha(Y)]$  and the statement of the proposition still holds.

The second statement follows by the Bhatia-Davis inequality:  $(\bar{Y} - E[Y])(E[Y] - \underline{Y}) \geq Var[Y]$ .

To show that both inequalities are tight, suppose that  $Y$  takes on the values  $\underline{Y}$  and  $\bar{Y}$  only, with  $a = Pr(Y = \bar{Y})$ . Since  $\alpha(Y)$  must trivially be a linear function of  $Y$  when  $Y$  has binary support, and since the Cauchy-Schwarz inequality reduces to an equality when one random variable is a linear transformation of the other, this implies  $Var[Y]Var[\alpha(Y)] = Cov[Y, \alpha(Y)]^2$ . Moreover,

$$\begin{aligned} Var[Y] &= a(\bar{Y} - a\bar{Y} - (1 - a)\underline{Y})^2 + (1 - a)(\underline{Y} - a\bar{Y} - (1 - a)\underline{Y})^2 \\ &= a(1 - a)^2(\bar{Y} - \underline{Y})^2 + (1 - a)a^2(\bar{Y} - \underline{Y})^2 \\ &= a(1 - a)(\bar{Y} - \underline{Y})^2 \end{aligned}$$

At the same time,

$$\begin{aligned} (\bar{Y} - E[Y])(E[Y] - \underline{Y}) &= (\bar{Y} - a\bar{Y} - (1-a)\underline{Y})(a\bar{Y} + (1-a)\underline{Y} - \underline{Y}) \\ &= (1-a)(\bar{Y} - \underline{Y})a(\bar{Y} - \underline{Y}), \end{aligned}$$

which shows that  $(\bar{Y} - E[Y])(E[Y] - \underline{Y}) = \text{Var}[Y]$  for a distribution with binary support.  $\square$

## C.2 Proof of Proposition 2

We start with the more general statement.

**Proposition C.1.** *Let  $Y$  be a random variable supported on  $[\underline{Y}, \bar{Y}]$ . Then*

$$\Pr(Y > y) \geq \frac{E[(Y - \underline{Y})^2] - (y - \underline{Y})E[Y - \underline{Y}]}{(\bar{Y} - y)(\bar{Y} - \underline{Y})} \quad (22)$$

$$\Pr(Y < y) \geq \frac{E[(\bar{Y} - Y)^2] - (\bar{Y} - y)E[\bar{Y} - Y]}{(y - \underline{Y})(\bar{Y} - \underline{Y})} \quad (23)$$

and both bounds are tight.

*Proof.* For shorthand, set  $\alpha = \Pr(Y > y)$ . Suppose first that  $\underline{Y} = 0$ . Now for  $y \in [\underline{Y}, \bar{Y}]$ :

$$\begin{aligned} E[(Y - \underline{Y})^2] &= (1 - \alpha)E[(Y - \underline{Y})^2 | Y \leq y] + \alpha E[(Y - \underline{Y})^2 | Y > y] \\ &\leq (1 - \alpha)(y - \underline{Y})E[Y - \underline{Y} | Y \leq y] + \alpha(\bar{Y} - \underline{Y})E[Y - \underline{Y} | Y > y] \\ &= (1 - \alpha)(y - \underline{Y})E[Y - \underline{Y} | Y \leq y] + \alpha(y - \underline{Y})E[Y - \underline{Y} | Y > y] + \alpha(\bar{Y} - y)E[Y - \underline{Y} | Y > y] \\ &= (y - \underline{Y})E[Y - \underline{Y}] + \alpha(\bar{Y} - y)E[Y - \underline{Y} | Y > y] \\ &\leq (y - \underline{Y})E[Y - \underline{Y}] + \alpha(\bar{Y} - y)(\bar{Y} - \underline{Y}) \end{aligned}$$

Consequently,

$$\alpha \geq \frac{E[(Y - \underline{Y})^2] - (y - \underline{Y})E[Y - \underline{Y}]}{(\bar{Y} - y)(\bar{Y} - \underline{Y})}$$

Similarly, for shorthand, set  $\beta = \Pr(Y < y)$ , then for  $y > \underline{Y}$ ,

$$\begin{aligned} E[(\bar{Y} - Y)^2] &= (1 - \beta)E[(\bar{Y} - Y)^2 | Y \geq y] + \beta E[(\bar{Y} - Y)^2 | Y < y] \\ &\leq (1 - \beta)(\bar{Y} - y)E[(\bar{Y} - Y) | Y \geq y] + \beta(\bar{Y} - \underline{Y})E[(\bar{Y} - Y) | Y < y] \\ &= (1 - \beta)(\bar{Y} - y)E[(\bar{Y} - Y) | Y \geq y] + \beta(\bar{Y} - y)E[(\bar{Y} - Y) | Y < y] + \beta(y - \underline{Y})E[(\bar{Y} - Y) | Y < y] \\ &= (\bar{Y} - y)E[\bar{Y} - Y] + \beta(y - \underline{Y})E[(\bar{Y} - Y) | Y < y] \\ &\leq (\bar{Y} - y)E[\bar{Y} - Y] + \beta(y - \underline{Y})(\bar{Y} - \underline{Y}) \end{aligned}$$

Consequently,

$$\beta \geq \frac{E[(\bar{Y} - Y)^2] - (\bar{Y} - y)E[\bar{Y} - Y]}{(y - \underline{Y})(\bar{Y} - \underline{Y})}$$

Both bounds are tight. For the first one, consider a random variable that puts weight  $\alpha$  on  $Y = \bar{Y}$ , weight  $\beta$  on  $Y = y$ , and weight  $1 - \alpha - \beta$  on  $Y = \underline{Y}$ . Then

$$\begin{aligned} E[(Y - \underline{Y})^2] - (y - \underline{Y})E[Y - \underline{Y}] &= \alpha(\bar{Y} - \underline{Y})^2 + \beta(y - \underline{Y})^2 \\ &\quad - (y - \underline{Y})[\alpha(\bar{Y} - \underline{Y}) + \beta(y - \underline{Y})] \\ &= \alpha(\bar{Y} - \underline{Y})^2 - \alpha(y - \underline{Y})(\bar{Y} - \underline{Y}) \\ &= \alpha(\bar{Y} - \underline{Y})(\bar{Y} - \underline{Y} - y + \underline{Y}) \\ &= \alpha(\bar{Y} - \underline{Y})(\bar{Y} - y) \end{aligned}$$

and thus

$$\frac{E[(Y - \underline{Y})^2] - (y - \underline{Y})E[Y - \underline{Y}]}{(y - \underline{Y})(\bar{Y} - \underline{Y})} = \alpha.$$

Similarly, for a distribution that places weight  $\beta$  on  $Y = \underline{Y}$ , weight  $\alpha$  on  $Y = y$ , and weight  $1 - \alpha - \beta$  on  $Y = \bar{Y}$ ,

$$\begin{aligned} E[(\bar{Y} - Y)^2] - (\bar{Y} - y)E[\bar{Y} - Y] &= \beta(\bar{Y} - \underline{Y})^2 + \alpha(\bar{Y} - y)^2 \\ &\quad - (\bar{Y} - y)[\beta(\bar{Y} - \underline{Y}) + \alpha(\bar{Y} - y)] \\ &= \beta(\bar{Y} - \underline{Y})^2 - \beta(\bar{Y} - y)(\bar{Y} - \underline{Y}) \\ &= \beta(\bar{Y} - \underline{Y})(\bar{Y} - \underline{Y} - \bar{Y} + y) \\ &= \beta(\bar{Y} - \underline{Y})(y - \underline{Y}) \end{aligned}$$

from which the conclusion follows. □

We obtain Proposition 2 as a corollary. When  $\underline{Y} = 0$  and  $y = 1$ , equation (22) translates to

$$Pr(Y > 1) \geq \frac{E[Y^2] - E[Y]}{\bar{Y}(\bar{Y} - 1)}.$$

When  $\bar{Y} = 1$  and  $y = 0$ , equation (23) translates to

$$\begin{aligned} Pr(Y < 0) &\geq \frac{E[(1 - Y)^2] - E[1 - Y]}{(-\underline{Y})(1 - \underline{Y})} \\ &= \frac{E[Y^2] - 2E[Y] + 1 - (1 - E[Y])}{-\underline{Y}(1 - \underline{Y})} \\ &= \frac{E[Y^2] - E[Y]}{\underline{Y}(\underline{Y} - 1)} \end{aligned}$$

## D Predictions 3 and 4 with heterogeneity in attention costs

We set  $\tilde{\theta}_i(\sigma)$  to be individual  $i$ 's expected revealed valuation weight at stakes  $\sigma$ . The realized weight  $\theta_i(\sigma) = \tilde{\theta}_i(\sigma) + \delta$ , where  $Cov[\delta, \tilde{\theta}_i] = 0$ . We set  $\eta_i(\sigma) = -\frac{d(1-\tilde{\theta}_i)}{d\sigma} \frac{\sigma}{1-\tilde{\theta}_i}$  to be the elasticity of misreaction with respect to the stakes  $\sigma$ . The costly attention models imply that  $\eta_i(\sigma) \geq 0$  for all consumers  $i$ . Our key question is when

$$Cov\left[\tilde{\theta}_i(\sigma), \frac{d\tilde{\theta}_i}{d\sigma}\right] = \frac{1}{\sigma} Cov[\tilde{\theta}_i(\sigma), \eta_i(\sigma)(1 - \tilde{\theta}_i(\sigma))]$$

is negative. To that end, first note that if  $\eta_i(\sigma) \perp \tilde{\theta}_i(\sigma)$  then

$$\frac{1}{\sigma} Cov[\tilde{\theta}_i(\sigma), \eta_i(\sigma)(1 - \tilde{\theta}_i(\sigma))] = -\frac{E[\eta_i(\sigma)]}{\sigma} Var[\tilde{\theta}_i(\sigma)] < 0$$

That is, if the elasticities of misreaction are independent of the valuation weights, then it is guaranteed that consumers with the highest valuation weights will on average increase those valuation weights the least.

The condition  $\eta_i(\sigma) \perp \tilde{\theta}_i(\sigma)$  is satisfied when, for example, a consumer's  $\tilde{\theta}_i$  can be approximated by  $\tilde{\theta}_i(\sigma) = (1 - A_i w(\sigma)) + \theta_d^i A_i w(\sigma)$ , where  $w(\sigma)$  is a function of stakes and  $A_i$  is a constant determining sensitivity to stakes, and  $\theta_d^i$  is the valuation weight that results in the absence of any attention. In this case,  $1 - \tilde{\theta}_i = A_i w(\sigma)(1 - \tilde{\theta}_d^i)$ , and

$$\frac{d\tilde{\theta}_i}{d\sigma} = A_i w'(\sigma)(\theta_d^i - 1) = -(1 - \tilde{\theta}_i)w'(\sigma)/w(\sigma)$$

which implies that  $\eta_i = -\sigma w'(\sigma)/w(\sigma)$ , homogeneous.

To examine the implications of correlated heterogeneity in  $\eta_i$ , we now suppose that the following linear approximation is valid:

$$\eta_i = a_0 + a_1 \tilde{\theta}_i + \varepsilon \tag{24}$$

where  $\varepsilon \perp \theta$  and  $E[\varepsilon] = 0$ . A positive  $a_1$  means that  $\eta_i$  is on average positively related to  $\theta_i$ , while a negative  $a_1$  means that  $\eta_i$  is negatively related to  $\theta_i$ .

For the remainder of this appendix, we will often omit writing  $\eta_i$  and  $\tilde{\theta}_i$  as functions of  $\sigma$  to economize on notation. We now have that

$$\begin{aligned} Cov[\theta_i, \eta_i(1 - \tilde{\theta}_i)] &= Cov[\theta_i, (a_0 + a_1 \tilde{\theta}_i)(1 - \tilde{\theta}_i)] \\ &= -(a_0 - a_1)Var[\tilde{\theta}_i] - a_1 Cov[\tilde{\theta}_i, \tilde{\theta}_i^2] \end{aligned} \tag{25}$$

Now suppose that, as in our data, that  $E[\eta_i(1 - \tilde{\theta}_i)] = \sigma \frac{dE[\tilde{\theta}_i]}{d\sigma} > 0$ . Multiplying (24) by  $1 - \tilde{\theta}_i$  and taking

expectations yields

$$\begin{aligned} 0 &< a_0 E[1 - \tilde{\theta}_i] + a_1 [\tilde{\theta}_i(1 - \tilde{\theta}_i)] \\ &= a_0(1 - E[\tilde{\theta}_i]) + a_1(E[\tilde{\theta}_i] - E[\tilde{\theta}_i^2]) \end{aligned}$$

This implies that if  $E[\tilde{\theta}_i^2] = \text{Var}[\tilde{\theta}_i] + E[\tilde{\theta}_i]^2 > 1$ , as in our data, then

$$a_0 > \frac{a_1(E[\tilde{\theta}_i^2] - E[\tilde{\theta}_i])}{(1 - E[\tilde{\theta}_i])} > a_1 \quad (26)$$

In other words, if there is large variance in  $\tilde{\theta}_i$ , then  $\eta_i$  and  $\tilde{\theta}_i$  cannot be too strongly positively related—else it would imply that decreasing stakes would decrease the average  $\tilde{\theta}_i$ .

Next, observe that given an upper bound  $\bar{\theta}$  on  $\tilde{\theta}_i$ ,

$$\begin{aligned} \text{Cov}[\tilde{\theta}_i, \tilde{\theta}_i^2] &= E[(\tilde{\theta}_i - E[\tilde{\theta}_i])(\tilde{\theta}_i^2 - E[\tilde{\theta}_i^2])] \\ &\leq E[(\bar{\theta} - E[\tilde{\theta}_i])(\tilde{\theta}_i^2 - E[\tilde{\theta}_i^2])] \\ &\leq (\bar{\theta} - E[\tilde{\theta}_i])\text{Var}[\tilde{\theta}_i] \end{aligned}$$

And because the restriction that  $\eta_i \geq 0$  implies that  $a_0 + \bar{\theta}a_1 \geq 0$ , we thus have that  $a_1 \geq -a_0/\bar{\theta}$  and thus that

$$\begin{aligned} -(a_0 - a_1)\text{Var}[\tilde{\theta}_i] - a_1\text{Cov}[\tilde{\theta}_i, \tilde{\theta}_i^2] &\leq -(a_0 - a_1)\text{Var}[\tilde{\theta}_i] + \frac{a_0}{\bar{\theta}}(\bar{\theta} - E[\tilde{\theta}_i])\text{Var}[\tilde{\theta}_i] \\ &< -(a_0 - a_1)\text{Var}[\tilde{\theta}_i] + a_0\text{Var}[\tilde{\theta}_i] \\ &= a_1\text{Var}[\tilde{\theta}_i] \end{aligned} \quad (27)$$

Putting (26) and (27) together to sign (25), we see that: (i) if  $a_1$  is positive, then both terms of (24) must be negative by (26). On the other hand, (ii) if  $a_1$  is negative, then the expression in (25) must be negative by (27).

### D.1 Simple example with binary attention costs

For concreteness, consider the simple in Section 2.2. In this example, if  $\theta_i < 1$  then either  $\frac{d\theta_i}{d\sigma} = 0$  or  $\frac{d\theta_i}{d\sigma} = \frac{\lambda_i}{\sigma^2 r} = (1 - \theta_i)/\sigma$ . Similarly, if  $\theta_i \geq 1$  then either  $\frac{d\theta_i}{d\sigma} = 0$  or  $\frac{d\theta_i}{d\sigma} = (1 - \theta_i)/\sigma$ . Thus the elasticity of misreaction is either  $\eta_i = 0$  or  $\eta_i = 1$ . Simple algebra reveals that

$$\begin{aligned} \text{Cov}[\theta_i, \eta_i(1 - \theta_i)] / \text{Pr}(\eta_i = 1) &= \text{Cov}[\theta_i, (1 - \theta_i) | \eta_i = 1] + E[(1 - \theta_i) | \eta_i = 1] (E[\theta_i | \eta_i = 1] - E[\theta_i]) \\ &= -\text{Var}[\theta_i | \eta_i = 1] + \text{Pr}(\eta_i = 1)(1 - \text{Pr}(\eta_i = 1))(E[\theta_i | \eta_i = 1] - E[\theta_i | \eta_i = 0]) \end{aligned}$$

Thus when  $\text{Pr}(\eta_i = 1)$  is close enough to 1—i.e., sufficiently many individuals are at least somewhat elastic to stakes—the covariance is negative.

## E Relation to sales tax literature

Table A.1 summarizes the prior literature on misreaction to sales taxes. We mark a cell with “Yes” if the prediction is confirmed, with “Unclear” if the evidence is inconclusive, and we leave the cell blank if the prediction is not tested. We include the following papers: Chetty, Looney, and Kroft (2009, **CK**); Goldin and Homonoff (2013, **GH**); Feldman and Ruffle (2015, **FR**); Taubinsky and Rees-Jones (2018, **TRJ**); Feldman, Goldin, and Homonoff (2018, **FGH**); Kroft et al. (2019, **KLLN**); Bradley and Feldman (2020; **BF**); and this paper (**MT**)

As summarized in the first row, many papers have documented that consumers misreact to sales taxes both in lab or field experiments (Chetty et al., 2009; Feldman and Ruffle, 2015; Taubinsky and Rees-Jones, 2018; Feldman et al., 2018), and in quasi-experimental analysis (Chetty et al., 2009; Goldin and Homonoff, 2013; Kroft et al., 2020; Bradley and Feldman, 2020).

Far fewer papers are able to document meaningful individual differences in misreaction. Goldin and Homonoff (2013) find that low-income consumers reduce demand for cigarettes when the sales tax increases, but that high-income consumers have no statistically significant change in demand. TRJ use self-reported attention to the tax to document variation in  $\theta$  and to estimate a lower bound on  $Var[\theta]$ ; however, their lower-bound is at least an order of magnitude too loose, as we comment on below.

In row 3 we review the evidence on incorrect beliefs as a source of misreaction. Chetty et al. (2009), Feldman and Ruffle (2015) and Taubinsky and Rees-Jones (2018) elicit beliefs about the tax rate after the experiment and find that approximately 70-75% participants have nearly accurate beliefs about the tax rate in the study.

Rows 4 and 5 summarize evidence on how misreaction changes as the stakes increase. As discussed in Section 2.4.1, we verify in our paper that average misreaction decreases both as tax rates increase and as the posted prices increase. TRJ find that average underreaction is lower in their high-tax treatment than in their standard-tax treatment, but they generally lack statistical power to find differences across posted prices, and can only detect a difference between small prices and all other prices in the high-tax-rate condition. Feldman and Ruffle (2015) provide an indirect test of the effect of changing prices in their Table 5, but caveat that there are potential confounds and thus “do not push these particular results too far.” This test is not a main focus of their paper. Like TRJ, Feldman et al. (2018) vary tax rates in a between-subject design. However, Feldman et al. (2018) report results that are statistically imprecise and consistent both with the null of exogenous attention as well as the results we report in this paper. The Feldman et al. (2018) is not set-up to identify  $E[\theta]$  or  $Var[\theta]$ , and otherwise features the same weaknesses as TRJ, which we discuss below.

Testing the predictions in rows 6-9 requires within-consumer variation in stake size. Existing empirical work has not used such a design, and thus our paper is the first to find empirical support for these predictions.

Table A.1: Summary of prior literature

Prediction	CLK	GH	FR	TRJ	FGH	KLLN	BF	MT
1 Consumers misreact to shrouded sales taxes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2 There are individual differences in misreaction	–	Yes	–	Yes	–	–	–	Yes
3 Most consumers know the tax rates	Yes	–	Yes	Yes	–	–	–	Yes
4 Avg. misreaction decreases as the posted price increases	–	–	Unclear	Unclear	–	–	–	Yes
5 Avg. misreaction decreases as the tax rate increases	–	–	–	Yes	Unclear	–	–	Yes
6 Individual differences persist across stakes (P2 in MT)	–	–	–	–	–	–	–	Yes
7 Consumers with largest $\theta$ increase it the least in response to larger stakes (P3 in MT)	–	–	–	–	–	–	–	Yes
8 Consumers who react the least to higher stakes have the highest levels of $\theta$ at all stakes (P4 in MT)	–	–	–	–	–	–	–	Yes
9 Consumers with $\theta > 1$ decrease their $\theta$ with higher stakes (P5 in MT)	–	–	–	–	–	–	–	Yes

## E.1 Detailed comparisons to TRJ

**Design.** Our experimental design shares two similarities with TRJ, and differs in all other respects. First, we use the same products and product descriptions, and second, we exogenously increase the tax rate to triple its standard size. Unlike TRJ we use simple “yes-no” buying decisions rather than a less natural Becker-DeGroot-Marschak (BDM) bidding mechanism, and we create experimental stores that consumers have to “enter.” The BDM mechanism generates additional complexity, and in particular requires subjects to divide by “one plus the tax rate” when computing the before-tax price they should be willing to pay. In contrast, natural shopping decisions, as well as decisions in our experiment, involve looking at the posted price and thinking about what the final post-tax price would be—a simpler operation that involves multiplication. In particular, the rules-of-thumb and attention strategies that people employ in their day-to-day shopping decisions are much more likely to be faithfully captured by the experimental design in this paper, rather than that of TRJ.

**Within-person variation in tax rates.** TRJ do not have an experimental design that varies tax rates within-individual because of their focus on normative implications of tax salience. Their experimental design only varies tax rates between-subjects. Thus their data cannot be used to test predictions 2-5 examined in our paper.

**Statistical power.** TRJ’s design affords significantly less power in both estimates of average  $\theta$  and in particular how it varies with posted prices. For example, as column 1 of TRJ’s Table 3 shows, TRJ have no evidence that attention varies by price in the standard tax condition. There is some evidence that  $\theta$  is increasing in price in the high tax condition, but TRJ can only detect a difference between small prices and all other prices. By contrast, our study allows us to relatively precisely trace out how  $\theta$  varies by price at a much more granular level.

The differences in statistical power are likely a consequences of two design differences. First, TRJ have

more “noise” in any given decision created by the apparently more confusing BDM mechanism. Second, TRJ do not fully randomize the order of all the tax conditions within-subject. Consequently, they must use a “control” arm to estimate the order effects, which decreases statistical power.

**Bounds on individual differences at a given tax rate.** TRJ’s normative focus leads them to study  $Var[\theta]$  at a given tax rate, for which they estimate a lower bound of 0.1. Although  $Var[\theta]$  is not a direct focus of our paper—we study *individual differences in attentional responses to changes in tax rates*—we use a better-suited experimental design and new econometric methods to estimate a new lower bound on the variance that is about an order of magnitude higher. Thus, our experimental design presents a substantial advance over TRJ even for the statistic that is a key focus of TRJ but not of our paper.

**Appropriate designs for applying our econometric methods.** TRJ create data that violates the key assumption underlying our approach to individual differences, stated in Section 5.1: that  $(\hat{\theta}_{ijk} \perp \hat{\theta}_{ij'k'}) | \theta_{ijk}$  when  $j' \neq j$ . Although relatively weak, the validity of this assumption does rely on an important design feature: that all decisions are presented in random order. In the absence of this design feature, “order effects” that, for instance, lead to declining valuations over time as in TRJ would lead to correlated measurement error and violate our assumption. Consequently, our new methods for bounding  $Var[\theta]$  are not applicable to experimental datasets such as those in TRJ or Feldman et al. (2018) that do not vary the order of tax environments.

**Ensuring comprehension of the experiment.** TRJ have three arms in their experiment, and screen out participants who fail the comprehension questions before starting the shopping decisions in each round. However, because some comprehension questions are harder than others, this leads to differentially selected samples in the three arms: 35% of the sample is screened out in the no-tax arm and 22% of the sample is screened out in the triple tax arm. Our fully within-subject design avoids this potential confound, generates higher comprehension rates, and utilizes more relevant tests of comprehension. Our higher comprehension rates are due to our experiment presenting participants with questions twice. Once before the decisions, when we review the rules for the different stores if participants answer the questions incorrectly, and once after participants are finished making decisions. We screen out participants if they fail to correctly report back the rules after they finish their decisions—a more relevant screener because what matters is whether participants had knowledge of the rules all the way through their last decisions.

## F Counterfactual demand curve construction

Formally, let  $p_n$  denote the  $n$ th lowest price on the price list. Recall that we constructed the price list such that  $p_1 = 4$  and  $p_n = 1.15 \cdot p_{n-1}$  for  $n > 1$ . We thus estimate the counterfactual demand  $\tilde{D}_{jB}(p_n)$  for store B at price  $p_n$  as  $\tilde{D}_{jB}(p_n) := \sum_i \left[ \frac{0.15 - \tau_i}{0.15} D_{ijA}(p_n) + \frac{\tau_i}{0.15} D_{ijA}(p_{n+1}) \right]$ , where  $\tau_{iB}$  is the tax rate faced by the person in store B, and  $D_{ijA}(p) \in \{0, 1\}$  is an indicator for whether the consumer bought the product at price  $p$  in store A.<sup>36</sup> For store C, if  $\tau_{iC} < 0.15$  we use the same interpolation as in the store B counterfactual demand; if  $\tau_{iC} > 0.15$ , we calculate  $\tilde{D}_{jC}(p_n) := \frac{0.30 - 3\tau_i}{0.15} D_{jA}(p_{n+1}) + \frac{3\tau_i - 0.15}{0.15} D_{jA}(p_{n+2})$ . To construct  $\tilde{D}_{jC}(p_9)$ , we use the self-reported maximum willingness to pay to see if individuals willing to purchase at price  $p_{10}$  would be

<sup>36</sup>All sales tax rates in our sample are less than 15%.



willing to purchase at price  $1.15p_{10}$ .

## G Interpreting coefficients in the probit regression

We have that person  $i$  chooses to buy product  $j$  in store  $k$ , with probability  $F\left(\frac{\mu_j - \log p - \theta_i \log(1 + \tau_{ik})}{\sigma_j}\right)$ , where  $F$  is the standard normal CDF. Let  $f$  denote the standard normal density function. Here we formally verify that

$$E_i F\left(\frac{\mu_j - \log p - \theta_i \log(1 + \tau)}{\sigma_j}\right) \approx F\left(\frac{\mu_j - \log p - E[\theta_i] \log(1 + \tau)}{\sigma_j}\right)$$

with negligible error terms. For shorthand, we set  $\alpha := \log(1 + \tau)$ . A first-order Taylor expansion around  $y := \frac{\mu_j - \log p}{\sigma_j} - \frac{E[\theta_i] \alpha}{\sigma_j}$  yields

$$\begin{aligned} E\left[F\left(\frac{\mu_j - \log p - \theta_i \alpha}{\sigma_j}\right)\right] &= F(y) + E[\theta_i - E[\theta_i]] f\left(x_j - \frac{E[\theta_i] \alpha}{\sigma_j}\right) + O(\alpha^2) \\ &= F(y) + O(\alpha^2) \end{aligned}$$

Thus, the estimated population  $\theta$  in our probit model corresponds to the average  $\theta$  up to terms of order  $\alpha^2$ . These are certainly negligible in store B. To more carefully assess the impact of second order-terms, we now compute a second-order Taylor expansion, around  $y := \frac{\mu_j - \log p}{\sigma_j} - \frac{E[\theta_i] \alpha}{\sigma_j}$ , using the fact that for a normal distribution,  $f'(x) = -xf(x)$ :

$$\begin{aligned} E\left[F\left(\frac{\mu_j - \log p - \theta_i \alpha}{\sigma_j}\right)\right] &= F(y) + E[\theta_i - E[\theta_i]] f(y) \\ &\quad + \frac{1}{2} E\left[\left(\frac{\theta_i \alpha - E[\theta_i] \alpha}{\sigma_j}\right)^2\right] f'(y) + O(\alpha^3) \\ &= F(y) - \frac{1}{2} y \alpha^2 \frac{\text{Var}[\theta_i]}{\sigma_j^2} f(y) + O(\alpha^3) \\ &= F\left(y - \frac{1}{2} y \alpha^2 \frac{\text{Var}[\theta_i]}{\sigma_j^2}\right) + O(\alpha^3) \\ &= F\left(\frac{\mu_j - \log p}{\sigma_j} - \left(\frac{\mu_j - \log p}{2\sigma_j^3} \alpha \text{Var}[\theta_i] + \frac{E[\theta_i]}{\sigma_k}\right) \alpha\right) + O(\alpha^3) \end{aligned}$$

If we instead assume that the probability is given by  $F\left(\frac{\mu_j - \log p}{\sigma_j} - \frac{E[\theta_i] \alpha}{\sigma_j}\right)$ , how much bias do we get from this model specification? The answer depends on the average value of  $x_k$ , which determine the extent to which introducing taxes leads to a lower probability of buying. Note that we can estimate  $1/\sigma_j$  and  $\mu_j/\sigma_j$  from the probit regression in which there are no taxes, which on average are given by 2.073 and 3.897, respectively. Using those estimates, we can find that the average value of  $\frac{\mu_j - \log p}{\sigma_j}$  is given by  $-0.24$ . This means that our representative population estimates produce slight underestimates of the

actual population average, and that the degree of underestimation is greater for triple taxes than for standard taxes. Under the conservative upper bound on  $\text{Var}[\theta|\alpha]$  of 1, this implies that the margin of error is about  $-0.24 \cdot (1/2) \cdot 2.073^2 \cdot E[\alpha] = -0.52E[\alpha]$ . For standard taxes, this gives a margin of error of about  $-.036$  and for triple taxes this gives a margin of error of about  $-0.101$ . When studying how a particular covariate affects  $E[\theta]$ , the margin of error is even smaller, since the difference in variances should be smaller than 1. If the covariate does not affect variances, then the margin of error vanishes to be of order three or higher.

One way of assessing whether our model produces estimates close to the average is to consider estimates  $\hat{\theta}_{pop|X}$  for a binary instrument  $X \in \{0, 1\}$ . If the probit model produces estimates close to the average, then we should have  $\hat{\theta}_{pop} = (1 - \text{Pr}(X = 1))\hat{\theta}_{pop|X=0} + \text{Pr}(X = 1)\hat{\theta}_{pop|X=1}$ . To the extent that we underestimate taxes significantly due to the variance, notice that because the average of variances of two distributions is lower than the variance of their mixture,<sup>37</sup> the average of the  $\theta$  estimates from two samples should be lower than our estimate of the overall average. We do not find this to be a large effect. For our binary proxies, we compare the estimates in tables 2 and 3 for the triple tax. Recall that the estimate of  $E[\theta]$  for the triple tax from the baseline regression is 0.79. When we average the two values in table 1 we get  $0.25 \times 1.20 + 0.75 \times 0.25 = 0.78$ . When we average the two values in table 2 we get  $0.25 \times .86 + 0.75 \times 0.76 = 0.785$ . These results suggest that there is not a significant bias.

Finally, note that the bias induced by the approximation works against our results on how  $\theta$  changes with the price. This is because  $\frac{\mu_j - \log p}{2\sigma_j^3}$  is decreasing in  $p$ , which dampens our findings about how  $E[\theta]$  varies with price.

## H Point estimates and confidence intervals for Figures 2a and 2b

Table A.2: Average revealed valuation weights in Figure 2a

Price cutoff	Avg. revealed val. wgt.: standard tax	95% CI	Avg. revealed val. wgt.: triple tax	95% CI
4.60	0.23	[0.10, 0.35]	0.40	[0.34, 0.47]
5.29	0.27	[0.12, 0.42]	0.55	[0.48, 0.63]
6.08	0.27	[0.11, 0.44]	0.64	[0.56, 0.71]
7.00	0.34	[0.17, 0.51]	0.72	[0.65, 0.80]
8.05	0.39	[0.22, 0.56]	0.77	[0.69, 0.85]
9.25	0.43	[0.26, 0.59]	0.80	[0.72, 0.87]
10.64	0.46	[0.30, 0.62]	0.81	[0.74, 0.88]
12.24	0.47	[0.31, 0.62]	0.80	[0.73, 0.87]
14.07	0.48	[0.32, 0.63]	0.79	[0.72, 0.86]

Table A.2 presents the estimates for  $E[\theta]$  and average tax owed displayed in figure 2a.  $\theta$  is defined as the revealed valuation weight that consumers place on the sales tax, with  $\theta = 0$  corresponding to complete neglect of the tax and  $\theta = 1$  corresponding to putting the same weight on the tax as on the salient posted price. Each price cutoff corresponds to a different posted price on the price list presented to consumers. The results are estimated using equation (4) for prices below the cutoff. Standard errors are clustered at the subject level.

<sup>37</sup>The variance of a mixture  $X$  of random variables  $X_i$  with weights  $w_i$  is given by  $E[(X - \mu)^2] = \sigma^2 = \sum_{i=1}^n w_i(\mu_i^2 + \sigma_i^2) - \mu^2$ .

Table A.3: Average revealed valuation weights in figure 2b

Bin	Avg. price	Avg. tax rate	Avg. tax owed	Avg. revealed val. wgt.	95% CI
1	4.30	7.24%	0.31	0.23	[0.11, 0.35]
2	5.69	7.24%	0.41	0.27	[0.05, 0.49]
3	7.52	7.24%	0.54	0.52	[0.32, 0.71]
4	9.95	7.24%	0.72	0.65	[0.44, 0.87]
5	13.15	7.24%	0.95	0.72	[0.40, 1.05]
6	4.30	21.72%	0.93	0.41	[0.35, 0.47]
7	5.69	21.72%	1.24	0.80	[0.69, 0.90]
8	7.52	21.72%	1.63	0.87	[0.77, 0.97]
9	9.95	21.72%	2.16	0.87	[0.75, 0.99]
10	13.15	21.72%	2.66	0.91	[0.73, 1.08]

Table A.3 presents store-specific estimates of  $E[\theta]$  by the average tax owed within each bin.  $\theta$  is defined as the revealed valuation weight that consumers place on the sales tax, with  $\theta = 0$  corresponding to complete neglect of the tax and  $\theta = 1$  corresponding to putting the same weight on the tax as on the salient posted price. For each tax environment—store B and store C—each bin is constructed by dividing the 10 prices in the experiment into 5 ordered pairs. The average tax owed is constructed by taking the average of the two prices in each bin, and multiplying it by the average tax rate in stores B and C, respectively. The estimating equation is an extension of equation (4), described in equation (5). Standard errors are clustered at the subject level.

## I Relationship between average revealed valuation weights and marginal utility of money

One potential concern with our estimation procedure in Section 4.3 is that the set of consumers on the margin at each price mechanically have different product valuations. If the valuation for the product is correlated with attention, this would confound our results about how average valuation weights covary with price.

In this appendix, we present additional evidence that our results are robust to this concern. Specifically, we utilize the split-sample techniques described in Section 5.1.1 to analyze whether, holding price constant, participants with lower marginal utility of money, and hence higher propensity to pay for the products in our experiment, are more attentive to taxes.

First, we use one product to divide consumers into two groups. The *low* marginal utility of money (MU) group consists of those with above-median values of willingness to pay in the no-tax environment and the *high* MU group consists of those with below-median values of willingness to pay. The intuition is that participants who have a higher valuation of the product were willing to forego more money to obtain the product, or equivalently have a lower marginal utility of money. We then estimate equation (5) *on the other two products* to estimate average valuation weights at each of the five price sets for the high and low MU groups. We then repeat this process for the other two products, and average the resulting estimates to obtain average valuation weights at each price pair  $P_n$ ,  $E[\theta_{ijk}|k = K, p \in P_n]$ , for both the low and high groups. To hold prices constant, we compute separate average valuation weights for each price-pair.

More concretely, we index each of the three products seen by each person with  $j \in \{1, 2, 3\}$ . First, we start with  $j = 1$  and we split the sample into two groups: those with  $p_{ijA}^*$  in the top 50% of the population and those with  $p_{ijA}^*$  in the bottom 50% of the population, where  $p_{ijA}^*$  denotes the WTP for product  $j$  in store

A.<sup>38</sup> We then use decisions in the *other two* products to estimate the average valuation weights  $E[\theta_{ijk}|k = K, LowMU_{i1}, p \in P_n, j \neq 1]$  for each price pair  $P_n$  using equation (5), where  $K \in \{B, C\}$  and  $LowMU_{i1}$  is defined as an indicator for the low MU group. We repeat the procedure twice using products 2 and 3 to generate  $LowMU_{i2}$  and  $LowMU_{i3}$ , and estimate  $E[\theta_{ijk}|k = K, LowMU_{i2}, p \in P_n, j \neq 2]$  and  $E[\theta_{ijk}|k = K, LowMU_{i3}, p \in P_n, j \neq 3]$ . Finally, we average the estimates from each of these three iterations to get an overall average estimate of  $\theta_{ijk}$  for those in the high and low groups:

$$E[\theta_{ijk}|k = K, LowMU_{ij}, p \in P_n,] = \frac{1}{3}(E[\theta_{ijk}|k = K, LowMU_{i1}, p \in P_n, j \neq 1] \\ + E[\theta_{ijk}|k = K, LowMU_{i2}, p \in P_n, j \neq 2] + E[\theta_{ijk}|k = K, LowMU_{i3}, p \in P_n, j \neq 3])$$

We then compute the average difference high valuation group and the low valuation group across all five price pairs  $P_n$ :

$$\frac{1}{10} \sum_{K \in \{B, C\}} \sum_{n=1}^5 (E[\theta_{ijk}|k = K, LowMU_{ij} = 1, p \in P_n,] - E[\theta_{ijk}|k = K, LowMU_{ij} = 0, p \in P_n,])$$

We compute confidence intervals using the percentile bootstrap, clustered at the individual level. This average difference is both small in magnitude and not statistically significant (0.11, 95% CI [-0.09, 0.28]).

To confirm we are separating participants by their marginal utility of money, we also check whether participants with above-median WTP for product 1 have higher WTP for products 2 and 3. Specifically, we estimate the following equations:

$$p_{ijA}^* = \alpha_{ij} + \beta^1 \cdot LowMU_{i1} + \varepsilon_{ij}, j \neq 1$$

$$p_{ijA}^* = \alpha_{ij} + \beta^2 \cdot LowMU_{i2} + \varepsilon_{ij}, j \neq 2$$

$$p_{ijA}^* = \alpha_{ij} + \beta^3 \cdot LowMU_{i3} + \varepsilon_{ij}, j \neq 3$$

In each regression, we exclude the product used to divide the sample into the low and high MU groups. We then average the  $\beta^j$  coefficients to obtain a single estimate  $\beta := \beta^1 + \beta^2 + \beta^3$ . The resulting coefficient is \$4.00 (95% CI [3.76, 4.24]), which implies that the low MU group has a \$4.00 higher WTP for any given product.

In summary, we find that participants who have an above-median WTP for one product have a \$4.00 higher WTP for the other two products, but this translates to only a 0.11 difference in average  $\theta$  at any given price.

<sup>38</sup>Section 5.1.1 details the methodology used to construct  $p_{ijA}^*$ .

## J Covariates of attention

### J.1 Local tax rate variation

We first divide the sample into those whose local tax rate is above 7.00%, the median in our sample (“high tax group”), and those below 7.00% (“low tax group”). We then run the regression in equation (4) separately for the above-median and below-median tax groups to create figures analogous to Figure 2a. We include state fixed effects to capture some of the geographic variation.

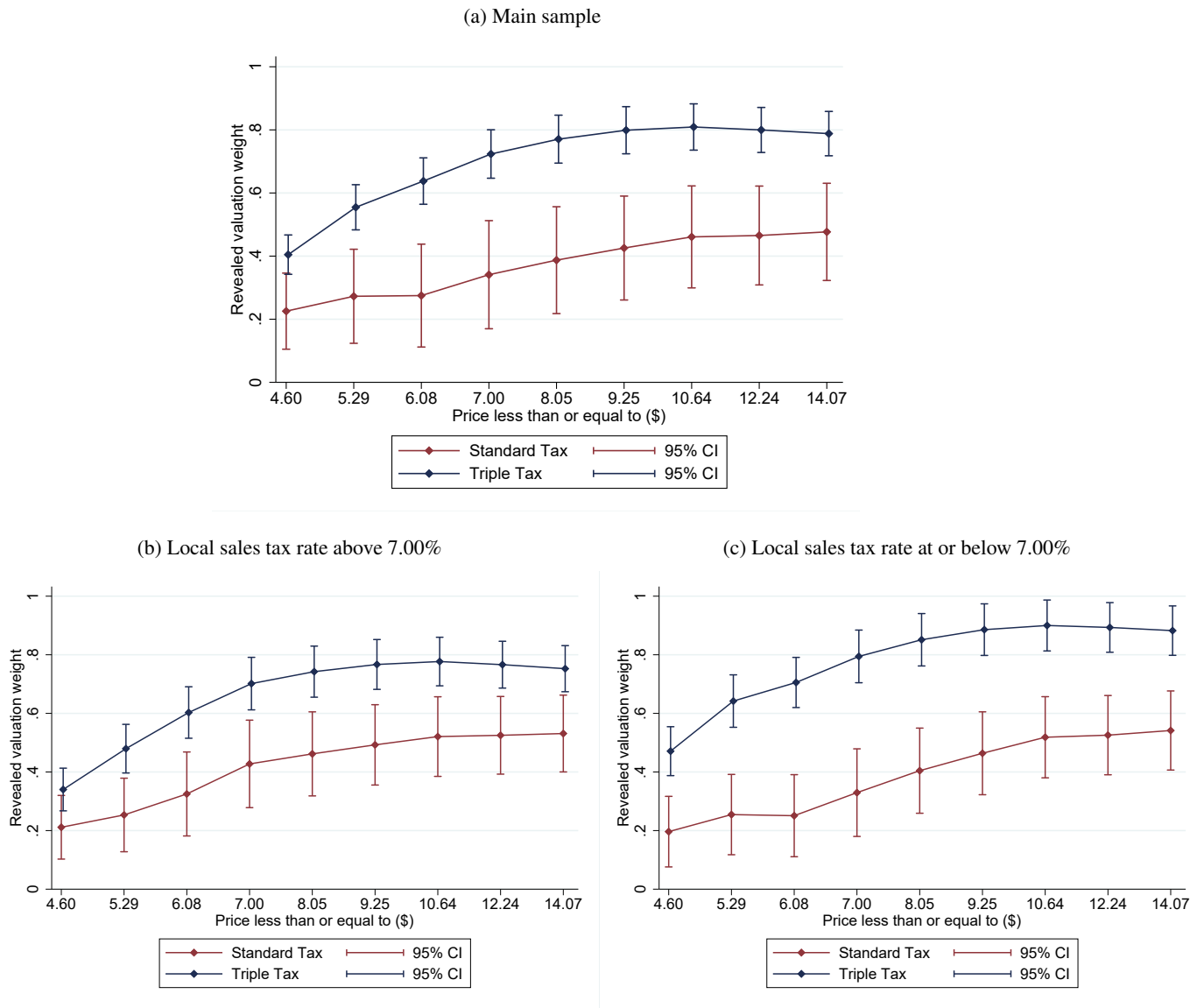
Figure A.1 presents the results. Panel (a) uses the main sample and is identical to Figure 2a. Panel (b) restricts to participants with a local sales tax rate above 7.00%, the median of the sample. Panel (c) restricts to participants with a local sales tax rate at or below 7.00%. The results provide some evidence that participants in high sales tax locations have lower revealed valuation weights than those from low sales tax locations.

Table A.4 presents store-specific estimates of average  $\theta$  by tax group using all prices. These estimates match the rightmost points of the series in Figure A.1. The third column presents the difference, which is statistically significant for store C.

We next move from splitting the sample at the median tax rate to dividing the sample into quartiles of local tax rates. The median tax in our sample is 7.00% and the interquartile range is 6.00%-8.15%. Table A.5 presents the results. The differences in tax rates are statistically significant for store C ( $p = 0.001$ ) and just outside the 10% significance level for store B ( $p = 0.105$ ). For both stores B and C, the lowest tax quartile sample has the highest average valuation weights, 0.62 and 0.86, respectively. We similarly see that the highest tax quartile group has the lowest average revealed valuation weights of 0.34 and 0.57, respectively.

Although this association between valuation weights and local tax rates appears to be opposite to the relationship seen in Figure 2a, we note that it is likely a confounded test of costly attention models because local tax variation could be related to a number of differences in geography, including consumers’ views and preferences about tax rates, or consumers’ attention to tax rates. For example, higher-tax rate jurisdictions tend to be more urban and in more liberal states, and the observed differences in average valuation weights may reflect sorting into urban versus rural jurisdictions. Moreover, to the extent that states and counties set their tax rates that follow some version of a standard inverse-elasticity rule, this will tend to lead to higher tax rates being set in places where consumers tend to be least attentive to taxes (a “reverse-causality” mechanism).

Figure A.1: Average revealed valuation weight for posted prices at or below a cutoff



Panels (b) and (c) of Figure A.1 recreate Figure 2a, restricting to participants above and below the median local sales tax rate respectively.

Table A.4: Average revealed valuation weights by tax group

	Standard	Triple
High tax group	0.53 [0.40, 0.66]	0.75 [0.67, 0.83]
Low tax group	0.56 [0.42, 0.70]	0.89 [0.81, 0.98]
Difference	-0.04 [-0.23, 0.16]	-0.14 [-0.26, -0.03]

Table A.1 presents estimates of store-specific estimates  $E[\theta]$  by tax group.  $\theta$  is defined as the revealed valuation weight that consumers place on the sales tax, with  $\theta = 0$  corresponding to complete neglect of the tax and  $\theta = 1$  corresponding to the equal weight of the tax and salient price. Individuals with a local sales tax rate above 7.00% are classified as high tax, and individuals with a local sales tax rate at or below 7.00% are classified as low tax. The results are estimated using equation (4), interacting the covariate with price and tax. Standard errors are clustered at the subject level.

Table A.5: Average revealed valuation weights by tax rate quartile

	Standard	Triple
Top quartile	0.62 [0.43, 0.81]	0.86 [0.74, 0.98]
Second quartile	0.45 [0.29, 0.62]	0.77 [0.67, 0.87]
Third quartile	0.52 [0.35, 0.70]	0.70 [0.60, 0.80]
Bottom quartile	0.34 [0.20, 0.48]	0.57 [0.48, 0.67]

Table A.5 presents store-specific estimates of  $E[\theta]$  by tax rate quartile.  $\theta$  is defined as the revealed valuation weight that consumers place on the sales tax, with  $\theta = 0$  corresponding to complete neglect of the tax and  $\theta = 1$  corresponding to the equal weight of the tax and salient price. The median income in our sample is 7.00% and the interquartile range is 6.00%-8.15%. The results are estimated using equation (4), interacting the covariate with price and tax. Standard errors are clustered at the subject level.

## J.2 Demographics

Here we analyze how revealed valuation weights vary according to observed demographics. We separately analyze the effects of political party, education, income, and beliefs.

**Political party:** Table A.6 presents average  $\theta$  estimates for self-identified Republicans (28.5% of our sample), Democrats (32.1% of our sample), and individuals with independent or other political beliefs (39.4% of our sample).<sup>39</sup> Republicans and Democrats have an average  $\theta_B$  of 0.52 and 0.51 respectively (95% CI for difference: [-0.39, 0.39]). Republicans have a slightly larger  $\theta_C$  than do Democrats in our sample (0.86 vs. 0.74), but the difference is not statistically significant (95% CI for difference: [-0.06, 0.29]).

**Education:** Table A.7 compares the average  $\theta$  estimates between college graduates (35.3% of our sample) and participants with no or some college experiences (64.7% of our sample; includes associate's degree recipients). College graduates have a slightly higher  $\theta_B$  (0.51 vs. 0.46), but the difference is not

<sup>39</sup>30.6% of participants self-identify as independent and 8.9% of participants self-identify as other.

statistically significant (95% CI for difference  $[-0.27, 0.38]$ ). Both education groups have the same estimate for  $\theta_C$  (0.79, 95% CI for difference  $[-0.14, 0.15]$ ). In Table A.8, we similarly split the sample into those with advanced degrees versus those without advanced degrees. We again do not detect statistically significant differences in average valuation weights between the groups.

**Income:** Table A.9 presents average  $\theta$  estimates for each income quartile. Individuals in the top income quartile have self-reported annual income above \$80000, in the second quartile from \$49000-\$80000, in the third quartile from \$25000-\$49000, and in the bottom quartile below \$25000.

All quartiles have average  $\theta_C$  point estimates in the 0.77-0.79 range. Individuals in the top three quartiles have average  $\theta_B$  estimates in the 0.39-0.52 range, while individuals in the bottom quartile have an average  $\theta_B$  of 0.52 (95% CI:  $[0.22, 0.81]$ ). A test of equivalence between the  $\theta$  estimates in all quartiles yields  $\chi^2 = 0.44$ , ( $p = 0.93$ ) for store B and  $\chi^2 = 0.51$ , ( $p = 0.92$ ) for store C.

**Beliefs:** Table A.10 presents average  $\theta$  estimates separately for (i) participants who exactly know their local tax rate (51.0% of our sample), (ii) participants who know their local tax rate with one percentage point but do not know it exactly (31.0% of our sample), and (iii) participants who do not know their tax rate within one percentage point (18.0% of our sample). The means of participants' estimates of their sales tax rates in these three groups are 7.08%, 7.46%, and 8.41%, respectively. Compared to participants who do not know their local tax rate within one percentage point, we see that participants with exact knowledge have higher average  $\theta_B$  (0.54 vs. 0.54, 95% CI for difference:  $[-0.21, 0.77]$ ) and  $\theta_C$  (0.85 vs. 0.61, 95% CI for difference:  $[0.03, 0.45]$ ), though these differences are only statistically significant in the triple-tax environment. The results provide some evidence that participants with less accurate knowledge about their local tax rates underreact to taxes more on average than do participants with more accurate knowledge about them.

Table A.6: Average revealed valuation weights by political party

	Standard	Triple
(1): Republicans	0.52 [0.25, 0.78]	0.86 [0.74, 0.99]
(2): Democrats	0.51 [0.23, 0.80]	0.74 [0.62, 0.87]
(3): Independent/Other	0.42 [0.18, 0.67]	0.77 [0.66, 0.89]
(4): (1) - (2)	0.00 [-0.39, 0.39]	0.12 [-0.06, 0.29]

Table A.6 presents store-specific estimates of  $E[\theta]$  by political party affiliation.  $\theta$  is defined as the revealed valuation weight that consumers place on the sales tax, with  $\theta = 0$  corresponding to complete neglect of the tax and  $\theta = 1$  corresponding to the equal weight of the tax and salient price. Individuals were asked to select which of independent, Republican, Democrat, or other best described their political party affiliation. The results are estimated using equation (4), interacting the covariate with price and tax. Standard errors are clustered at the subject level.



Table A.7: Average revealed valuation weights by education: college graduates versus not college graduates

	Standard	Triple
College graduate	0.51 [0.26, 0.77]	0.79 [0.68, 0.90]
Not college graduate	0.46 [0.27, 0.65]	0.79 [0.70, 0.88]
Difference	0.06 [-0.27, 0.38]	0.00 [-0.14, 0.15]

Table A.7 presents store-specific estimates of  $E[\theta]$  by education level.  $\theta$  is defined as the revealed valuation weight that consumers place on the sales tax, with  $\theta = 0$  corresponding to complete neglect of the tax and  $\theta = 1$  corresponding to the equal weight of the tax and salient price. Not college graduate includes participants with associate's degrees or with some years in college. The results are estimated using equation (4), interacting the covariate with price and tax. Standard errors are clustered at the subject level.

Table A.8: Average revealed valuation weights by education: graduate school versus no graduate school

	Standard	Triple
Graduate school	0.64 [0.13, 1.15]	0.77 [0.55, 0.99]
No graduate school	0.46 [0.30, 0.62]	0.79 [0.72, 0.86]
Difference	0.18 [-0.36, 0.71]	-0.02 [-0.25, 0.21]

Table A.8 presents store-specific estimates of  $E[\theta]$  by education level.  $\theta$  is defined as the revealed valuation weight that consumers place on the sales tax, with  $\theta = 0$  corresponding to complete neglect of the tax and  $\theta = 1$  corresponding to the equal weight of the tax and salient price. Graduate includes participants with master's degrees or more advanced degrees. The results are estimated using equation (4), interacting the covariate with price and tax. Standard errors are clustered at the subject level.

Table A.9: Average revealed valuation weights by income quartile

	Standard	Triple
Top quartile	0.39 [0.07, 0.71]	0.77 [0.62, 0.93]
Second quartile	0.50 [0.16, 0.84]	0.84 [0.68, 1.00]
Third quartile	0.52 [0.25, 0.79]	0.79 [0.67, 0.91]
Bottom quartile	0.52 [0.22, 0.81]	0.77 [0.64, 0.90]

Table A.9 presents store-specific estimates of  $E[\theta]$  by income quartile.  $\theta$  is defined as the revealed valuation weight that consumers place on the sales tax, with  $\theta = 0$  corresponding to complete neglect of the tax and  $\theta = 1$  corresponding to the equal weight of the tax and salient price. The median income in our sample is \$49,000 and the interquartile range is \$25,000-\$80,000. The results are estimated using equation (4), interacting the covariate with price and tax. Standard errors are clustered at the subject level.

Table A.10: Average revealed valuation weights by knowledge of local sales tax rate

	Standard	Triple
Exact knowledge	0.54 [0.33, 0.75]	0.85 [0.75, 0.95]
Belief error $\in (0\%, 1\%]$	0.48 [0.24, 0.73]	0.79 [0.68, 0.91]
Belief error $> 1\%$	0.27 [-0.17, 0.71]	0.61 [0.43, 0.79]

Table A.10 presents store-specific estimates of  $E[\theta]$  by participants' knowledge of their local tax rate.  $\theta$  is defined as the revealed valuation weight that consumers place on the sales tax, with  $\theta = 0$  corresponding to complete neglect of the tax and  $\theta = 1$  corresponding to the equal weight of the tax and salient price. The first row includes participants who know their local sales tax rate exactly, the second row includes participants who have an error in their beliefs of less than one percentage point, and the third row includes participants who do know their local sales within one percentage point. The results are estimated using equation (4), interacting the covariate with price and tax. Standard errors are clustered at the subject level.

### K Alternative construction of proxies for valuation weights

Figure A.2: Average revealed valuation weight by posted price

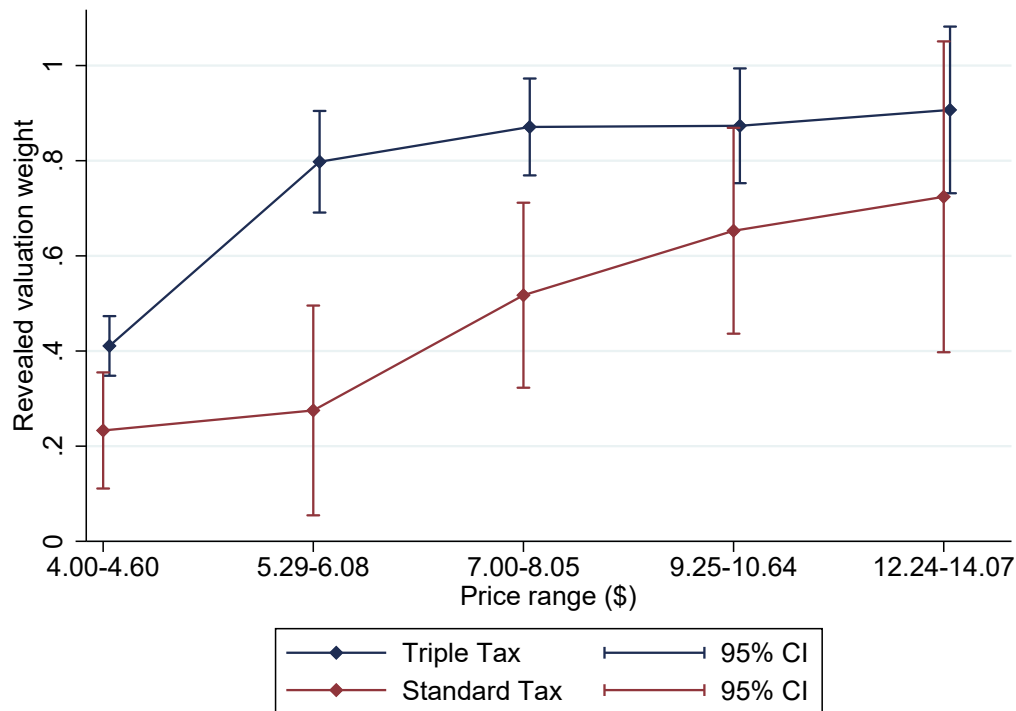


Figure A.2 presents store-specific estimates  $E[\theta]$  by the posted price.  $\theta$  is defined as the revealed valuation weight that consumers place on the sales tax, with  $\theta = 0$  corresponding to complete neglect of the tax and  $\theta = 1$  corresponding to the equal weight of the tax and salient price. Each point is estimated using equation (4) for the specified posted prices. Standard errors are clustered at the subject level.

Table A.11: Average revealed valuation weights by group: using a 50th percentile cutoff

	Standard	Triple	Triple – Standard
(1): High valuation wgt.	0.84 [0.66, 1.05]	1.09 [1.00, 1.19]	0.25 [0.09, 0.40]
(2): Low valuation wgt.	0.32 [0.13, 0.49]	0.65 [0.58, 0.73]	0.33 [0.20, 0.47]
(3): (1) – (2)	0.53 [0.31, 0.77]	0.44 [0.34, 0.56]	–0.08 [–0.28, 0.09]

Table A.11 repeats Table 1 with an alternative split of consumers into high and low valuation weight groups. In this table high valuation weight individuals are those with  $F(\hat{\theta}_{iJB}) > 0.50$  and low valuation weight individuals are those with  $F(\hat{\theta}_{iJB}) \leq 0.50$ .

Table A.12: Average revealed valuation weights by valuation weight group: using an 80th percentile cutoff

	Standard	Triple	Triple – Standard
(1): High valuation wgt	1.24 [0.97, 1.51]	1.32 [1.20, 1.45]	0.08 [–0.14, 0.27]
(2): Low valuation wgt	0.29 [0.14, 0.44]	0.67 [0.59, 0.73]	0.38 [0.26, 0.49]
(3): (1) – (2)	0.95 [0.66, 1.24]	0.65 [0.52, 0.78]	–0.30 [–0.53, –0.08]

Table A.12 repeats Table 1 with an alternative split of consumers into high and low valuation weight groups. In this table high valuation weight individuals are those with  $F(\hat{\theta}_{iJB}) > 0.80$  and low valuation weight individuals are those with  $F(\hat{\theta}_{iJB}) \leq 0.80$ .

Table A.13: Average revealed valuation weights by valuation weight group: using an 85th percentile cutoff

	Standard	Triple	Triple – Standard
(1): High valuation wgt	1.23 [0.91, 1.54]	1.31 [1.17, 1.45]	0.08 [–0.17, 0.32]
(2): Low valuation wgt	0.35 [0.18, 0.50]	0.70 [0.63, 0.77]	0.35 [0.24, 0.48]
(3): (1) – (2)	0.88 [0.52, 1.22]	0.61 [0.47, 0.76]	–0.27 [–0.53, –0.01]

Table A.13 repeats Table 1 with an alternative split of consumers into high and low valuation weight groups. In this table high valuation weight individuals are those with  $F(\hat{\theta}_{iJB}) > 0.85$  and low valuation weight individuals are those with  $F(\hat{\theta}_{iJB}) \leq 0.85$ .

Table A.14: Average revealed valuation weights by adjustment group: using a 50th percentile cutoff

	Standard	Triple	Triple – Standard
(1): Low Adj.	0.72 [0.52, 0.93]	0.84 [0.75, 0.93]	0.12 [−0.04, 0.28]
(2): High Adj.	0.38 [0.21, 0.56]	0.78 [0.70, 0.86]	0.39 [0.26, 0.52]
(3): (1) – (2)	0.34 [0.13, 0.57]	0.06 [−0.03, 0.16]	−0.28 [−0.45, −0.10]

Table A.14 repeats Table 2 with an alternative split of consumers into high and low adjustment groups. For this table high adjustment individuals are those with  $F(\hat{\theta}_{i1C} - \hat{\theta}_{i1B}) > 0.50$  and low adjustment individuals are those with  $F(\hat{\theta}_{i1C} - \hat{\theta}_{i1B}) \leq 0.50$ .

Table A.15: Average revealed valuation weights by adjustment group: using a 20th percentile cutoff

	Standard	Triple	Triple – Standard
(1): Low Adj.	0.85 [0.60, 1.09]	0.85 [0.74, 0.96]	0.01 [−0.17, 0.19]
(2): High Adj.	0.38 [0.21, 0.53]	0.77 [0.70, 0.84]	0.39 [0.29, 0.52]
(3): (1) – (2)	0.47 [0.22, 0.73]	0.08 [−0.03, 0.19]	−0.39 [−0.60, −0.21]

Table A.15 repeats Table 2 with an alternative split of consumers into high and low adjustment groups. For this table high adjustment individuals are those with  $F(\hat{\theta}_{i1C} - \hat{\theta}_{i1B}) > 0.20$  and low adjustment individuals are those with  $F(\hat{\theta}_{i1C} - \hat{\theta}_{i1B}) \leq 0.20$ .

Table A.16: Average revealed valuation weights by adjustment group: using a 15th percentile cutoff

	Standard	Triple	Triple – Standard
(1): Low Adj.	1.23 [0.91, 1.54]	1.31 [1.17, 1.45]	0.08 [−0.17, 0.32]
(2): High Adj.	0.35 [0.18, 0.50]	0.70 [0.63, 0.77]	0.35 [0.24, 0.48]
(3): (1) – (2)	0.88 [0.52, 1.22]	0.61 [0.47, 0.76]	−0.27 [−0.53, −0.01]

Table A.16 repeats Table 2 with an alternative split of consumers into high and low adjustment groups. For this table high adjustment individuals are those with  $F(\hat{\theta}_{i1C} - \hat{\theta}_{i1B}) > 0.15$  and low adjustment individuals are those with  $F(\hat{\theta}_{i1C} - \hat{\theta}_{i1B}) \leq 0.15$ .

## L Replication of results restricting to participants with nearly-accurate beliefs and high computational ability

Do participants know their true sales tax rate, and if not, are incorrect beliefs a mechanism driving the results? Consistent with Chetty et al. (2009) and Taubinsky and Rees-Jones (2018), we find that participants generally have correct beliefs: 51.0% of our sample know their tax rate exactly, 70.3% within 0.5 percentage

points, and 82.0% within one percentage point.<sup>40</sup> We also do not find any evidence of systematic underestimation of tax rates. The mean of participants' estimates of their sales tax rates is 7.44%, which is negligibly higher than the actual mean of 7.24%. We refer to the 70.3% who know their tax within 0.5 percentage points as the "nearly-accurate beliefs" sample.

Another potential mechanism driving the results is the inability of participants to compute the sales tax they would need to pay for an item. We test for this mechanism by asking participants to report how much sales tax they would owe for an \$8.00 item purchased in their city of residence. 44.1% of participants are able to calculate their tax burden within \$0.01, and 62.9% are able to compute their tax burden within \$0.05.<sup>41</sup>

We first separately examine the effects of the two possible mechanisms, by estimating average revealed valuation weights restricting to (1) the 70.3% who know their sales tax rate within 0.5 percentage points, and (2) the 62.9% who can estimate the sales tax burden on an \$8.00 item purchased in their city of residence within \$0.05. Figures A.3 and A.4 present the results.

We next repeat our individual differences analysis, restricting to the "nearly-accurate beliefs and computation" sample. Tables A.17-A.19 recreate tables 1-3 restricting to this sample. Consistent our main results, the low valuation weight group exhibits a larger increase in the revealed valuation weights than the high valuation weight group when tax rates are tripled. The adjustments and their difference are similar in magnitude to our main sample results.

When dividing consumers by adjustment group, we still find that there are significant individual differences: consumers in the low adjustment group increase their valuation weights by an average of 0.04 (95% CI [-0.16, 0.24]), and those in the high adjustment group increase their revealed valuation weights by an average of 0.43 (95% CI [0.28, 0.58]). Consistent with our main prediction, and the possibility that some consumers might be overreacting, we find that consumers in the low adjustment group have higher valuation weights in both the standard tax regime (0.91 vs. 0.43; 95% CI for difference [0.76, 0.96]) and in the triple tax regime (0.95 vs. 0.48; 95% CI for difference [-0.04, 0.21]). The average valuation weight estimates are all slightly higher in this sample than in our main sample, but the differences all have a magnitude within 0.05 of our main results for both the standard tax regime and the triple tax regime.

<sup>40</sup>We asked participants to enter their answer as a percent rather than a decimal, and gave them the following example: "For example, if you think that the tax rate is 1 percent, please enter 1, rather than 0.01." 159 participants still entered a number less than 0.15. We attribute these low estimates to misunderstanding the instructions, and multiply these estimates by 100 when analyzing their beliefs.

<sup>41</sup>We did not explicitly remind participants to exclude the \$8.00 they would have to pay for the item from their answer. In our sample, there are 228 participants who entered an answer between 8 and 12. We attribute these high estimates to misunderstanding the instructions, and subtract 8 from these estimates in the analysis. We also observe 69 participants who entered an answer over 20. We attribute this to confusion as to whether answers should be entered as dollars (as we specified) or as cents. We divide these estimates by 100.

Figure A.3: Average revealed valuation weight for posted prices at or below a cutoff: nearly-accurate beliefs subsample

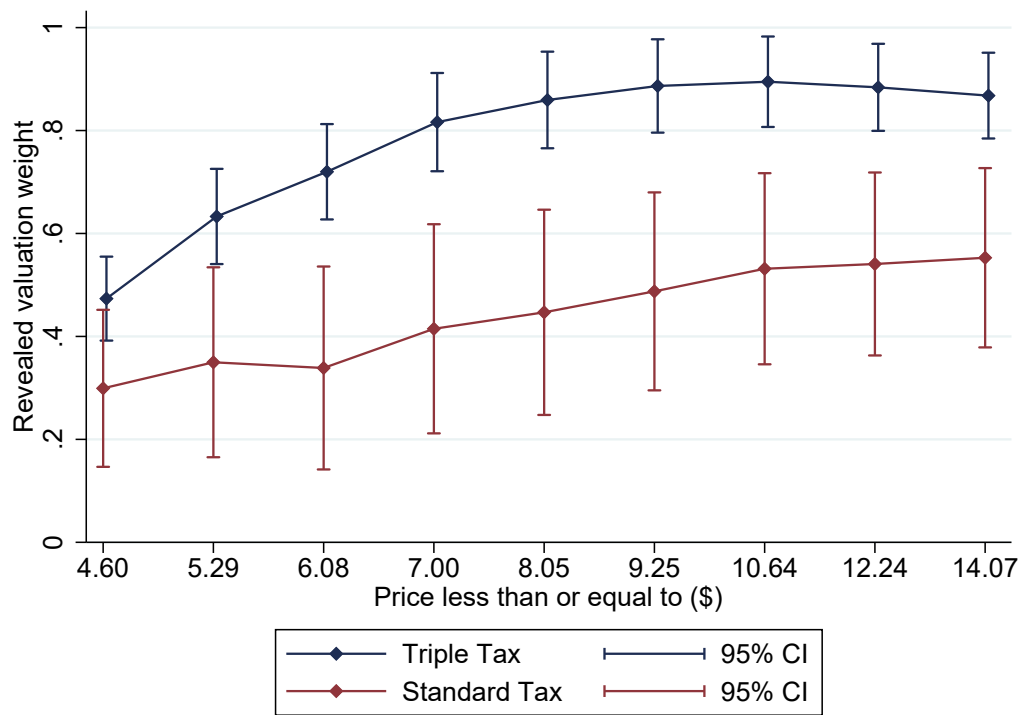


Figure A.3 recreates figure 2a, restricting to the 70.3% of the main sample who could identify their local sales tax rate within 0.5 percentage points.

Figure A.4: Average revealed valuation weight for posted prices at or below a cutoff: restricting to participants with strong computational ability

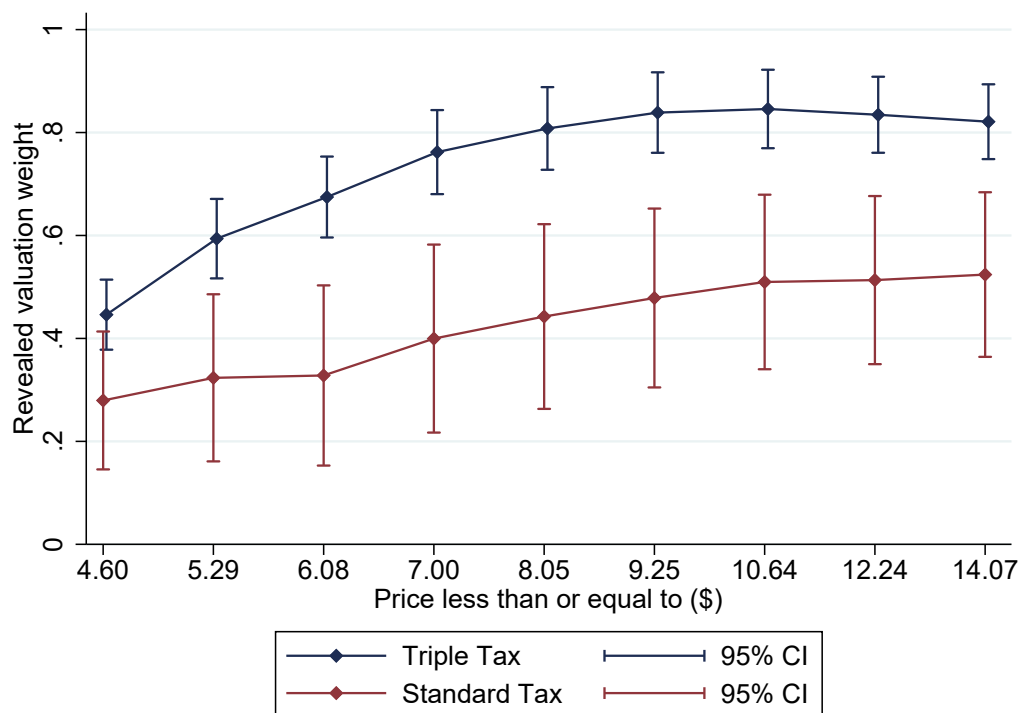


Figure A.4 recreates figure 2a, restricting to the 62.9% of the main sample who could identify their tax burden within \$0.05 on an \$8.00 item purchased in their city of residence.

Table A.17: Average revealed valuation weights by group: restricting to participants with nearly-accurate beliefs and strong computational ability

	Standard	Triple	Triple – Standard
(1): High valuation wgt.	1.11	1.32	0.40
	[0.85, 1.37]	[1.18, 1.45]	[0.24, 0.55]
(2): Low valuation wgt.	0.34	0.73	0.20
	[0.13, 0.55]	[0.64, 0.83]	[-0.01, 0.42]
(3): (1) – (2)	0.77	0.58	-0.19
	[0.47, 1.08]	[0.44, 0.73]	[-0.44, 0.06]
(4): Full sample	0.56	0.88	0.33
	[0.37, 0.74]	[0.79, 0.97]	[0.19, 0.47]

Table A.17 repeats Table 1, restricting to the 59.9% of the main sample who could identify their local sales tax rate within 0.5 percentage points and compute the sales tax they would owe for an \$8.00 item purchased in their city of residence within \$0.05.

Table A.18: Average revealed valuation weights by adjustment group: restricting to participants with nearly-accurate beliefs and strong computational ability

	Standard	Triple	Triple – Standard
(1): Low adj.	0.91 [0.64, 1.17]	0.95 [0.82, 1.07]	0.04 [-0.16, 0.24]
(2): High adj.	0.43 [0.23, 0.63]	0.48 [0.20, 0.76]	0.43 [0.28, 0.58]
(3): (1) – (2)	0.86 [0.76, 0.96]	0.08 [-0.04, 0.21]	-0.39 [-0.62, -0.17]
(4): Full sample	0.56 [0.37, 0.74]	0.88 [0.79, 0.97]	0.33 [0.19, 0.47]

Table A.18 repeats Table 2, restricting to the 59.9% of the main sample who could identify their local sales tax rate within 0.5 percentage points and compute the sales tax they would owe for an \$8.00 item purchased in their city of residence within \$0.05.

Table A.19: Bounds on the dispersion of revealed valuation weights: restricting to participants with nearly-accurate beliefs and strong computational ability

	Standard	Triple	Standard-Triple
Variance (Lower Bound)	0.71 [0.41]	0.75 [0.60]	0.84 [0.24]
Supremum (Lower Bound)	1.84 [1.31]	1.74 [1.56]	0.93 [0.05]

Table A.19 repeats Table 3, restricting to the 59.9% of the main sample who could identify their local sales tax rate within 0.5 percentage points and compute the sales tax they would owe for an \$8.00 item purchased in their city of residence within \$0.05.

## M Replication of main results without excluding study participants failing comprehension questions or violating monotonicity

In our primary analyses we exclude 255 respondents who incorrectly answered one or more of the comprehension questions and an additional 47 respondents who had monotonicity violations within a price list. Figure A.5 repeats Figure 2a including these 302 participants.<sup>42</sup> We again find strong evidence for Prediction 1, indicating that poor computational ability was not the sole mechanism driving consistency with the prediction. The estimates are of smaller magnitude than the full sample results, but are consistent with the theory which predicts average valuation weights are increasing in the absolute size of the tax. Using all prices we estimate an average revealed valuation weight of 0.36 (95% CI [0.22, 0.51]) for the standard tax environment in the restricted sample compared to 0.48 (95% CI [0.32, 0.63]) in the main sample. Similarly, we estimate an average revealed valuation weight of 0.67 (95% CI [0.60, 0.74]) for the triple tax environment in the restricted sample, which is only slightly lower than the estimate in the main sample 0.79 (95% CI [0.72, 0.86]).

Tables A.20-A.22 replicate tables 1-3 including the respondents who failed the comprehension checks.

<sup>42</sup>We continue to exclude one of these participants who reported being under age 18.



We still exclude participants with monotonicity violations, as our estimation procedure in Section 5.1 assumes monotonic preferences in estimating a willingness-to-pay.

As with our main results, the low valuation weight group exhibits a larger increase in the revealed valuation weights than the high valuation weight group when tax rates are tripled (0.39 vs. 0.13; 95% CI for difference  $[-0.42, -0.08]$ ). The adjustments and their difference are similar in magnitude to our main sample results (0.16 vs. 0.39; 95% CI for difference  $-0.43$  to  $-0.04$ ).

When dividing consumers by adjustment group, the estimates are also very similar in magnitude: consumers in the low adjustment group increase their valuation weights by an average of  $-0.00$  (95% CI  $-0.15$ - $0.15$ ) compared to  $0.01$  (95% CI  $[-0.15, 0.17]$ ) in our main sample. Similarly, those in the high adjustment group increase their revealed valuation weights by an average of  $0.42$  (95% CI  $[0.30, 0.54]$ ) compared to  $0.43$  (95% CI  $[0.30, 0.55]$ ) in our main sample. Consistent with our main results, we find that consumers in the low adjustment group have higher valuation weights in both the standard tax regime ( $0.77$  vs.  $0.24$ ; 95% CI for difference  $[0.32, 0.74]$ ) and in the triple tax regime ( $0.77$  vs.  $0.66$ ; 95% CI for difference  $[0.01, 0.20]$ ). The average valuation weight estimates are all slightly lower in this sample than in our main sample, but the differences all have a magnitude within  $0.01$  of our main results for both the standard tax regime and the triple tax regime.

Including participants who failed comprehension checks leads to a lower variance bound on adjustment ( $0.32$ , 5% confidence bound of  $0.18$ ) than the bound of ( $0.86$ , 5% confidence bound of  $0.31$ ) in our main sample. Additionally, we estimate an upper bound on  $\underline{\Delta}$  to be  $0.16$  (95% confidence bound of  $-0.06$ ), which is smaller than the bound from our main sample  $0.94$  (95% confidence bound of  $0.16$ ) and not statistically significantly below 0.

Figure A.5: Average revealed valuation weight for posted prices at or below a cutoff: including participants who fail comprehension checks

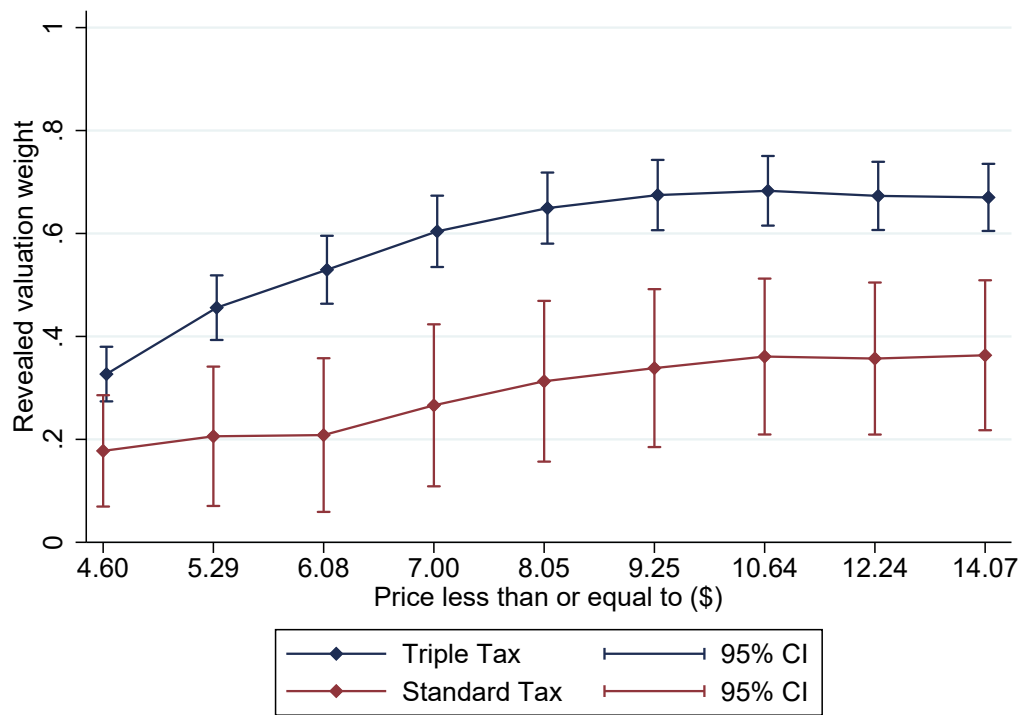


Figure A.5 recreates figure 2a, including the 302 participants who failed comprehension checks or had monotonicity violations in purchase decisions.

Table A.20: Average revealed valuation weights by group: including participants who fail comprehension checks

	Standard	Triple	Triple – Standard
(1): High valuation wgt.	0.96	1.09	0.39
	[0.76, 1.16]	[0.99, 1.19]	[0.27, 0.50]
(2): Low valuation wgt.	0.16	0.54	0.13
	[0.00, 0.52]	[0.48, 0.61]	[–0.02, 0.29]
(3): (1) – (2)	0.80	0.55	–0.25
	[0.57, 1.03]	[0.44, 0.65]	[–0.42, –0.08]

Table A.20 repeats Table 1 including 255 participants who were excluded from our main sample solely for failing our comprehension check.

Table A.21: Average revealed valuation weights by adjustment group: including participants who fail comprehension checks

	Standard	Triple	Triple – Standard
(1): Low adj.	0.77 [0.57, 0.97]	0.77 [0.68, 0.86]	–0.00 [–0.15, 0.15]
(2): High adj.	0.24 [0.09, 0.40]	0.66 [0.59, 0.74]	0.42 [0.30, 0.54]
(3): (1) – (2)	0.53 [0.32, 0.74]	0.11 [0.01, 0.20]	–0.42 [–0.59, –0.26]

Table A.21 repeats Table 2 including 255 participants who were excluded from our main sample solely for failing our comprehension check.

Table A.22: Bounds on the dispersion of revealed valuation weights: including participants who fail comprehension checks

	Standard	Triple	Standard-Triple
Variance (Lower Bound)	0.73 [0.51]	0.71 [0.59]	0.32 [0.18]
Supremum (Lower Bound)	2.29 [1.70]	1.73 [1.55]	0.16 [–0.06]

Table A.22 repeats Table 3 including 255 participants who were excluded from our main sample solely for failing our comprehension check.

## N Replication of main results excluding participants who always or never buy a product in at least one store

In this appendix we replicate the main results dropping all participants who either always buy or never buy at least one product in at least one store. This sample restriction excludes 47% of our main sample.

Table A.23 presents the results for average valuation weights. The estimates are of smaller magnitude than the full sample results, but are consistent with our prediction that average valuation weights are increasing in the absolute size of the tax. Using all prices we estimate an average revealed valuation weight of 0.36 (95% CI [0.22, 0.51]) for the standard tax environment in the restricted sample compared to 0.48 (95% CI [0.32, 0.63]) in the main sample. Similarly, we estimate an average revealed valuation weight of 0.67 (95% CI [0.60, 0.74]) for the triple tax environment in the restricted sample, which is only slightly lower than the estimate from our main sample, 0.79 (95% CI [0.72, 0.86]).

Tables A.24-A.25 replicate Tables 1-2 excluding the respondents who either always buy or never buy at least one product in at least one store. As with our main results, the low valuation weight group exhibits a larger increase in the revealed valuation weights than the high valuation weight group when tax rates are tripled (0.20 vs. 0.28; 95% CI for difference [–0.27, 0.13]). The adjustments and their difference are similar in magnitude to our main sample results (0.16 vs. 0.39; 95% CI for difference [–0.43, –0.04]).

When dividing consumers by adjustment group, the estimates are also broadly similar: consumers in the low adjustment group increase their valuation weights by an average of 0.15 (95% CI [–0.02, 0.31])

compared to 0.01 (95% CI  $[-0.15, 0.17]$ ) in our main sample. Similarly, those in the high adjustment group increase their revealed valuation weights by an average of 0.28 (95% CI  $[0.16, 0.40]$ ) compared to 0.43 (95% CI  $[0.30, 0.55]$ ) in our main sample. Consistent with our main results, we find that consumers in the low adjustment group have higher valuation weights in both the standard tax regime (0.47 vs. 0.28; 95% CI for difference  $[-0.02, 0.43]$ ) and in the triple tax regime (0.62 vs. 0.55; 95% CI for difference  $[-0.03, 0.17]$ ). In summary, we find that the estimates for this subsample conform to Predictions 1-4.

Table A.23: Average revealed valuation weights excluding those who always or never buy

Sample	Store B	Store C
(1): Excl. always or never buy	0.36	0.60
	[0.22, 0.50]	[0.53, 0.67]
(2): Full sample	0.48	0.79
	[0.32, 0.63]	[0.72, 0.86]

This table presents store-specific estimates of the average valuation weight. Row (1) presents estimates excluding 726 individuals who either always choose to purchase the product or never choose to purchase the product in a given store, while row (2) presents results for the full sample of 1534 individuals.  $\theta$  is defined as the revealed valuation weight that consumers place on the sales tax, with  $\theta = 0$  corresponding to complete neglect of the tax and  $\theta = 1$  corresponding to the equal weight of the tax and salient price. The results are estimated using equation (4). Standard errors are clustered at the subject level.

Table A.24: Average revealed valuation weights by group: excluding those who always or never buy

	Standard	Triple	Triple – Standard
(1): High valuation wgt.	0.76	0.96	0.20
	[0.54, 0.99]	[0.85, 1.07]	[0.01, 0.37]
(2): Low valuation wgt.	0.21	0.49	0.28
	[0.06, 0.37]	[0.42, 0.55]	[0.16, 0.39]
(3): (1) – (2)	0.55	0.47	-0.08
	[0.30, 0.81]	[0.36, 0.59]	[-0.27, 0.13]

This table repeats Table 1 excluding 726 individuals who either always choose to purchase the product or never choose to purchase the product in a given store.

Table A.25: Average revealed valuation weights by adjustment group: excluding those who always or never buy

	Standard	Triple	Triple – Standard
(1): Low Adj.	0.47	0.62	0.15
	[0.27, 0.68]	[0.52, 0.72]	[-0.02, 0.31]
(2): High Adj.	0.28	0.55	0.28
	[0.12, 0.43]	[0.48, 0.63]	[0.16, 0.40]
(3): (1) – (2)	0.19	0.07	-0.13
	[-0.02, 0.43]	[-0.03, 0.17]	[-0.33, 0.05]

This table repeats Table 2 excluding 726 individuals who either always choose to purchase the product or never choose to purchase the product in a given store.

## O Order Effects

A potential concern with our within-subject experimental design is that purchase decisions could be influenced by the order in which the nine purchase decisions are presented to consumers. For example, individuals might be more likely to buy in store A when store A preceded by store B rather than comes after store B, since in the former scenario store A seems like a particularly good deal. In Table A.26, we test four potential order effects. First, we examine whether the tax environment first shown to consumers impacts their buy probability. We test for this effect via the following model:

$$1 - Pr(buy_{ijk}|p) = \Phi \left( \frac{\alpha_j + \ln(p) + \bar{\theta}_B \ln(1 + \tau_{ik}) \cdot I(k = B) + \bar{\theta}_C \ln(1 + \tau_{ik}) \cdot I(k = C)}{\sigma_j} + \frac{\gamma^B First_i^B + \gamma^C First_i^C}{\sigma_j} \right) \quad (28)$$

This model modifies equation (4) by adding the terms  $First_i^B$  and  $First_i^C$ .  $First_i^k$  is an indicator variable which equals one if the consumer's first purchase decision occurred in store  $k$  and equals zero otherwise. We compute the Wald statistic for  $\gamma^B = \gamma^C = 0$ , which has a corresponding p-value of 0.95.

In our next three tests, we examine product-specific order effects, or whether a consumer's buy probability for product  $j$  is affected by the store order in which the consumer shops for product  $j$ . For our second test, we construct indicator variables  $First_{ij}^k$  which equal one if the consumer's first purchase decision for product  $j$  occurred in store  $k$  and equals zero otherwise. We then repeat equation (28), using  $First_{ij}^k$  instead of  $First_i^k$ :

$$1 - Pr(buy_{ijk}|p) = \Phi \left( \frac{\alpha_j + \ln(p) + \bar{\theta}_B \ln(1 + \tau_{ik}) \cdot I(k = B) + \bar{\theta}_C \ln(1 + \tau_{ik}) \cdot I(k = C)}{\sigma_j} + \frac{\gamma^B First_{ij}^B + \gamma^C First_{ij}^C}{\sigma_j} \right)$$

We compute the Wald statistic for  $\gamma^B = \gamma^C = 0$ , which has a corresponding p-value of 0.70.

For our third test, we examine whether the last store shown to consumers for a product affects their purchase decision. We construct indicator variables  $Last_{ij}^k$  which equal one if the consumer's last purchase decision for product  $j$  occurred in store  $k$  and equals zero otherwise. We then repeat equation (28), using  $Last_{ij}^k$  instead of  $First_i^k$ :

$$1 - Pr(buy_{ijk}|p) = \Phi \left( \frac{\alpha_j + \ln(p) + \bar{\theta}_B \ln(1 + \tau_{ik}) \cdot I(k = B) + \bar{\theta}_C \ln(1 + \tau_{ik}) \cdot I(k = C)}{\sigma_j} + \frac{\gamma^B Last_{ij}^B + \gamma^C Last_{ij}^C}{\sigma_j} \right)$$

We compute the Wald statistic for  $\gamma^B = \gamma^C = 0$ , which has a corresponding p-value of 0.28.

For our fourth test, we construct indicator variables for each possible combination stores A, B, and C were presented to consumer  $i$  for product  $j$ . We then estimate the following model for  $\kappa_1, \kappa_2, \kappa_3 = \{A, B, C\}$ :<sup>43</sup>

$$1 - Pr(buy_{ijk}|p) = \Phi \left( \frac{\alpha_j + \ln(p) + \bar{\theta}_B \ln(1 + \tau_{ik}) \cdot I(k = B) + \bar{\theta}_C \ln(1 + \tau_{ik}) \cdot I(k = C) \cdot I(\tau_{ik} = 3\tau_i)}{\sigma_j} \right. \\ \left. + \frac{\sum_{\kappa_2 \neq \kappa_1; \kappa_3 \neq \kappa_2, \kappa_1} \gamma^{\kappa_1 \kappa_2 \kappa_3} I(First_{ij} = \kappa_1, Second_{ij} = \kappa_2, Third_{ij} = \kappa_3)}{\sigma_j} \right)$$

We compute the Wald statistic for  $\gamma^{ACB} = \gamma^{BCA} = \dots = 0$ , which has a corresponding p-value of 0.17.

As a test of whether attention is altered by the within-subject nature of our design, in Table A.27 we report estimates of average  $\theta$  using the first  $N = 1, 2, \dots, 9$  decisions that consumers make. Although the results are noisy for the standard tax condition for low values of  $N$ , the triple tax condition provides us with greater statistical power, and shows that there is little variation in the estimates of average  $\theta$  when we use only initial decisions or all decisions.

Table A.26: Tests for the impact of order effects on buy probability

Order effect tested	p-value
Tax env. of first purchase decision	0.95
Tax env. of first purchase decision (by product)	0.70
Tax env. of last purchase decision (by product)	0.28
Ordering of tax env. (by product)	0.17

Table A.26 presents p-values of Wald statistics for the impact of order effects on buy probabilities. The Wald statistics and p-values are calculated by adding indicators for the different orderings tested to equation (4). For the first row, we add two indicators for whether the tax environment of the first purchase decision shown to consumers was standard tax or triple tax. For the second (third) row, we add two indicators for whether the tax environment of the first (last) purchase decision for product  $j$  was standard tax or triple tax. For the fourth row, we add five indicators for each of the possible orders in which store A, B, and C were presented to the consumer for product  $j$  (order A, B, C was omitted due to collinearity).

<sup>43</sup>We omit the store ordering A, B, C due to collinearity.

Table A.27: Average revealed valuation weights using the first N purchase decisions

N	Store B	95% CI	Store C	95% CI
1	0.27	[−0.54, 1.07]	0.57	[0.28, 0.85]
2	0.10	[−0.43, 0.63]	0.72	[0.52, 0.91]
3	0.05	[−0.36, 0.47]	0.72	[0.57, 0.87]
4	0.25	[−0.09, 0.59]	0.73	[0.61, 0.86]
5	0.30	[0.02, 0.59]	0.75	[0.64, 0.86]
6	0.49	[0.24, 0.73]	0.82	[0.73, 0.92]
7	0.54	[0.33, 0.75]	0.84	[0.76, 0.93]
8	0.52	[0.33, 0.70]	0.81	[0.73, 0.89]
9	0.48	[0.32, 0.63]	0.79	[0.72, 0.86]

Table A.27 presents store-specific estimates of  $E[\theta]$  for the first N purchase decisions made by participants.  $\theta$  is defined as the revealed valuation weight that consumers place on the sales tax, with  $\theta = 0$  corresponding to complete neglect of the tax and  $\theta = 1$  corresponding to the equal weight of the tax and salient price. The final row includes all nine purchases decisions, and matches the estimates from row (4) in Table 1. The results are estimated using equation (4), interacting the covariate with price and tax. Standard errors are clustered at the subject level.

## P Comparison of demand curves to Amazon.com prices

Participants in our experiment made online purchase decisions for goods that were also available in a variety of online stores, including Amazon.com. When in our online shopping experiment, consumers might then incorporate the prices for the online stores into their WTP and potentially elect not to buy in our experiment whenever they could purchase the product at a cheaper price in an online store.

While our experiment is not designed to check consumers' awareness of prices at other online stores, we do observe how frequently consumers are willing to purchase above the Amazon.com prices. In Figure A.6, we plot product-specific demand curve in the no-tax environment compared to the prices listed on Amazon.com near the time of our experiment. These prices range from \$7.73 to \$12.99, which exceed five and nine, respectively, of the prices on our MPL. Depending on the product, approximately 10-30% of consumers choose to buy the product at a price above the Amazon.com price.

There are several caveats in comparing the consumer demand curves to the Amazon.com prices. First, purchasing a product on Amazon.com often requires consumers to pay additional costs such as shipping fees or, in select states, sales taxes.<sup>44</sup> Second, the Amazon.com prices we report are from February 2015, as documented in Taubinsky and Rees-Jones (2018). They may vary over time or by geographic region. Third, we only report the price available on Amazon.com, even though consumers could also buy the products from other online or physical stores.

<sup>44</sup>At the time of our experiment (September 2016), Amazon did not collect sales taxes from most states, which is why we choose to compare Amazon.com prices to the demand curves in the no-tax store.

Figure A.6: Product-specific demand curves from the no-tax environment compared to Amazon.com prices

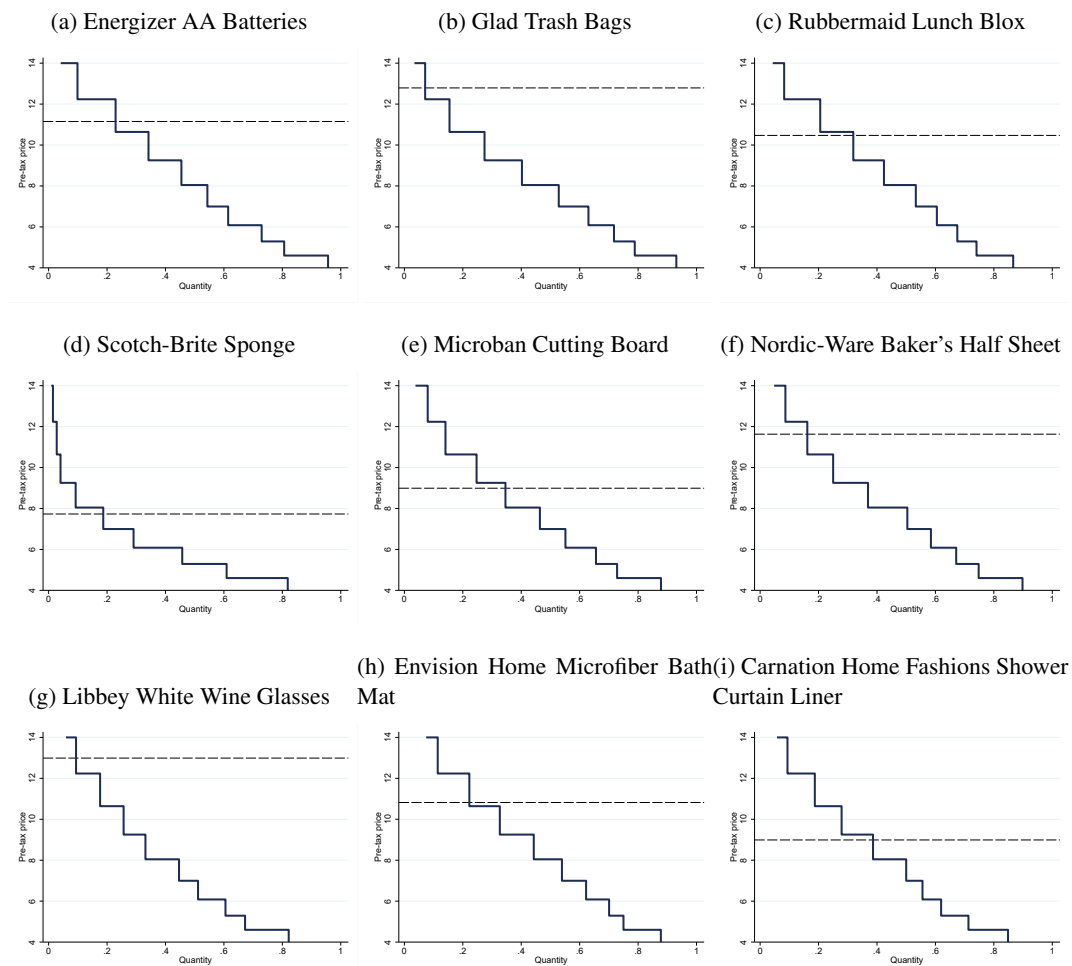


Figure A.6 presents product-specific demand curves from the no-tax store. The Amazon.com price is indicated by the dashed line. Prices are from February 2015, as documented in Taubinsky and Rees-Jones (2018). They may vary over time or by geographic region.

## Q Welfare implications of overreaction

While the primary focus of this paper is to link our our detailed results about misreaction to models of costly attention, in this appendix we briefly note that our findings also have substantial policy implications. Our evidence suggests that shrouding taxes can generate significant deadweight loss for two reasons. First, because it leads consumers to exert costly attention to compute the post-tax prices. Second, because consumers' highly heterogeneous reactions to these taxes can lead to misallocation: variation in price perceptions, due to underreaction by some and overreaction by others, creates a misallocation of products to consumers.

Concretely, Taubinsky and Rees-Jones (2018, TRJ) show that excluding attention costs, the deadweight loss from a small tax  $t$  in a competitive market with price-taking firms is given by



$$DWL(t) \approx \frac{t^2}{2} (\rho E[\theta|p,t]^2 + \text{Var}[\theta|p,t]) \frac{\varepsilon_{D,p}}{p+t}$$

where  $\varepsilon_{D,p}$  is the price elasticity of demand,  $E[\theta|p,t]$  and  $\text{Var}[\theta|p,t]$  correspond to the mean and variance of consumers marginal at price  $p$  and tax  $t$ ,<sup>45</sup> and  $\rho$  is the pass-through rate of producer taxes (e.g., excise taxes) to prices. The variance of misreaction is a sufficient statistic for the efficiency costs of heterogeneous tax misreaction. The formula generalizes the homogeneous underreaction case analyzed by Chetty et al. (2009), in which making taxes opaque always reduces deadweight loss under the assumption of quasilinear utility.<sup>46</sup> Note that we assume that the tax is sufficiently small that it does not affect the pass-through rate. If pass-through rates are decreasing in  $\theta$  this can attenuate (amplify) DWL due to overreaction (underreaction).

With heterogeneity, making taxes opaque increases deadweight loss if and only if  $\rho E[\theta|p,t]^2 + \text{Var}[\theta|p,t] > \rho$ . In the leading case of constant marginal costs of production (and thus full pass-through), this reduces to  $E[\theta|p,t]^2 + \text{Var}[\theta|p,t] > 1$ . But as shown in equation (8) of Proposition 1,  $E[\theta|p,t]^2 + \text{Var}[\theta|p,t] \leq E[\theta|p,t]$ , and thus deadweight loss is guaranteed to be smaller with shrouded taxes if  $\theta < 1$  for all consumers. For example, combining our experimental estimate of  $E[\theta] = 0.48$  with the presumption that all consumers underreact would imply that deadweight loss is *at least* 50% smaller when sales taxes are shrouded. Instead, we find significant overreaction, corresponding to  $\text{Var}[\theta] \geq 0.83$ . This implies that  $E[\theta|p,t]^2 + \text{Var}[\theta|p,t] \geq 1.06$ , and thus that shrouding taxes increases deadweight loss. Moreover, because this calculation uses the lower-bound variance estimate, considers full pass-through, and ignores the mental effort costs used to process the opaque taxes, the actual deadweight loss may be significantly higher if (i) the variance is significantly larger, (ii) mental costs of effort are taken into account and (iii)  $\rho$  is significantly lower than 1.<sup>47</sup>

## R Additional details of the experiment

The experiment proceeded in the following order:

- Consent form: participants were first shown a consent form, which is shown in Figure A.7.
- Questions on residence: participants answered the questions in Section R.2.1 about their residence. The city and state selected in Question 1 were entered in future questions.
- Instructions: participants were shown three screens containing the experiment instructions. Figures A.8-A.10 contain the screenshots.

<sup>45</sup>Concretely, this is mathematically equivalent to the formula in Proposition 2 of TRJ. See also Farhi and Gabaix (2020) for similar insights with applications to Ramsey taxation.

<sup>46</sup>Quasilinearity is a sensible assumption for small-stakes purchases such as those in our experiment, and which constitute a large share of people's typical consumption expenditures. For larger stakes purchases where quasilinearity does not apply, Chetty et al. (2009) show that making taxes opaque can reduce efficiency even with homogeneous misreaction, due to the poor budgeting that it causes.

<sup>47</sup>As discussed in the next section, TRJ's variance bound of 0.13 is far too loose to either deduce overreaction or to conclude that making sales taxes opaque reduces welfare. Instead, TRJ show that this bound implies that assuming homogeneity, as in Chetty et al. (2009), produces a deadweight loss estimate that is relatively too small.

- Pre-purchase comprehension questions: participants answered the questions in Section R.2.2. Participants must answer all questions correctly to proceed. They were informed if they answered a question incorrectly, and were given unlimited attempts to get the correct answers.
- Purchase decisions: participants made nine purchase decisions. Each participant was randomly assigned three products from the list in Section R.3, and shopped for each product in all three stores. For each purchase, participants were first shown a screen detailing which store they entered, as seen in Figure A.11a. They then filled out an MPL, as seen in Figure A.11b.
- Post-purchase comprehension questions: participants answered the questions in Section R.2.3.
- Additional closing questions: participants answered the questions in Section R.2.4.

Section R.1 contains screenshots of the instructions shown to participants. Section R.2 contains text of the questions asked of participants, and, where applicable, the correct answer displayed in parenthesis and italics. Section R.3 contains a list of the products used, along with the Amazon.com prices and product descriptions.

## R.1 Instructions

Figure A.7: Introduction Screen

You are being asked to take part in an online shopping experiment. We anticipate that the experiment will take less than 20 minutes to complete. Your participation is voluntary, and is greatly appreciated.

Please complete this study on your computer, not your mobile phone. The study will not display correctly on any device other than a computer.

**Compensation:**

You will receive \$2.00 for your participation in this study. Furthermore, participants who complete this study have a one in three chance of receiving \$16, to use as a shopping budget. If you receive this shopping budget, it is yours to keep, but you may choose to spend part of that budget to purchase an item in the course of the study.

**Contact information:**

This study is being conducted by economic researcher Dmitry Taubinsky (Dartmouth College). If you have any questions or comments, please contact Dmitry Taubinsky at [dmitry.taubinsky@dartmouth.edu](mailto:dmitry.taubinsky@dartmouth.edu).

If you agree to participate in this survey, please click on the continue button below to begin.

Continue

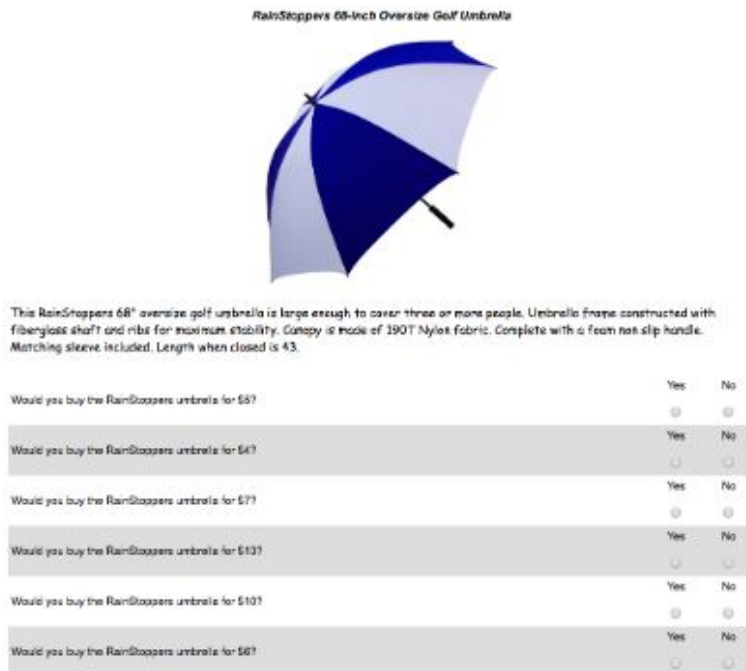
Figure A.8: Instructions (screen 1 of 3)

Instructions

In this study, you will answer questions about your willingness to buy various household products. We would like you to imagine that you are looking at these products in a local store, making a decision about whether or not to buy them at the price that is listed on their price tag.

For each product, you will be presented with a screen like the one below. You will see the product and read a brief description. Then at the bottom of the screen, you will answer whether you would buy the product at various prices.

For these questions, imagine that we are giving you a \$16 shopping budget to potentially purchase the item. You will get to keep whatever money you don't spend. As we will explain shortly, some respondents will actually receive the \$16 shopping budget and will have one of their purchasing decisions implemented.



To continue reading the instructions, please click the "Continue" button. You may click the "Back" button at any time to read the instructions on the previous page.

Note: Subjects did not shop for the Oversize Golf Umbrella in the experiment.

Figure A.9: Instructions (screen 2 of 3)

### Your shopping decisions

The purpose of the study is to understand how people make shopping decisions in their day-to-day lives. There are no right or wrong purchasing decisions in this study; different people have different preferences and behave in different ways. Some people may value the products very little, while others may value them a lot. Some people may choose to ignore sales taxes, while others may want to keep track of them.

You will be asked about your willingness to buy various products, at prices that will range from low to high, and in different kinds of stores. The differences between the products and the prices will be self-explanatory. The three stores are as follows:

When you purchase an item in Store A, you will pay **no sales tax** in addition to the price. Store A is like one of your local stores, with the taxes already included in the prices that you see on the tags of the items.

When you purchase an item in Store B, you will have to pay an additional sales tax, just like you typically do at the register at your local stores (on non-tax-exempt items). The sales tax rate in Store B is **the standard sales tax rate** that applies in your city of residence, [city], [state].

When you purchase an item in Store C, the sales tax that you have to pay in addition to the price is much higher than what you would have to pay at your local stores. The sales tax rate in Store C is **triple the standard sales tax rate** that applies in your city of residence, [city], [state].

Figure A.10: Instructions (screen 3 of 3)

### How your decisions in this study impact you

The decisions you make in this study are important. One out of every three participants will be selected to receive \$16, which you can use in one of the shopping decisions in this study. This is purely random and is in no way affected by the decisions you make.

If you are randomly selected to receive a \$16 shopping budget then:

The computer will randomly select one of the stores.

Then the computer will randomly select one of the items in that store.

Then the computer will randomly select one (and only one) of the prices for which you made a "yes/no" buying decision.

If you answer "no" in that randomly selected question, then you will keep the \$16 and not receive the item in the question.

If you answer "yes" in that randomly selected question, then you will receive the item in the question. The amount of money that you will keep will be \$16 minus the price, and minus the tax that you have to pay in the store selected by the computer.

Every store, item, and price has an equal probability of being selected.

### The bottom line:

The goal of these procedures is to ensure that you answer every question as carefully and honestly as possible. Although the procedures may seem complicated, **the most important thing to keep in mind is that it is in your best interest to answer each question honestly.**

Figure A.11: Screenshots of a purchase decision

(a) Introduction

You are now entering Store B to shop for: *Glad OdorShield Tall Kitchen Drawstring Trash Bags*.

Continue

(b) Purchase decision

*Glad OdorShield Tall Kitchen Drawstring Trash Bags, Fresh Clean, 13 Gallon, 80 Count*



**Product Description:** Glad OdorShield Tall Kitchen Drawstring Trash Bags backed by the power of Febreze are tough, reliable trash bags that neutralize strong and offensive odors for lasting freshness. These durable bags are great for use in the kitchen, home office, garage, and laundry room.

Would you buy the Glad OdorShield Trash Bags for \$8.05?	Yes	No
	<input type="radio"/>	<input type="radio"/>
Would you buy the Glad OdorShield Trash Bags for \$5.29?	Yes	No
	<input type="radio"/>	<input type="radio"/>
Would you buy the Glad OdorShield Trash Bags for \$7.00?	Yes	No
	<input type="radio"/>	<input type="radio"/>
Would you buy the Glad OdorShield Trash Bags for \$4.00?	Yes	No
	<input type="radio"/>	<input type="radio"/>
Would you buy the Glad OdorShield Trash Bags for \$10.64?	Yes	No
	<input type="radio"/>	<input type="radio"/>
Would you buy the Glad OdorShield Trash Bags for \$9.25?	Yes	No
	<input type="radio"/>	<input type="radio"/>
Would you buy the Glad OdorShield Trash Bags for \$12.24?	Yes	No
	<input type="radio"/>	<input type="radio"/>
Would you buy the Glad OdorShield Trash Bags for \$4.60?	Yes	No
	<input type="radio"/>	<input type="radio"/>
Would you buy the Glad OdorShield Trash Bags for \$6.08?	Yes	No
	<input type="radio"/>	<input type="radio"/>
Would you buy the Glad OdorShield Trash Bags for \$14.07?	Yes	No
	<input type="radio"/>	<input type="radio"/>

Back Continue

Figure A.11 shows an example of the two screenshots participants see for each of their nine purchase decisions. Subjects first saw a screen indicating the product for which they will be shopping and the relevant sales tax environment. Store A corresponds to a tax-free environment, store B to a standard sales tax environment, and store C to a triple-the-standard sales tax environment. On the second screen, participants saw an image and product description from Amazon.com, and were asked a series of questions about whether they would buy the product at various prices. The order of the prices was randomized. When filling out the price list, participants were able to click on a “back” button to revisit the first screen with the store information and an “instructions” button to reread the experiment instructions.

## R.2 Text of questions

### R.2.1 Residence

Question 1: Please select the state, county, and city in which you currently live:

Question 2: How long have you live in [city], [state]

- Less than 1 year
- Between 1 and 3 years
- Between 3 and 5 years
- 5 years or longer

Question 3: Where did you live prior to living in [city], [state]? (*Asked only if “5 years or longer” was not selected on the previous question*)

### R.2.2 Pre-purchase comprehension questions

Question 1: How big of a shopping budget do you have for each purchase decision? (*Correct answer: option 2*)

- \$10
- \$16
- \$30

Question 2: If you are selected to receive the shopping budget, how many of your purchase decisions will the computer randomly choose to implement? (*Correct answer: option 1*)

- The computer randomly chooses **one** decision to play out for real outcomes
- The computer randomly chooses **ten** decisions to play out for real outcomes
- The computer plays out **all** decisions

Question 3: At what prices do you see the products in this study? (*Correct answer: The prices vary*)

- The prices are always fixed at \$5
- The prices are always fixed at \$15
- The prices vary from low to high

Question 4: If you purchase an item for \$10 in Store A, then... (*Correct answer: option 1*)

- ... you will pay no sales tax in addition to the \$10 (the sales tax is included in the price).



- ... in addition to the \$10, you will pay the standard sales tax that you pay on a \$10 purchase in [city], [state].
- ... in addition to the \$10, you will pay **triple** the standard sales tax that you pay on a \$10 purchase in [city], [state].

Question 5: If you purchase an item for \$10 in Store B, then... (*Correct answer: option 2*)

- ... you will pay no sales tax in addition to the \$10 (the sales tax is included in the price).
- ... in addition to the \$10, you will pay the standard sales tax that you pay on a \$10 purchase in [city], [state].
- ... in addition to the \$10, you will pay **triple** the standard sales tax that you pay on a \$10 purchase in [city], [state].

Question 6: If you purchase an item for \$10 in Store C, then... (*Correct answer: option 3*)

- ... you will pay no sales tax in addition to the \$10 (the sales tax is included in the price).
- ... in addition to the \$10, you will pay the standard sales tax that you pay on a \$10 purchase in [city], [state].
- ... in addition to the \$10, you will pay **triple** the standard sales tax that you pay on a \$10 purchase in [city], [state].

### R.2.3 Closing comprehension questions

The questions below are about a hypothetical survey respondent Alex, who lives in [city], [state] just like you. **You must answer these three questions correctly to be eligible for the \$16 shopping budget.**

Question 1: If Alex purchases an item for \$10 in Store A, then... (*Correct answer: option 1*)

- ... Alex will pay no sales tax in addition to the \$10 (the sales tax is included in the price).
- ... in addition to the \$10, Alex will pay the standard sales tax that he pays on a \$10 purchase in [city], [state].
- ... in addition to the \$10, Alex will pay **triple** the standard sales tax that he pays on a \$10 purchase in [city], [state].

Question 2: If Alex purchases an item for \$10 in Store B, then... (*Correct answer: option 2*)

- ... Alex will pay no sales tax in addition to the \$10 (the sales tax is included in the price).
- ... in addition to the \$10, Alex will pay the standard sales tax that he pays on a \$10 purchase in [city], [state].

- ... in addition to the \$10, Alex will pay **triple** the standard sales tax that he pays on a \$10 purchase in [city], [state].

Question 3: If Alex purchases an item for \$10 in Store C, then... (*Correct answer: option 3*)

- ... Alex will pay no sales tax in addition to the \$10 (the sales tax is included in the price).
- ... in addition to the \$10, Alex will pay the standard sales tax that he pays on a \$10 purchase in [city], [state].
- ... in addition to the \$10, Alex will pay **triple** the standard sales tax that he pays on a \$10 purchase in [city], [state].

#### R.2.4 Additional closing questions

Question 1: 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 non-exempt good. If you're not sure, please make your best guess. (Note: Please enter your answer as a percent. For example, if you think that the tax rate is 1 percent, please enter 1, rather than 0.01. Do not include a percent sign in your answer.)

Question 2: Suppose that you bought a (standard, not tax-exempt) product for \$8. What would be the [city], [state] sales tax that you would have to pay for that product? (Note: Please enter your answer in dollar units. For example, if you think that the sales tax is 10 cents, please enter 0.10 and not 10. Do not include a dollar sign in your answer.)

Question 3: Suppose you had \$100 in a savings account and the interest rate was 2 percent per year. After 5 years, how much do you think you would have in the account if you left the money to grow? (*Correct answer: option 1*)

- More than \$102
- Exactly \$102
- Less than \$102
- Do not know

Question 4: Imagine that the interest rate on your savings account was 1 percent per year and inflation was 2 percent per year. After 1 year, would you be able to buy more than, exactly the same as, or less than today with the money in this account? (*Correct answer: option 3*)

- More than today
- Exactly the same as today
- Less than today
- Do not know

Question 5: Do you think that the following statement is true or false? “Buying a single company stock usually provides a safer return than a stock mutual fund.” (*Correct answer: option 2*)

- True
- False
- Do not know

Question 6: How many people are in your household (including yourself)?

Question 7: What is your marital status?

- Single
- Married or domestic partnership
- Widowed

Question 8: What was your total household income for the year 2015?

Question 9: What is the highest level of education that you have attained?

- Some high school
- High school graduate
- Some college or associate degree
- College graduate
- Master’s degree
- Doctoral degree (Ph.D., M.D., J.D., or equivalent)

Question 10: Are you currently a student?

- Yes
- No

Question 11: What is your age?

Question 12: Which best describes your political party affiliation?

- Independent
- Republican
- Democrat
- Other

**R.3 Items used in the study**

Product	Amazon.com price	Amazon.com product description
Energizer AA Batteries max Alkaline 20-Pack	\$11.15	Energizer AA max alkaline batteries 20 pack super fresh, Expiration Date: 2024 or better. Packed in original Energizer small box 4 batteries per box x 5 boxes total 20 batteries.
Glad OdorShield Tall Kitchen Drawstring Trash Bags, Fresh Clean, 13 Gallon, 80 Count	\$12.79	Glad OdorShield Tall Kitchen Drawstring Trash Bags backed by the power of Febreze are tough, reliable trash bags that neutralize strong and offensive odors for lasting freshness. These durable bags are great for use in the kitchen, home office, garage, and laundry room.
Rubbermaid Lunch Blox medium durable bag - Black Etch	\$10.47	The Rubbermaid 1813501 Lunch Blox medium durable bag - Black Etch is an insulated lunch bag designed to work with the Rubbermaid Lunch Blox food storage container system. The bag is insulated to achieve the maximum benefit of Blue Ice blocks and keep your food cold. The bag features a bottle holder, side pocket, comfort-grip handle and removable shoulder strap. The lunch Blox bag is durable and looks good for both the professional bringing their lunch to work or the kid taking their lunch to school.
Scotch-Brite Heavy Duty Scrub Sponge 426, 6-Count	\$7.73	O-Cel-O™ sponges and Scotch Brite scrubbers are truly a fashion-meets-function success story. The highly absorbent and durable sponges come in different sizes and scrub levels for the various surfaces around the home. Their assorted colors and patterns follow the current fashion trends to create the perfect accent in any room.
Microban Antimicrobial Cutting Board Lime Green - 11.5x8 inch	\$8.99	The Microban cutting board from Uniware is the perfect cutting board for the health conscious. The cutting board has a soft grip with handle and is dishwasher safe. The cutting board can be reversible, used on both sides, and is non-porous, non-absorbent. The rubber grips prevents slipping on countertop. Doesn't dull knives, juice-collecting groove. Microban is the most trusted antimicrobial product protection in the world. Built-In defense that inhibits the growth of stain and odor causing bacteria, mold, and mildew. Always works to keep the cutting board cleaner between cleanings. Lasts throughout the lifetime of the cutting board. Size: 11.5"x8" Color: Lime Green.
Nordic Ware Natural Aluminum Commercial Baker's Half Sheet	\$11.63	Nordic Ware's line of Natural Commercial Bakeware is designed for commercial use, and exceeds expectations in the home. The durable, natural aluminum construction bakes evenly and browns uniformly, while the light color prevents over-browning. The oversized edge also makes getting these pans in and out of the oven a cinch. Proudly made in the USA by Nordic Ware.
Libbey 14-Ounce Classic White Wine Glass, Clear, 4-Piece	\$12.99	Great for any party, this set includes four 14-ounce clear classic white wine glasses which match perfectly with the classic collection by libbey. The glasses are dishwasher safe and made in the USA.

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

Note: Prices are from February 2015, as documented in Taubinsky and Rees-Jones (2018). They may vary over time or by geographic region.

