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

We are especially grateful to Alex Rees-Jones for many helpful conversations and input that lead 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, John Leahy, Daniel Martin, Josh Schwartzstein, and Philipp Strack, as well as conference and seminar participants, for helpful comments. Afras Sial provided excellent research assistance. Taubinsky thanks the 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.

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
JEL No. D9,H2

ABSTRACT

This paper develops a methodology for testing whether attention costs are a source of consumers' misreaction to opaque prices. We show that costly attention models make a series of predictions about how individual differences in misreaction respond to stakes. We then test and confirm these predictions in an experiment on consumers' online shopping decisions in the presence of shrouded sales taxes that are exogenously varied within consumer over time. The empirical results point to a model in which consumers use heterogeneous rules of thumb to compute the opaque tax when the stakes are low, but use costly mental effort to increase their accuracy when the stakes increase. In particular, some consumers systematically underreact to sales taxes while others systematically overreact. But when the stakes increase, consumers who tend to underreact become more sensitive to sales taxes, while consumers who tend to overreact becomes less sensitive to sales taxes. We establish the results both by using simple reduced-form tests as well as by developing novel econometric techniques for quantifying individual differences. The results are inconsistent with models in which attention is exogenous and models in which all consumers are either fully attentive to the tax or ignore it completely.

William Morrison
University of California, Berkeley
Department of Economics
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 and cognitive scientists 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 (Caplin, 2016; Maćkowiak et al., 2018; Gabaix, 2019).[1] 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*. Intuition suggests, however, that when purchasing an expensive electronics product or an automobile, 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.

Such deliberate, and plausibly elastic, use of cognitive shortcuts may not only play an important role in recent findings of misreaction to sales taxes (Chetty et al., 2009; Feldman and Ruffle, 2015; Taubinsky and Rees-Jones, 2018), but may also play a role in misreaction to shipping and handling charges (Hossain and Morgan, 2006), energy prices (Allcott and Taubinsky, 2015), various features of health insurance contracts (Handel and Kolstad, 2015; Bhargava et al., 2017; Abaluck and Adams, 2017), left-digit bias (Lacetera et al., 2012; Shlain, 2019), shrouded financial fees (Heidhues et al., 2017), and add-on charges (Gabaix and Laibson, 2006).

This relatively recent but quickly growing empirical literature has provided compelling reduced-form tests that consumers misreact to opaque prices in various settings, but the mechanisms still remain largely unexplored. One possibility, as in the example above, is that consumers are aware of the opaque prices, are in principle capable of correctly incorporating them into their decisions, but choose not to do so to avoid cognitive costs. This mechanism 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). The resource rationality framework decomposes "the value of having applied a heuristic into the utility of the judgment, decision, or belief update that results from it and the computational cost of its execution" (Lieder and Griffiths, 2019).

However, avoidance of attention costs is not the only plausible source of misreaction. Other possible sources of misreaction may include complete unawareness of the opaque price, incorrect beliefs generated by systematic mislearning[2] or misleading marketing, forgetting, or a systemic lack of financial literacy that prevents consumers from reaching the right answer no matter how hard they try.

[1] See also Lieder and Griffiths (2019) for a review of the work in the cognitive sciences.
[2] See, e.g., Schwartzstein (2014) and Hanna et al. (2014).

This paper develops a theoretical, statistical, and experimental methodology for testing whether costly attention is a source of misreaction to opaque prices. We begin in Section 2 by formalizing the economic environment and several types of costly inattention 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,[3] and the attention weight adjustment cost function of Gabaix (2014).

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.

This simplification is in the spirit of Matejka and McKay (2015) who show that rational inattention with Shannon entropy leads to choices that follow the multinomial logit model. It is also in the spirit of Caplin et al. (2018), who produce a reduced-form representation of rational inattention by drawing a link between the Psychometric Weber curve and the theory of competitive supply, and show how basic microeconomic tools can be used to quantify attention costs. We complement these papers by showing how theories of costly attention can be linked to the reduced-form models used by empirical economists to study opaque prices.

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 demanding 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 misreaction varies with the stakes, we develop

[3] 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.

a new experimental design in which the size of the tax rate is varied exogenously within consumers over time. Existing empirical work, such as that of Chetty et al. (2009), Feldman and Ruffle (2015), Feldman et al. (2018), or Taubinsky and Rees-Jones (2018) does not feature within-consumer variation in stake size, and thus cannot be used to test these predictions.

Our experiment features 1,534 consumers who approximate the U.S. adult population on age, gender, and income, and who are drawn 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 type 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 3×3 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 purchased stores, and purchased products are shipped to their homes.

We begin our analysis in Section 4 by computing average underreaction to taxes of varying size, exploiting both the exogenous variation in prices and the exogenous variation in tax rates. We find striking evidence for the prediction that misreaction depends on stakes. The average valuation weight is 0.23 for the smallest price at standard tax rates, and is 0.79 for the largest price at triple tax rates. 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 move to reduced-form tests of predictions about individual differences. As we discuss in that section, a key challenge for tests of individual differences is that it is not possible to compute estimates of valuation weights at the *individual* level with a reasonable degree of precision, since individual-level valuations of the products can only be measured with substantial noise. These challenges are not unique to our setting, and would pose problems for most within-subject experiments seeking to quantify individual differences.[4] 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 form "leave-out instruments" for consumers' attention. We use one product to divide consumers by whether their willingness to buy dropped significantly or not in response to the standard tax rate, and we then estimate those two groups' valuation weights using decisions in the other two products.

These instruments are powerful, which is consistent with significant individual differences. Consumers in the high valuation weight group have an average valuation weight of 1.04 (95% CI 0.83-1.26) for standard taxes, while consumers in the low valuation weight group have an average valuation weight of 0.25 (95% CI 0.08-0.42) for standard taxes. Consistent with the prediction that individual differences are persistent across stakes, we find that in the triple tax store, consumers classified as having

[4]This includes many experiments in which authors choose to report individual-level estimates nonetheless.

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).

Using this methodology we 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 in the low valuation weight group exhibit a significantly larger increase in their valuation weights than consumers in the high valuation weight group (0.39 vs. 0.16; 95% CI for difference 0.03-0.43). Second, when we construct analogous instruments that divide consumers into groups based on how much they adjust their valuation weight as stakes increase, we find that consumers in the low adjustment group have significantly higher valuation weights in both the standard tax regime (0.85 vs. 0.34; 95% CI for difference 0.28-0.75) *and* in the triple tax regime (0.86 vs. 0.76; 95% CI for difference -0.01-0.20). 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 three additional 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? And finally, how large is the variance of the valuation weights, which Taubinsky and Rees-Jones (2018) show is a key input in efficiency cost calculations?

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 mis-measured right-hand-side variables in regressions (e.g., Hausman, 2001; Gillen et al., forthcoming). 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 taxes, the variance of individual differences is at least 0.83 (5% confidence bound of 0.52), and that 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.[5] And 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 inattention (Gabaix et al., 2006; Bartos et al., 2016; Martin, 2016; Dean and Neligh, 2018; Ambuehl et al., 2018; Caplin et al., 2018). With the exception of Bartos et al.

[5]Taubinsky and Rees-Jones (2018) use observable covariates to estimate a lower bound on heterogeneity, but the results in our paper show that their methodology produces a lower bound that is about an order of magnitude off, and cannot be used to establish that some consumers overreact.

(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 and policy relevant domain of behavior, and asking whether the "mistakes" identified by reduced-form empirical work in that domain fit the patterns of costly attention models. In that sense, our paper is most similar in spirit to Bartos et al. (2016), who conduct experiments testing whether discrimination in the job market and rental housing market conforms to patterns predicted by a costly attention model.[6]

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. Taubinsky and Rees-Jones (2018) provide partial evidence that average underreaction decreases with stakes, but they are underpowered to quantify how average underreaction varies by posted price, particularly in the standard tax regime. Consequently, they cannot rule out that their results are driven by other possibilities such as consumers over-reacting to a surprising change that violates their shopping "norms" (Bordalo et al., 2017). More importantly, the lack of within-consumer variation in tax rates in the Taubinsky and Rees-Jones (2018) data makes testing core predictions about individual differences in how attention responds to stakes infeasible.

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., forthcoming), we introduce techniques for bounding 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 could be used to provide formal statistical evidence for other questions about individual differences, such as whether some individuals are risk-loving or future-biased.

As we discuss in Section 7, a better understanding of the mechanisms of misreaction to opaque incentives informs policy implications and improves predictions about market structure. In that section, we also discuss limitations and future directions.

2 Theoretical framework for hypothesis development

In this section, we provide several formal examples of costly attention models based on two prominent modeling approaches in the literature, which we use to motivate our hypotheses. The several variations we provide are far from exhaustive of the full set of costly attention models. Our goal is to show that our hypotheses can be derived from tractable and generally applicable models, and to give some indication of the robustness of our predictions by showing that they are not tied to a single modeling framework. A fully exhaustive characterization of all possible costly attention models is beyond the scope of this paper.

[6] See also Hoopes et al. (2015) and Coibion and Gorodnichenko (2015) for tests of costly information acquisition in observational data.

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 $vx - p$, 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 σ is a parameter that is known to the consumers and represents the "stakes" involved, while q_o is the part that may be mis-processed. 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 σ. 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.[7] 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

[7] 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.

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. 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. Processed-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.

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. After using this example to draw out the main intuitions, we present two richer and more commonly used models of attention choice—the Shannon model used in rational inattention models (Sims, 2003; Matejka and McKay, 2015; Caplin et al., forthcoming) and Gabaix's (2014) attention weight adjustment model—and show that they deliver similar sets of testable implications. The simple model in this section is a special case of Gabaix (2014).

In this simple example, we suppose that computing the opaque price correctly is a binary decision: consumers can rely on their initial perceptions or they can pay a cost λ 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.

Summing up, 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

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

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

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

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 θ that is independent of the price p_s. We call θ 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 θ. Let Δp 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 θ—generated by individual differences in priors ($\hat{t}$ and r) and in the cost of attention λ—must be persistent across stakes σ. In particular, θ is increasing in $\hat{t}$, and $|1-\theta|$ is decreasing in r and increasing in λ.

Second, θ is increasing in σ when $\hat{t} < t$, and is decreasing in σ when $\hat{t} > t$, with $\lim_{\sigma\to\infty}\theta = 1$. 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 σ.[8]

The fact that $|1-\theta| \to 0$ as $\sigma \to \infty$ 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, θ falls with σ for individuals who overreact. More generally, this implies that the extent to which θ increases with stakes σ is decreasing with the baseline level of θ. Finally, if some individuals overreact, the individuals whose θ fall the most as σ increases from σ_1 to $\sigma_2 > \sigma_1$ should on average have the highest θ at both σ_1 to σ_2.

Although mathematically straightforward, this last prediction is particularly demanding. In essence, it is saying that if the distribution of θ at two stakes levels σ_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 Δ. 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."

[8] For $\hat{t} < t$, for example,

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

2.3 Shannon model with heterogeneous priors

In contrast to our simple example above, 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 π over signals and q_o. Without loss of generality, we associate each signal with a posterior belief ρ 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 π 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 π is $c(\pi) = \lambda \left(H(r) - E_\pi[H(\rho)]\right)$, 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 ρ iff $v - p_s - \sigma(\rho t + (1-\rho)\hat{t}) > 0$. We will use $b(\rho) \in \{0, 1\}$ to note whether it is optimal for a consumer to buy given ρ.

4. Consumers thus choose π 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. (forthcoming).

Proposition 1. *For each triplet $\Xi = (\lambda, r, \hat{t})$ and stakes σ in the Shannon model, there exists a distribution $F_{\Xi,\sigma}$ such that the behavior of all consumers with parameters Ξ 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. $\lim_{\sigma\to\infty} F_{\Xi,\sigma} \xrightarrow{d} 1$. 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 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 $\theta = \bar{\theta} + \nu$, where ν 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 ν. 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 $\bar{\theta}+\nu$ (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 1 shows that when stakes are large, the stable components $\bar{\theta}$ converge to 1 and the stochasticity vanishes (i.e., ν converges to 0 in distribution). The last part of Proposition 1 also shows that there will be stable individual differences in θ 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 under-estimation, the Shannon model predicts that a large increase in stakes lowers θ for some consumers and increases θ for other consumers, in line with our binary attention example.

2.4 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.[9] 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 λm^{α}, 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 $B = \min\left((1-r)(v-p_s-t), r(p_s+t-v)\right)$. Consumers then approximate the benefit of choosing $m \in [0,1)$ by the quadratic approximation $B-(1-m)^2B$.

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 2. *For each triplet $\Xi = (\lambda, r, \hat{t})$ and stakes σ in the Gabaix (2014) model, there exists a $\theta_{\Xi,\sigma} \in \mathbb{R}$ such that consumers with parameters Ξ 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 σ and converges to zero as $\sigma \to \infty$. 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.

[9] 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.

Like Proposition 1, Proposition 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 θ 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.

2.5 Extensions

While we have restricted our analysis to binary priors to simplify exposition and amplify intuition, the results hold more generally. In the appendix we consider priors given by $\hat{t} + \varepsilon$, where $E[\varepsilon] = 0$ and $\hat{t}$ varies. When $\hat{t}$ is equal to the true value t, consumers behave as if $\theta = 1$. However, $\hat{t} < t$ generates underreaction and $\hat{t} > t$ generates overreaction. As before, increasing stakes decreases relative misreaction.

2.6 Empirical tests of costly attention theories

Our theoretical results motivate five empirical tests, following the intuition provided in the special case described in Section 2.2. 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 σ) and a "high stakes regime" (higher value of σ). All of the empirical tests are grounded in the core idea that individual differences persist across stakes, and that the revealed valuation 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 θ in the standard stakes regime will also have higher values of θ in the high stakes regime.*

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

Prediction 4. *Consumers whose θ increases the least in response to the high stakes regime have the highest values of θ 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 θ downward when put in the high stakes regime.*

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.

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; Heidhues et al., 2017), as well as frameworks 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.[10]

3 Experimental design

3.1 Overview

The experiment had three parts: (1) instructions and comprehension questions, (2) nine shopping decisions, and (3) end-of-study survey questions. Decisions were incentivized: study participants had a chance to receive a \$16 shopping budget to actually enact their purchasing decisions, and they received any products purchased. Participants retained any unspent portion of the budget.

Each consumer was randomly assigned three of nine household products 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 shopping decisions was randomized within subject.

3.2 Recruitment

The experiment was implemented 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. This

[10] 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., 2015) 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 price 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. (2015) 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.

platform is frequently used by firms that ship products to consumers to elicit product ratings, but is additionally available to researchers for academic use.[11] Two key features of this platform made it appropriate for our experimental design. First, ClearVoice provided samples that match the U.S. population on basic demographic characteristics. Second, ClearVoice maintained an infrastructure for easily shipping products to consumers, which facilitates an incentive-compatible online shopping experiment.

Our experimental design used language referring to the sales tax rate that study participants pay in their city of residence. To avoid confusion, we asked ClearVoice to only recruit panel members from states 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.[12] 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.[13]

3.3 Shopping Decisions

Each purchase decision appeared on a separate screen. Figure 1 provides an example of decisions screens participants would see for each purchase decision. Consumers first saw an image telling them which store they were entering and for which product they were shopping. Consumers were then shown a picture and a product description drawn from Amazon.com, along with a price list containing ten prices. These ten 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.[14] Consumers were asked to select the prices at which 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 shown the following message:

> 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

[11] For other economic research using ClearVoice Research, see Benjamin et al. (2014), Taubinsky and Rees-Jones (2018), and Rees-Jones and Taubinsky (forthcoming).

[12] 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.

[13] Local tax rate data is drawn from the September 2016 update of the "zip2tax" tax calculator.

[14] 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.

answers.

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 J lists the nine products, their Amazon.com prices, and their Amazon.com product descriptions.

3.4 Incentive Compatibility

Decisions in the experiment were incentive compatible. All study participants who passed the necessary comprehension questions (described below) had a 1/3 chance of being selected to receive a \$16 budget. 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. Consumers who were selected to receive the \$16 budget had one tax environment and one product randomly chosen. Outcomes were determined by randomly selecting one of the prices on the price list. 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. Consumers additionally received $16 - p_s(1 + \tau)$ dollars, where τ is the experimentally induced tax rate. The product was shipped to the consumer by ClearVoice, and the remainder of the budget was included in experimental compensation. Participants received a full explanation of the 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."

3.5 Comprehension Questions

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.[15] 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 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 the comprehension checks, so as not to confound comprehension of study rules with actual attention costs.

3.6 Survey 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.

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.

3.7 Sample

1,846 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 47 respondents who had monotonicity violations within a price list. Our main results in Section 4 hold when including 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 F 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 1,534 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 $49,000, an average income of $60,838, and an interquartile range of $32,000-$59,000. 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. 50% of our sample is female.

[15]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. Figure J.3 contains a screenshot of these questions.

8.2% of participants are high school graduates and 35.3% are college graduates. 55.1% of participants are married or in a domestic partnership, and 80.8% live with at least one other individual. The mean sales tax rate charged in participants' city of residence is 7.24% (median 7.00%), with a standard deviation of 1.26%.

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 Summary of behavior

Figure 2 provides a summary of the demand curves as a function of the posted prices. To construct the figure, we start by constructing demand curves $D_{jk}(p)$ where j indexes products and k indexes the store type, A, B, or C. These correspond to the fraction of consumers willing to buy product j at each price on the price list.

Because there are nine products in the study in total, we summarize the data by plotting the average demand curves $D_{avg,k}(p) := \frac{1}{9}\sum_j D_{jk}(p)$ as a function of the transparent price p for each tax-environment (we omit the subscript s for to ease notation). 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 under-react 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. Formally, let p_n denote the nth 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 $\tilde{D}_{jB}(p_n) := \sum_i \left[\frac{\tau_i}{0.15}D_{ijA}(p_n) + \frac{0.15-\tau_i}{0.15}D_{ijA}(p_{n+1})\right]$, where τ_{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.[16] 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 $\tilde{D}_{jC}(p_n) := \frac{3\tau_i-0.15}{0.15}D_{jA}(p_{n+1}) + \frac{0.30-3\tau_i}{0.15}D_{jA}(p_{n+2})$. 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 willing to purchase at price $1.15p_{10}$.

Panel (b) reports the results for the standard tax environment, and panel (c) for the triple tax environment. Comparing the counterfactual demand to the observed demand in the same store, we see evidence of under-reaction. The under-reaction is particularly noticeable at low posted prices.

4.2 Estimating average revealed valuation weights

Recall that the definition of the revealed valuation weight θ_{ijk} for consumer i considering product j in store $k \in \{A, B, C\}$ is that the consumer is θ_{ijk} as responsive to a change in the tax as he is to

[16] All sales tax rates in our sample are less than 15%.

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_{ij} is such that $\ln v_{ij} \geq \ln p + \ln(1 + \theta_{ijk}\tau_{ik})$. 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

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

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

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

where Φ is the standard normal CDF. By allowing both α_j and σ_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. We also allow for separate coefficients on the single tax and triple tax terms by separately measuring β^B and β^C.

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 θ is heterogeneous within each store. In Appendix C we verify that this approximation error is negligible, and that it works against the results that follow about how the 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 τ (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 σ, and our models predict that the revealed valuation weights should increase in σ.

Figure 3 plots $E[\theta|p \leq p^\dagger]$ against a price cutoff $p^\dagger$, such that all prices less than or equal to the cutoff 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 3 are reported in Appendix D.

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 θ in the standard tax store is 0.48 (95% CI 0.32-0.63), and the average valuation weight θ 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 τ. As is immediately evident from figure 3, 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 0.32-0.45) 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 somehow 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. 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, or because the surprising and unusual environment simply draws more attention to itself (Bordalo et al., 2017).

Although our results on individual differences in reaction to the tripling of the tax rate generate a number of additional tests of costly attention, we present one more test on 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 responds to a tripling of the tax rate because it triggers tax aversion, or because it draws attention to an unusual environment, 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.

To do this, we divide the ten different prices in each store into five pairs of prices, for a total of $5 \times 2 = 10$ pairs. For each pair, we estimate the average valuation weight using an extension of model (4) with a separate $\bar{\theta}$ parameter for each pair.[17] We plot this against the average tax owed in each pair: for two adjacent prices p_l and p_{l+1}, we compute the average tax owed in stores B and C, respectively, as $E\left[\frac{p_l+p_{l+1}}{2}\tau_{ik}|k=B\right]$ and $E\left[\frac{p_l+p_{l+1}}{2}\tau_{ik}|k=C\right]$.

[17] Specifically, we partition the prices into pairs P_n, $n = 1, \ldots, 5$ and estimate

Figure 4 presents the results (see Appendix D for exact point estimates and confidence intervals). We see no trend break between the two series. If anything, the deviation in the leftmost point in the store C series has the opposite sign predicted by concerns of tax aversion or unusual environments triggering a spike in attention. Figure 4 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

4.4.1 Order effects

A potential concern with our experimental design is that purchase decisions could be influenced by the order in which the nine purchase decisions are presented to consumers. In Appendix G we test for four potential order effects. First, we examine whether the tax environment (i.e., no tax, standard tax, or triple tax) first shown to consumers impacts their buy probability. Second, we test whether the tax environment first shown to a consumer for product j impacts the likelihood of purchasing product j. Third, we test whether the tax environment last shown to a consumer for product j impacts the likelihood of purchasing product j. Finally, we test whether the specific store ordering for product j impacts the buy probability for product j.

We test the impact of the potential order effects by adding indicators for the orderings to the regression specified in equation (4). We run four regressions, detailed in Appendix G, corresponding to four different order effects concerns outlined above. The resulting coefficients and standard errors provide strong evidence that the order in which consumers see the tax environments does not affect their decisions.

4.4.2 Re-including confused participants

Recall that we excluded 16% of our respondents because they were not able to correctly answer the comprehension questions about the tax rates charged in stores A, B and C. Because misunderstanding study instructions is unlikely to be an externally valid mechanism for misreacting to sales taxes, we exclude them from our primary analysis. At the same time, because these consumers may have the highest attention cost, we can view estimates that include these consumers as an upper bound on misreaction. As shown in Appendix F, re-inclusion of these participants increases the estimate of average underreaction, but does not change any of the comparative statics.

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

4.5 Mechanisms for underreaction

4.5.1 Incorrect beliefs

Do participants know their true sales tax rate, and if not, are incorrect beliefs a mechanism driving the results? To understand the role of incorrect beliefs, we asked participants to report their sales tax rate in their current city of residence. We find that participants generally have correct beliefs: 50.8% of our sample know their tax rate exactly, 70.3% within 0.5 percentage points, and 82.0% within one percentage point.[18] In addition to finding high knowledge among most participants, we also do not find any evidence of systematic underestimation of tax rates. The mean of participants' estimates of their sales tax rates is 7.43%, 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. Figure E.1 in Appendix E repeats figure 3 using the nearly-accurate beliefs subsample defined above. We again find strong evidence for Prediction 1. The average revealed valuation weights increase monotonically in the absolute size of the tax. The estimates are also of similar magnitude to the full sample results. Using all prices we estimate an average revealed valuation weight of 0.55 (95% CI 0.38-0.73) for the standard tax environment in the restricted sample compared to the 0.48 (95% CI 0.32-0.63) in the main sample. Similarly, we estimate an average revealed valuation weight of 0.87 (95% CI 0.78-0.95) for the triple tax environment in the restricted sample, which is only slightly higher than the estimate in the main sample (0.79, 95% CI 0.72-0.86).

4.5.2 Computational ability

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. 44.7% 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.[19]

Figure E.2 in Appendix E repeats figure 3, restricting to the 62.9% of participants who are able to compute their true tax burden within $0.05. We again find strong evidence for Prediction 1. The estimates are also of similar magnitude to the full sample results. Using all prices we estimate an average revealed valuation weight of 0.52 (95% CI 0.36-0.68) for the standard tax environment in the restricted sample compared to the 0.48 (95% CI 0.32-0.63) in the main sample. Similarly, we estimate an average revealed valuation weight of 0.82 (95% CI 0.75-0.89) for the triple tax environment in the restricted sample, which is only slightly higher than the estimate in the main sample (0.79, 95% CI

[18] 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.

[19] 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 258 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.

0.72-0.86).

The fact that complete inability to compute taxes does not seem to be driving our results is consistent with a key premise of our theoretical framework: that consumers are capable of correctly incorporating taxes into their decision. The fact that incorrect beliefs about the tax rates do not seem to be a major factor is consistent with the interpretation of our theoretical framework as capturing an internal process of forming tax rate perceptions. That is, consumers are not literally using costly mental effort to acquire information about the tax rates; rather, they are using mental effort to go through the computation of figuring out the post-tax price.

4.5.3 Other correlates

Appendix H 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. The appendix also shows that we are underpowered for detecting meaningful differences in how average underreaction covaries with differences in local tax rates. This is not surprising, as the standard deviation of the tax rates applied to consumers' city of residence is only 1.26 percentage points—variation that is over an order of magnitude smaller than our exogenously induced experimental variation in tax rates. Although this correlation may be interesting to explore in better-powered designs, we note that it is likely a confounded test of costly attention models because local tax variation is likely related to a number of differences in geography, including consumers' views and preferences about tax rates.

5 Reduced-form results on individual differences

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 θ, 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 θ. We use this methodology to provide evidence for Predictions 2-4.

5.1 Methodology for testing predictions 2 and 3

A key feature of our design is that consumers make purchase decisions for the same product in different tax environments. We can thus construct individual-level estimates $\hat{\theta}_{ijk}$ for θ_{ijk} for each consumer i, product j, and store k combination via a two step process. First, we approximate the maximum pre-tax price p^*_{ijk} at which consumer i is willing to buy product j in store $k \in \{A, B, C\}$. Since we only observe purchase decisions at ten pre-tax prices, we approximate willingness to pay using the midpoint of each price interval. Formally, $\ln p^*_{ijk} = 0.5\left(\ln p^0_{ijk} + \ln p^1_{ijk}\right)$, where p^0_{ijk} is the highest price at which consumer i buys product j in store k and p^1_{ijk} 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), recall that we utilized a non-incentivized question that solicited the maximum price at which they would be willing to buy the product. Given the free-response and unincentivized nature of this question, some of these responses are very high. To reduce the impact of outliers, we therefore compute the median self-reported maximum pre-tax price by product and store among those censored, and assign this value to p^*_{ijk}. Using the buying condition in equation (3), we construct the estimate $\hat{\theta}_{ijk}$ for $k \in \{B, C\}$ as

$$\hat{\theta}_{ijk} = \frac{\ln(p^*_{ijk}) - \ln(p^*_{ijA})}{\ln(1 + \tau_{ik})} \tag{5}$$

However, we cannot directly use the $\hat{\theta}_{ijk}$ estimate to compute properties of the actual distribution of θ without making the unrealistically strong assumption that all within-person differences in choices between stores load on the θ_{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.[20] 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.

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, 24.6% 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 19.6% 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 θ would imply substantially negative θ 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. (forthcoming), it is prevalent in just about all experimental analysis of individual differences. Our finding is also in line with a long intellectual history of measuring and modeling noise in individual's decisions, starting from Block and Marschak (1960), continuing with Quantal Response Equilibrium (McKelvey and Palfrey, 1995), and recently gaining prominence in explicit modeling of mistakes (e.g., Woodford, 2012; Khaw et al., 2017; Natenzon, 2019).[21]

We instead use the $\hat{\theta}_{ijk}$ estimates to create proxies for high versus low valuation weight consumers.

[20] 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. Consequently, the BDM poses the same problems as a discrete set of prices, and comes at the additional cost of being more confusing.

[21] While the change in perceptions of product value may have deep cognitive foundations that can be modeled, this behavioral tendency is orthogonal to the estimation of the revealed valuation weights θ, and neither confounds our analysis nor reinforces it.

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 use our empirical model in (4) *on the other two products* to estimate average valuation weights for the low and high groups. Figure 5 illustrates the idea of this procedure. Using this system of dividing consumers into high and low valuation weight groups according to store B behavior, we test Predictions 2 and 3 by examining the valuation weights of the consumers in those groups in both stores B and C.

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. We define $x_{i1}^{75} = I\left[F(\hat{\theta}_{i1B}) > 0.75\right]$ to be an indicator for the high group. We then use decisions in 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 an overall average estimate of θ_{ijk} for those in the high and low groups:

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

We compute compute confidence intervals using percentile bootstrap, clustering by subject.

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

Assumption. $\left(\hat{\theta}_{ijk} \perp \hat{\theta}_{ij'k'}\right)|\theta_{ijk}$ 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 classical measurement error), or even that they are orthogonal to the true underlying θ_{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 and perform our actual estimation of $E[\theta_{ijk}]$ using only incentivized decisions.

5.2 Methodology for testing prediction 4

We also use the procedure in Section 5.1 to analyze heterogeneity in how the valuation weights respond to stakes. This provides a test of Prediction 4. Specifically, we divide consumers according to how much they adjust their valuation weight when the tax rate increases, rather than by how attentive they are to the tax.

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 $d_{ij}^{25} = I\left[F(\hat{\Delta}_{ij}) \leq 0.25\right]$ 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 $j' \in \{1, 2, 3\}$ and for each store $k \in \{B, C\}$. We then average to estimate $E[\Delta_{ij}|d_i^{25}]$ separately for the high and low adjustment groups.

5.3 Results: Predictions 2 and 3

Table 1 presents the $E[\theta_{ijk}]$ estimates using the instruments defined in Section 5.1. Rows (1) and (2) present estimates of $E[\theta_{ijk}]$ for the high and low valuation weight groups respectively, and row (3) presents estimates of the difference. Columns (1) and (2) correspond to the different tax environments, and column (3) to differences between the environments.

Consumers in the high valuation weight group have an average revealed valuation weight of 1.04 (95% CI 0.83-1.26) 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. Consistent with Prediction 2, these individual differences are 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). The predictive power of the instruments implies strong individual differences in the standard tax environment, and hence heterogeneous revealed valuation weights.

We next examine individual differences in adjustment. Consistent with Prediction 3, the low valuation weight group exhibits a significantly larger increase in the valuation weights than the high valuation weight group when tax rates are tripled (0.39 vs. 0.16; 95% CI for difference 0.03-0.43).

5.4 Results: Prediction 4

Using the methodology described in Section 5.2, we compare $E[\theta_{ijk}]$ estimates from the high adjustment group and the low adjustment group. 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 Δ_{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 0.28-0.75) *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 θ_{ijB} and θ_{ijC} took the form $\theta_{ijC} = a_0 + \theta_{ijB} + \varepsilon_{ij}$ for some constant a_0 and some random variable ϵ_{ij} independent of θ_{ijB}, then $E[\theta_{ijC}|\theta_{ijC} - \theta_{ijB} = \Delta]$ would be increasing in Δ, not decreasing. Intuitively, small values of Δ would imply a small indiosyncratic component, and thus a smaller value of θ_{ijC}. In general, any stochasticity in $\theta_{ijC} - \theta_{ijB}$ that is independent of the value of θ_{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.

5.5 Robustness

In Appendix E, 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. These results also conform with Predictions 2-4. In Appendix F we confirm that the results hold for the full sample of participants, including those failing comprehension checks.

In Appendix I, we verify our results still support these theoretical predictions when using alternative constructions of the proxies (tables I.1, I.2, I.3, I.4).

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) show is a key input in efficiency cost calculations?

In this section we develop novel econometric techniques for computing lower bounds on individual heterogeneity, which enable us to answer the three questions above.

6.1 Intuition - strength of the instruments

A key moment that we have not exploited in the analysis in the previous section is how well correlated the binary instruments are with each other. In the subsections that follow, we will show how this moment, combined with the results in Section 5, helps generate a lower bound on the variance of θ. The lower bound on the variance of θ 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 6.2 below.

To obtain intuition about the importance of the correlation between the instruments, consider a linear regression of a variable Y on a binary proxy X. The regression coefficient is given by $\beta = Cov[Y,X]/Var[X]$, which can be rewritten as $Corr[Y,X] \cdot \frac{\sqrt{Var[Y]}}{\sqrt{Var[X]}}$. The coefficient β here is essentially equivalent to our reduced-form results above about how the average θ varies with our binary instruments. Now holding β and $Var[X]$ constant, note that the smaller is $Corr[Y,X]$, the larger must be $Var[Y]$. Intuitively, the regression coefficient β can be increased either by scaling up the dependent variable Y, which increases its dispersion, or it can be increased by increasing the correlation between Y and X. Consequently, learning about the regression coefficient and the correlation is informative of the dispersion in Y.

If we regard Y as θ and X as our binary instruments from Section 5, then one indication of the correlation between Y and X is the correlation between our binary instruments. If our instruments are not well correlated with each other, then they also cannot be well correlated with Y.

We now compute the average correlation of the binary instruments utilized in the last section: $\frac{1}{3}\sum_{j\neq j'} Corr\left[x_{ij}^{75}, x_{ij'}^{75}\right]$. The resulting estimate of 0.18 (s.e. 0.02) implies these instruments are weakly correlated with each other. We similarly compute the average correlation for the adjustment instruments, $\frac{1}{3}\sum_{j\neq j'} Corr\left[d_{ij}^{25}, d_{ij'}^{25}\right]$. We estimate an average correlation of 0.14 (s.e. 0.02), indicating that the adjustment instruments are also weakly correlated with each other. The fact that these instruments predict large differences in the average θ as shown in Section 5, but are not well correlated with each other implies large underlying dispersion in θ, as we show below.

6.2 Methodology: Inequalities for dispersion of random variables

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

Proposition 3. *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*

$$Var[Y] \geq \frac{Cov[Y, X_1] \cdot Cov[Y, X_2]}{Cov[X_1, X_2]} \tag{6}$$

and

$$(\bar{Y} - E[Y])(E[Y] - \underline{Y}) \geq Var[Y]. \tag{7}$$

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

Result (6) formalizes the intuition in 6.1: the less well-correlated the proxies 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 line (7) is the Bhatia and Davis (2000) inequality. The intuition for this result 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 3 enables us to use the types of binary instruments utilized in Section 5 to compute bounds on the variance and support of θ_{ij}. Define $x_{ijk}^q = I\left[F\left(\hat{\theta}_{ij} > 0.01q\right)\right]$ as an indicator for $\hat{\theta}_{ijk}$ being in the qth 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 θ_{ijk} is supported on $[0, \bar{\theta}]$, where θ_{ijk} is the revealed valuation weight for product j of individual i in store k. Then given instruments $x_{ij'k}^q$ and $x_{ij''k}^q$ computed for products j' and j'' (with no two of j, j', and j'' equal), the variance of θ_{ijk} in store K is*

$$Var[\theta_{ijk}|k = K] \geq \frac{Cov[\theta_{ijk}, x_{ij'k}^q|k = K] \cdot Cov[\theta_{ijk}, x_{ij''k}^q|k = K]}{Cov[x_{ij'k}^q, x_{ij''k}^q]} \tag{8}$$

and

$$\bar{\theta} \geq E[\theta_{ijk}|k=K] + \frac{Cov[\theta_{ijk}, x^q_{ij'k}|k=K] \cdot Cov[\theta_{ijk}, x^q_{ij''k}|k=K]}{E[\theta_{ij'k}|k=K] \cdot Cov[x^q_{ij'k}, x^q_{ij''k}|k=K]} \tag{9}$$

The assumption that $\theta_{ijk} \geq 0$ is natural, as it is equivalent to assuming that no consumers perceive the tax to be a subsidy. As in Section 5, we continue to examine how an instrument $x_{ij'k}$ computed from product j' covaries with the average valuation weights for some other product $j \neq j'$. We never examine how an instrument computed for product j covaries with the average valuation weights for that same product j—doing so would bias our results because of correlated measurement error.

We can also use Proposition 3 to derive bounds on adjustment $\Delta_{ij} = \theta_{ijC} - \theta_{ijB}$. We define $d^q_{ij} = I\left[F\left(\hat{\Delta}_{ij} < 0.01q\right)\right]$ as a binary indicator for $\hat{\Delta}_{ij}$ being in qth decile or lower. This is analogous to Section 5, where we used $q = 25$.

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

$$Var[\Delta_{ij}] \geq \frac{Cov[\Delta_{ij}, d^q_{ij'}] \cdot Cov[\Delta_i, d^q_{ij''}]}{Cov[d^q_{ij'}, d^q_{ij''}]} \tag{10}$$

and

$$\underline{\Delta} \leq E[\Delta_{ij}] + \frac{Cov[\Delta_{ij}, d^q_{ij'}] \cdot Cov[\Delta_{ij}, d^q_{ij'}]}{(E[\Delta_{ij}] - 1) \cdot Cov[d^q_{ij}, d^q_{ij'}]} \tag{11}$$

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

While these results generate bounds on the supremum of the support, they do not quantify how many consumers overreact. We next derive a bound for the fraction of over-reactors, $Pr(\theta_{ij} > 1)$, 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 B.4, which can be seen as a converse of sorts to Chebyshev's inequality.

Proposition 4. *Assume θ_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 Δ_i is supported on $[\underline{\Delta}, 1]$.Then*

$$Pr(\theta_i > 1) \geq \frac{Var[\theta_i] + E[\theta_i]^2 - E[\theta_i]}{(\bar{\theta} - 1)\bar{\theta}} \tag{12}$$

and

$$Pr(\Delta_i < 0) \geq \frac{Var[\Delta_i] + E[\Delta_i]^2 - E[\Delta_i]}{(\underline{\Delta})(\underline{\Delta} - 1)} \tag{13}$$

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 $\theta_i \in \{0, 1, \bar{\theta}\}$.

6.3 Bounds on the variance and the support: Estimation and results

Estimation: In our empirical implementation of the bounds, we construct instruments x^q_{ijk} 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 (8) is that

$$Var[\theta_{ijk}|k = K] \geq \max_q \left\{ \frac{Cov[\theta_{ijk}, x^q_{ij'k}|k = K] \cdot Cov[\theta_{ijk}, x^q_{ij''k}|k = K]}{Cov[x^q_{ij'k}, x^q_{ij''k}]} \right\}. \tag{14}$$

An analogous extension follows for $\underline{\Delta}$ in (11). In the empirical implementation, we take the maximum over $q \in \{10, 15, ..., 90\}$. To estimate $Cov[\theta_{ijk}, x^q_{ijk'}]$, for each q, we use the expansion

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

with the conditional means $E[\theta_{ij}|x^q_{ij'} = 0]$ estimated as in Section 5.

We analogously generate the bounds for $Var[\Delta_{ij}]$ and $\underline{\Delta}$ by taking the maximum over $q \in \{10, 15, ..., 90\}$.

We calculate bootstrapped percentile-based confidence intervals from 1000 replications, clustered at the subject level, and report the 5% confidence bound.[22]

Results: Table 3 presents our estimates. We estimate a lower bound of 0.83 (5% confidence bound of 0.52) for $Var[\theta_{ijB}]$ and of 0.71 (5% confidence bound of 0.59) for $Var[\theta_{ijC}]$. 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 over-reactors 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 (95% confidence bound of -0.16). This result is consistent with Prediction 5 of our theoretical results.

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

[22] 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.

establish overreaction. The methods in this section show that the alternative approach in Taubinsky and Rees-Jones (2018) generates a lower bound that is almost an order of magnitude off.

6.4 Bounds for the fraction of over-reactors and participants with negative adjustment: Results

By substituting the lower bound for $Var[\theta_{ijk}]$ from equation (14) into the lower bound from equation (12), we derive a bound for the fraction of over-reactors as a function of $\bar{\theta}$.

Figure 6 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 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 these bounds are decreasing in $\bar{\theta}$: for $\bar{\theta} = 4.25$, we bound the fraction of over-reactors at 4.2% (5% confidence bound of 2.0%) of the population in the standard tax environment and at 3.9% (5% confidence bound of 3.0%) in the triple tax environment.

Using equation (13), we derive an analogous bound for the fraction of consumers who adjust their valuation weight downward in response to higher stakes. Figure 7 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%).

6.5 Robustness

In Appendix E, 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 F we confirm that the results hold for the full sample of participants, including those failing comprehension checks.

6.6 Discussion of our methods and alternative approaches

A key advantage of our approach to documenting heterogeneity is that it makes minimal assumptions about measurement error at the individual level, and provides non-parametric bounds once the conditional means $E[\theta_{ijk}|x^q]$ are given. The key assumption underlying this approach, stated in Section 5.1, is that $\left(\hat{\theta}_{ijk} \perp \hat{\theta}_{ij'k'}\right)|\theta_{ijk}$ 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 Taubinsky and Rees-Jones (2018), would lead to correlated measurement error and violate our assumption. Consequently, our approach is not applicable to experimental datasets such as those in Taubinsky and Rees-Jones (2018) or Feldman et al. (2018).

Another key feature of our approach is that it enables formal hypothesis tests about the support of valuation weights (e.g., whether some consumers overreact). Alternative approaches that rely on shrinkage estimators or deconvolution methods to estimate the distribution of a variable measured with error do not permit formal hypothesis tests of this kind.

7 Concluding remarks

In this paper, we provide some of the first tests of costly attention models in a concrete and policy-relevant setting. By focusing on the concrete setting of opaque sales taxes, and opaque prices more broadly, our paper also contributes to the empirical work in those settings. A better understanding of the mechanisms in these concrete settings 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 (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 other 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.

Of course, the fact that we find evidence for costly attention models does not preclude the possibility that other mechanisms could be at play. For example, some consumers might simply forget about sales taxes when shopping at a store. While this possibility seems less plausible in our specific setting, since consumers are reminded what store they are entering right before they make their decisions (see figure 1), it could be a more important factor in other shopping environments. Similarly, while we do not find that systematically incorrect beliefs explain underreaction in our experiment, they may play a more important role in other settings, such as the applications discussed in the shrouded prices models of Gabaix and Laibson (2006) and Heidhues et al. (2017).

There may also be important interactions between costly attention models and other mechanisms. One might expect the costly attention models to look better in our environment than in ones where misspecified models of the world are likely to play a more important role. 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. When considering other domains of behavior, simple tests of knowledge and computational skills, such as ours, could help indicate the extent to which mechanisms other than costly attention are a key force.

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 of the importance of costly attention mechanisms in a variety of other economically important settings.

References

Abaluck, J. and A. Adams (2017): "What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses," Working Paper 23566, National Bureau of Economic Research.

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

Ambuehl, S., A. Ockenfels, and C. Stewart (2018): "Attention and Selection Effects," *working paper.*

Bartos, V., M. Bauer, J. Chytilová, and F. Matêjka (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.

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.

Block, H. and J. Marschak (1960): "Random Orderings and Stochastic Theories of Response," in *Contributions to Probability and Statistics. Essays in Honor of Harold Hotelling*, ed. by I. Olkin, Stanford University Pres.

Bordalo, P., N. Gennaioli, and A. Shleifer (2013): "Salience and Consumer Choice," *Journal of Political Economy*, 121, 803–843.

——— (2017): "Memory, Attention, and Choice," *NBER Working Paper No. 23256.*

Bushong, B., J. Schwartzstein, and M. Rabin (2015): "A Model of Relative Thinking," *Working Paper.*

Caplin, A. (2016): "Measuring and Modeling Attention," *Annual Review of Economics*, 8, 379–403.

Caplin, A., D. Csaba, J. Leahy, and O. Nov (2018): "Rational Inattention, Competitive Supply, and Psychometrics," Working Paper 25224, working paper.

Caplin, A., M. Dean, and J. Leahy (forthcoming): "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," *Review of Economic Studies.*

Chetty, R., A. Looney, and K. Kroft (2007): "Salience and Taxation: Theory and Evidence," *NBER working paper No. 13330.*

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

Coibion, O. and U. Gorodnichenko (2015): "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review*, 105, 2644–2678.

Dean, M. and N. Neligh (2018): "Experimental Tests of Rational Inattention," *working paper*.

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

Feldman, N., J. Goldin, and T. Homonoff (2018): "Raisin 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.

Gabaix, X. and D. Laibson (2006): "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, 121.

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.

Gillen, B., E. Snowberg, and L. Yariv (forthcoming): "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study," *Journal of Political Economy*.

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.

Hausman, J. (2001): "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problemsfrom 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.

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

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

Khaw, M. W., Z. Li, and M. Woodford (2017): "Risk Aversion as a Perceptual Bias," .

Koszegi, B. and A. Szeidl (2013): "A Model of Focusing in Economic Choice," *Quarterly Journal of Economics*, 128, 53–104.

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. Matějka, and M. Wiederholt (2018): "Rational Inattention: A Disciplined Behavioral Model," Working paper.

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.

McKelvey, R. D. and T. R. Palfrey (1995): "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior*, 10, 6–38.

Natenzon, P. (2019): "Random Choice and Learning," *Journal of Political Economy*, 127, 419–457.

Ratcliff, R. (1978): "A Theory of Memory Retreival," *Psychological Review*, 85.

Rees-Jones, A. and D. Taubinsky (forthcoming): "Measuring Schmeduling'," *Review of Economic Studies*.

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," Tech. rep.

Sims, C. A. (2003): "Implications of Rational Inattention," *Journal of Monetary Economics*, 50, 665 – 690.

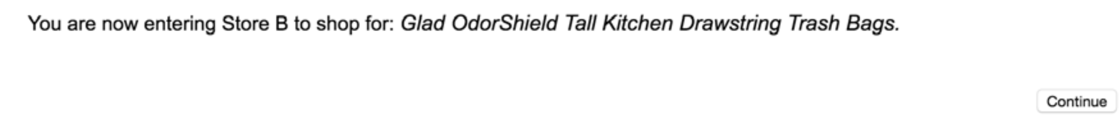
Taubinsky, D. and A. Rees-Jones (2018): "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment," *The Review of Economic Studies*, 85, 2462–2496.

Woodford, M. (2012): "Inattentive Valuation and Reference-Dependent Choice," *Working Paper*.

Figures and Tables

Figure 1: Screenshots of a purchase decision

(a) Introduction

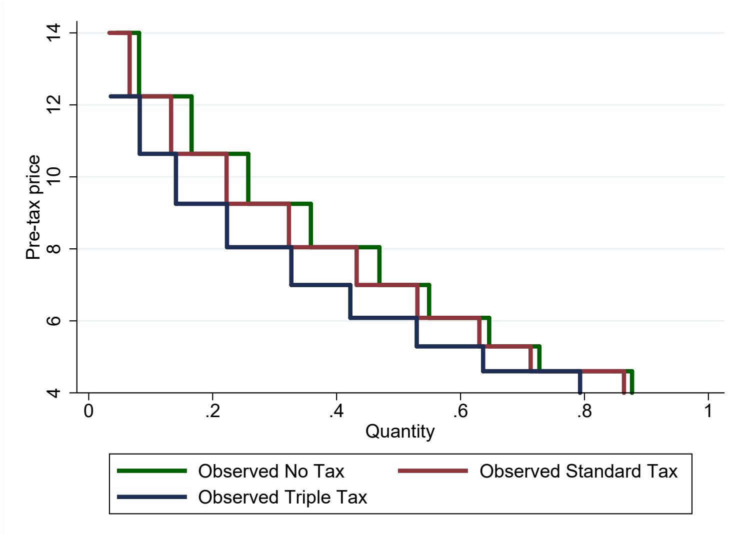


(b) Purchase decision

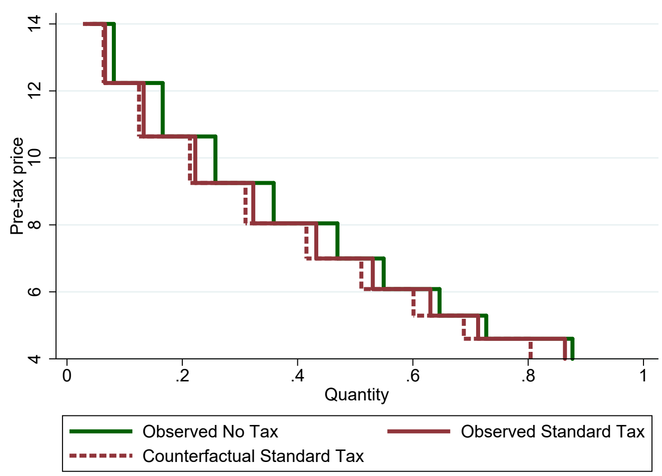
Figure 1 shows an example of the two screenshots participants see for each of their nine purchase decisions. Subjects first saw a screen indicating the product for which they will be shopping and the relevant sales tax environment. Store A corresponds to a tax-free environment, store B to a standard 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.

Figure 2: Demand curves

(a) Observed demand curves



(b) Observed vs. counterfactual demand: standard taxes



(c) Observed vs. counterfactual demand: triple taxes

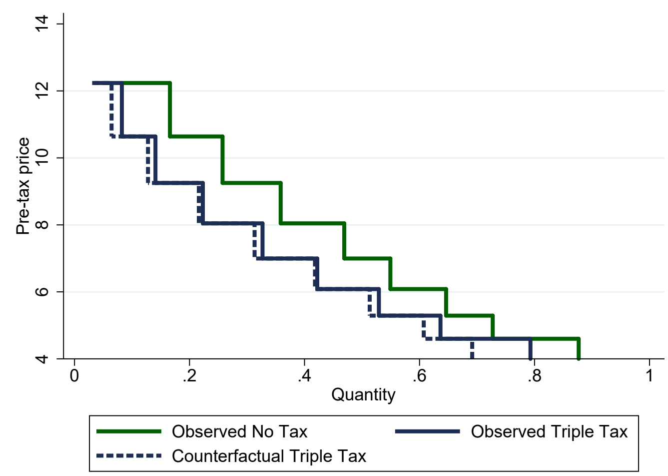


Figure 2 presents demand curves, averaging across all nine products. We start by constructing demand curves $D_{jk}(p)$ where j indexes products and k indexes the store type, A (no tax), B (standard tax), or C (triple tax). These correspond to the fraction of consumers willing to buy product j at each 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 3: Average revealed valuation weight for posted prices at or below a cutoff

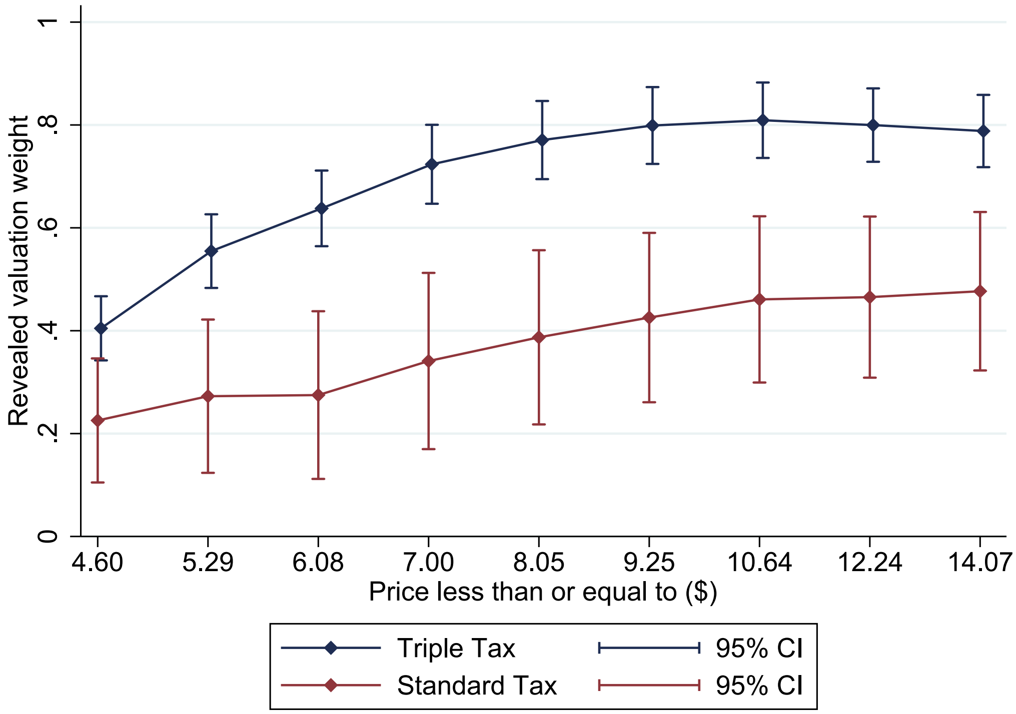


Figure 3 presents estimates of store-specific estimates $E[\theta]$ for prices less than or equal to the cutoff specified on the x-axis. θ is defined as the revealed valuation weight that consumers place on the sales tax, with $\theta = 0$ corresponding to complete neglect of the tax and $\theta = 1$ 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. Standard errors are clustered at the subject level.

Figure 4: Average revealed valuation weight by average tax owed

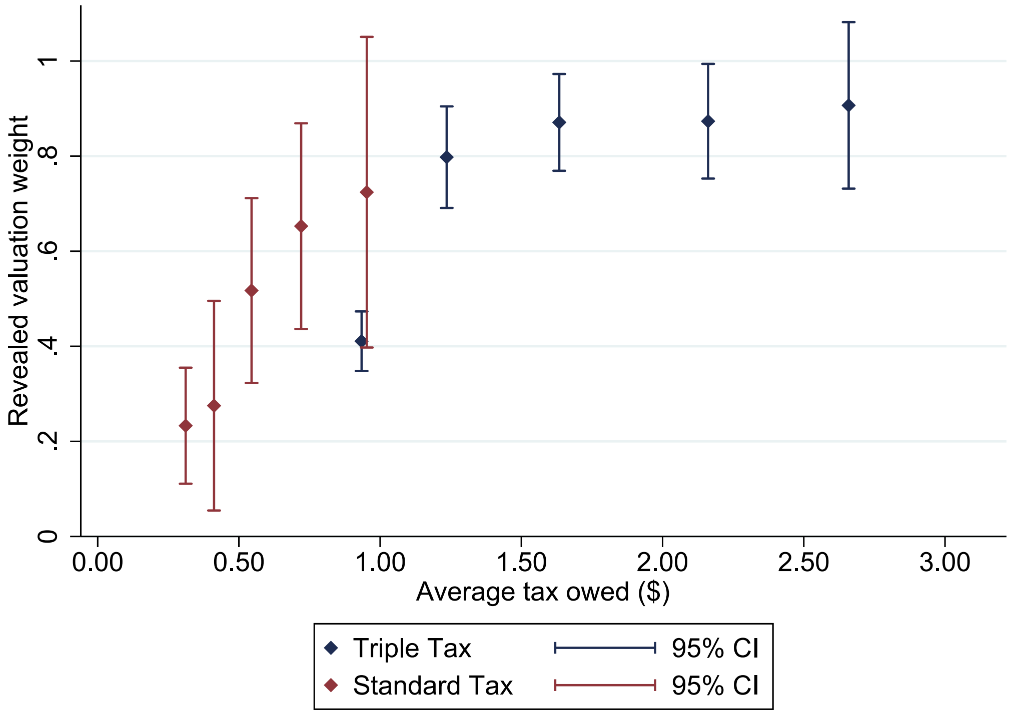


Figure 4 presents store-specific estimates $E[\theta]$ by the average tax owed within each bin. θ is defined as the revealed valuation weight that consumers place on the sales tax, with $\theta = 0$ corresponding to complete neglect of the tax and $\theta = 1$ corresponding to putting the same weight 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 footnote 17. Standard errors are clustered at the subject level.

Figure 5: Illustration of the estimation procedure for individual differences

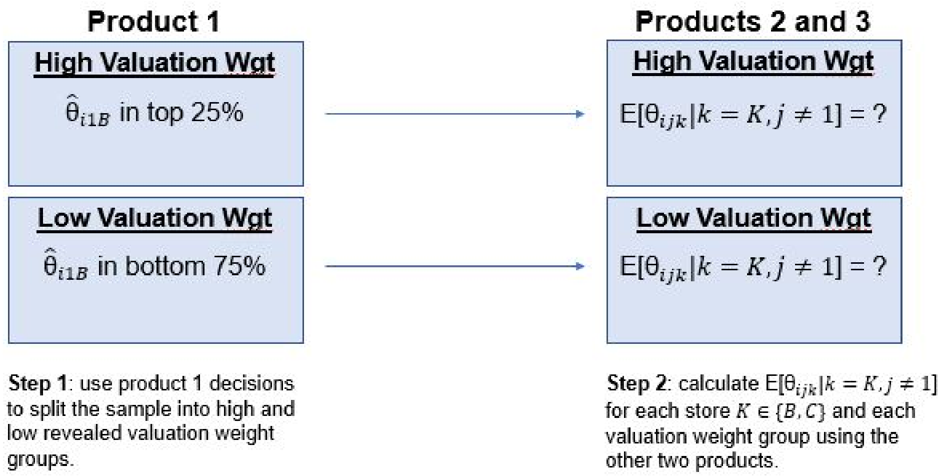
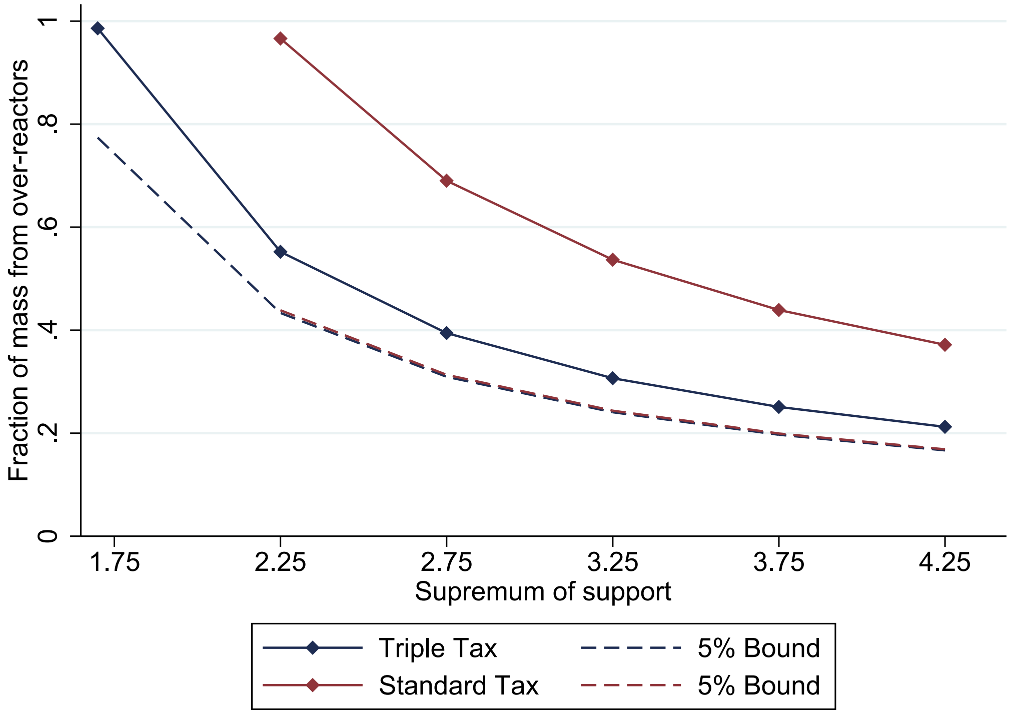


Figure 5 illustrates the estimation procedure for computing $E[\theta]$ for high and low valuation weight groups, as described in Section 5.1. θ is defined as the revealed valuation weight that consumers place on the sales tax, with $\theta = 0$ corresponding to complete neglect of the tax and $\theta = 1$ corresponding to putting the same weight on the tax as on the salient posted price.

Figure 6: Lower bound on the fraction of individuals who over-react to the sales tax



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 θ. θ 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 (12). The dashed lines present the 5% lower bound computed from a percentiled-based bootstrap procedure (1000 replications, clustered at the subject level).

Figure 7: Lower bound on the fraction of individuals whose valuation weight θ decreases in response to higher stakes

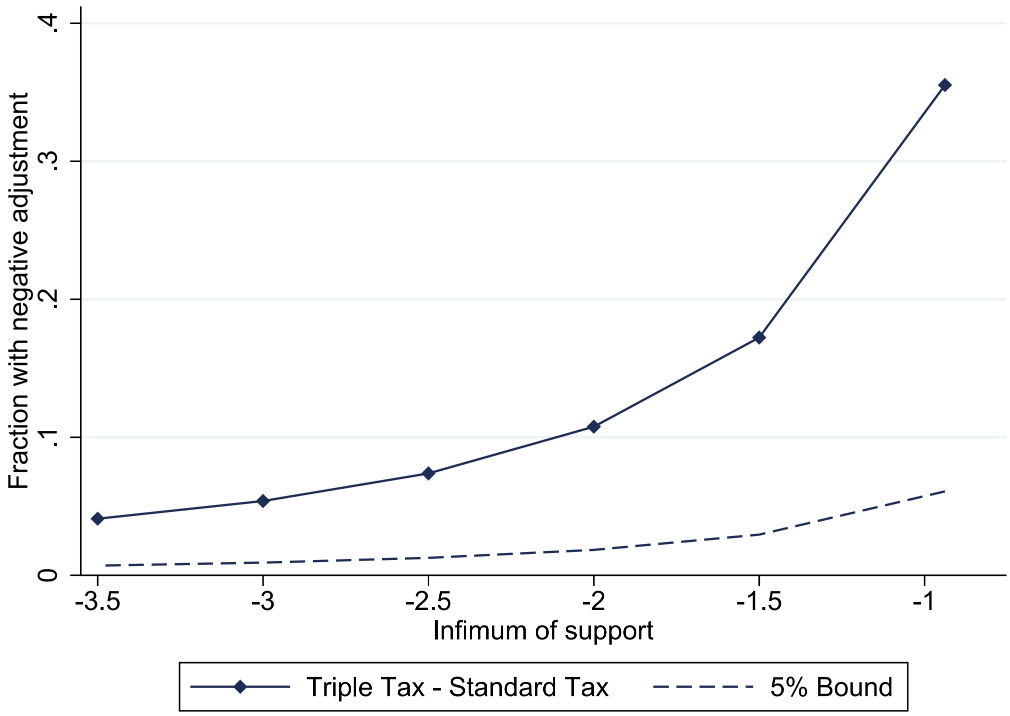


Figure 7 presents estimates for the lower bound on $Pr(\theta_{iC} - \theta_{iB}) < 0$ as a function of the infimum of the support. θ is defined as the revealed valuation weight that consumers place on the sales tax, with $\theta = 0$ corresponding to complete neglect of the tax and $\theta = 1$ corresponding to putting the same weight on the tax as on the salient posted price. The dashed lines present the 5% lower bound from bootstrapped percentile-based confidence intervals (1000 replications, clustered at the subject level).

Table 1: Average revealed valuation weights by group

	Standard	Triple	Triple − Standard
(1): High valuation wgt.	1.04	1.20	0.16
	[0.83, 1.26]	[1.10, 1.31]	[-0.02, 0.33]
(2): Low valuation wgt.	0.25	0.64	0.39
	[0.08, 0.42]	[0.57, 0.72]	[0.26, 0.52]
(3): (1) − (2)	0.79	0.56	-0.23
	[0.52, 1.06]	[0.44, 0.68]	[-0.43, -0.03]

Rows (1) and (2) of table 1 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. 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	0.86	0.01
	[0.64, 1.07]	[0.77, 0.96]	[-0.15, 0.17]
(2): High Adj.	0.34	0.76	0.43
	[0.17, 0.51]	[0.68, 0.84]	[0.30, 0.55]
(3): (1) − (2)	0.52	0.10	-0.42
	[0.28, 0.75]	[-0.01, 0.20]	[-0.60, -0.24]

Rows (1) and (2) of table 2 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. 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 (θ_B)	Triple (θ_C)	$\theta_B - \theta_C$
Variance (Lower Bound)	0.83	0.71	0.86
	[0.52]	[0.59]	[0.31]
Supremum (Lower Bound)	2.21	1.69	0.94
	[1.55]	[1.54]	[0.16]

Columns (1) and (2) of table 3 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 6.2 and the estimating equations are described in Section 6.3. Fifth percentile results from 1000 bootstrap replications, clustered by subject, are reported in brackets.

Online Appendices (Not for Publication)

A General results about costly attention models and preliminary lemmas

A.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 A1. *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. 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 θ 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 θ 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 θ corresponding to G_2 first order stochastically dominates the distribution of θ corresponding to G_1.

3. If $G(p_o - \delta, p_o) = 1 - G(p_o + \delta, p_o)$ for all δ, then the distribution of θ is symmetric about 1 and satisfies $E[\theta] = 1$.

Proof. Fix p_o, and let $F(\theta|p_o)$ be the distribution of θ 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, $E[\theta] = 1$. □

Lemma A1 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.

A.2 Models in the spirit of rational inattention

We consider a model in which $p_o = \sigma\omega$, where σ are the salient stakes, and $\omega \in \Omega$ is the initially unknown state. The set Ω includes the true value q_o. A consumer has a prior μ about ω. 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 $\pi : \Omega \to \Delta(\Gamma)$. The posterior γ 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 γ given state ω. The cost of selecting π is $K(\pi) \in \mathbb{R}^+$, where $\mathbb{R}^+$ denotes the non-negative reals. Given a posterior γ, 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 π is given by $V(\pi) = Q(\pi)(v - p_0) - R(\pi)$, 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 A2. *Let π be the information structure chosen for $u = v - p_s$ and let π' 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 π 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 $u > u'$. □

Lemma A2 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 A3. *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

$$Pr(buy|\omega = q_o) = \frac{Qe^{\frac{u - \sigma q_o}{\lambda}}}{(1 - Q) + Qe^{\frac{u - \sigma q_o}{\lambda}}} \tag{15}$$

Since the right-hand side of (15) is increasing in both Q and u, Lemma A2 implies that $Pr(buy|\omega = q_o)$ is increasing in u. □

The last result immediately leads to the following:

Proposition A1. *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 A3 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 A1 leads to the result. □

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 A2. *Suppose that the prior μ is symmetric around q_o; i.e., $\mu(\omega) = \mu(\omega')$ if $|w - q_o| = |w' - q_o|$. 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 ω 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 ω 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 A1 then implies the result. □

Generalization to other cost functions: In general, Proposition A2 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 A3. *Let the cost function be given by $\lambda(H(\mu) - E[H(\gamma)])$, where H denotes entropy. As $\lambda \to 0$ or as $\sigma \to \infty$, the distribution of θ approaches, in probability, a distribution that places unit mass on 1.*

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

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

Thus $Q_0 := \lim_{\lambda\to 0} Q \geq Pr(\omega \leq v-p_s)$. 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

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

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

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 λ/l. But since behavior approaches the full attentive benchmark when $\lambda\to 0$, 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 λ/k. Then the reasoning in the proof of Proposition A3 implies that any cost function that generates behavior that is continuous in λ at 0 will also generate the prediction that the distribution of θ approaches 1.

A.3 Gabaix (2014) model of attention adjustment

Again, we consider a model in which $p_o=\sigma\omega$, where σ are the salient stakes, and $\omega\in\Omega$ is the initially unknown state. The set Ω includes the true value q_o. The consumer has a prior μ about ω. 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 λm^α, 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,

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

Lemma A4. *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 A1 then implies:

Proposition A4. *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 λ and σ.

Proposition A5. *In the revealed valuation weight representation, $\theta = 1$ if $\bar{q}_o = q_o$. The relative misreaction $|1-\theta|$ is increasing in λ and is decreasing in σ, with $\lim_{\lambda\to 0}|1-\theta| = 0$ and $\lim_{\sigma\to\infty}|1-\theta| = 0$.*

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

$$u^\dagger = m(u^\dagger)\sigma q_o + (1 - m(u^\dagger))\sigma\bar{q}_o \tag{16}$$

Recall that Lemma A4 implies that there is a unique $u^\dagger$ satisfying this equation.

Then

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

Note that θ is a function of m and $u^\dagger$ only, and does not directly depend on stakes. The combination of (16) and (17) imply that $m(u)$ is decreasing in λ and increasing in σ 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 λ for all u, the assumption $q_o > \bar{q}_o$ implies that $q^* = mq_o + (1-m)\bar{q}_o$ is decreasing in λ for all values of u. Consequently, the solution $u^\dagger$ to equation (16) decreases in λ, and thus θ must be increasing in λ as well. Finally, since $m \to 1$ as $\lambda \to 0$, it follows that $\lim_{\lambda \to 0} \theta = 1$.

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

$$B(\sigma', u') = \int_{\omega \geq u'/\sigma'} (\sigma'\omega - u')\mu(\omega) = \frac{\sigma'}{\sigma}\int_\omega (\sigma\omega - u)\mu(\omega) = \frac{\sigma'}{\sigma}B(\sigma, u) \tag{18}$$

Since the perceived benefit of increasing m is linear in B, equation (18) above implies that increasing stakes to σ' has the same impact on m as decreasing attention costs from λm^α to $\frac{\sigma}{\sigma'}\lambda m^\alpha$. Thus, since m is decreasing in λ, it must be increasing in stakes σ. $\square$

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 θ. We show that for a family of distributions of prior beliefs indexed by the mean and the variance, the revealed valuation weight θ will be increasing in the mean and in the variance.

Proposition A6. *Suppose that prior beliefs are given by the random variable $d + l\varepsilon$, where ε is a mean-zero random variable. Then the revealed valuation weight θ 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 A5, 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 A1.

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 δ 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 A4, 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 A4. 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 Proofs of propositions in the body of the paper

B.1 Proof of Proposition 1

Proposition A1 establishes that the model has a revealed valuation weight representation. Proposition A3 establishes that $\lim_{\lambda\to 0}|1-\theta| = 0$ and $\lim_{\sigma\to\infty}|1-\theta| = 0$. 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, ρ_0 and ρ_1, such that $b(\rho_0) = 0$ and $b(\rho_1) = 1$. Now Proposition 2 of Caplin et al. (forthcoming) implies that the distribution π 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

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

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

$$\rho_0 = \frac{e^{\frac{u-\hat{t}}{\lambda}} - 1}{e^{\frac{u-\hat{t}}{\lambda}} - e^{\frac{u-t}{\lambda}}} \tag{19}$$

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 ρ_0. An alternative formulation is

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

Now clearly ρ_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, ρ_0 is increasing in u. Next, we see that ρ_0 is decreasing in $\hat{t}$ from the formulation in equation (19), since $e^{\frac{u-t}{\lambda}} < 1$ and $e^{\frac{u-\hat{t}}{\lambda}} > 1$ but is decreasing in $\hat{t}$.[23] Finally, ρ 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

[23]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 ρ_0 is monotone in $e^{\frac{u-\hat{t}}{\lambda}}$.

probability is increasing in u. Next, since ρ_0 is constant in r, the probability of buying is decreasing in r. And since ρ_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.2 Proof of Proposition 2

Proof. Proposition A4 establishes that the model has a revealed valuation weight representation. Proposition A5 establishes that the relative misreaction $|1-\theta|$ is increasing in λ and is decreasing in σ, with $\lim_{\lambda \to 0} |1-\theta| = 0$ and $\lim_{\sigma \to \infty} |1-\theta| = 0$.

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 A1, 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 $mt + (1-m)\bar{q}_o$, and thus increases the propensity to buy. □

B.3 Proof of Proposition 3

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

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

Again by the law of iterated expectations,

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

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 $E[X_1 X_2|Y] \geq E[X_1|Y]E[X_2|Y]$, 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 $(\bar{Y} - E[Y])(E[Y] - \underline{Y}) = Var[Y]$ for a distribution with binary support. $\square$

B.4 Proof of Proposition 4

We start with the more general statement.

Proposition B1. *Let Y be a random variable supported on $[\underline{Y}, \bar{Y}]$. Then*

$$Pr(Y > y) \geq \frac{E[(Y - \underline{Y})^2] - (y - \underline{Y})E[Y - \underline{Y}]}{(\bar{Y} - y)(\bar{Y} - \underline{Y})} \tag{21}$$

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

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 α on $Y = \bar{Y}$, weight β on $Y = y$, and weight $1 - \alpha - \beta$ on $Y = \underline{Y}$. Then

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

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 β on $Y=\underline{Y}$, weight α on $Y=y$, and weight $1-\alpha-\beta$ on $Y=\bar{Y}$,

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

from which the conclusion follows. $\square$

We obtain Proposition 4 as a corollary. When $\underline{Y}=0$ and $y=1$, equation (21) 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 (22) translates to

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

C 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 θ in our probit model corresponds to the average θ up to terms of order α^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, $f'(x) = -xf(x)$:

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

If we instead assume that the probability is given by $F\left(\frac{\mu_j - \log p}{\sigma_j} - \frac{E[\theta_i|\alpha]}{\sigma_j}\alpha\right)$, 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 μ_j/σ_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 $Var[\theta|\alpha]$ 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 proxy $X \in \{0, 1\}$. If the probit model produces estimates close to the average, then we should have $\hat{\theta}_{pop} = (1 - Pr(X = 1))\hat{\theta}_{pop}|_{X=0} + Pr(X = 1)\hat{\theta}_{pop}|_{X=1}$. 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,[24] the average of the θ 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 θ 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.

D Point estimates and confidence intervals for Figures 3 and 4

Table D.1: Average revealed valuation weights in Figure 3

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 D.1 presents the estimates for $E[\theta]$ and average tax owed displayed in figure 3. θ 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.

[24] 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 D.2: Average revealed valuation weights in figure 4

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, 0.36]
2	5.69	7.24%	0.41	0.28	[0.05, 0.50]
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 D.2 presents store-specific estimates of $E[\theta]$ by the average tax owed within each bin. θ 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 footnote 17. Standard errors are clustered at the subject level.

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

At the end of our experiment we asked participants (1) to report their sales tax rate in their current city of residence, and (2) to estimate the sales tax they would owe on an $8.00 (non-tax exempt) item purchased in their city of residence. We restricted the sample to the 60.0% of participants who (1) know their tax rate within 0.5 percentage points and (2) are able to calculate their tax burden within $0.05.

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 E.1 and E.2 present the results.

We next repeat our individual differences analysis, restricting to the "nearly-accurate beliefs and computation" sample. Tables E.1-E.3 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 (0.40 vs. 0.20; 95% CI for difference -0.06-0.44). The adjustments and their difference are similar in magnitude to our main sample results (0.39 vs. 0.16; 95% CI for difference -0.43 to -0.03).

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 over-reacting, 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 0.20-0.76) and in the triple tax regime (0.95 vs. 0.86; 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 (0.52, 95% CI 0.28-0.75) and the triple tax regime (0.10, 95% CI -0.01-0.20).

As with our main sample, the bounds on the variance of individual differences (0.84, 5% confidence bound of 0.24) imply at least some consumers overreact to taxes significantly. Consistent with the presence of overreaction in costly inattention models, we also estimate that some consumers reduce their valuation weights by at least 0.93 (5% confidence bound of 0.05) when shopping in the triple tax stores instead of the standard tax store. This estimate is very similar to the bound of 0.94 (5% confidence bound of 0.16) estimated from our main sample.

Figure E.1: Average revealed valuation weight for posted prices at or below a cutoff: nearly-accurate beliefs subsample

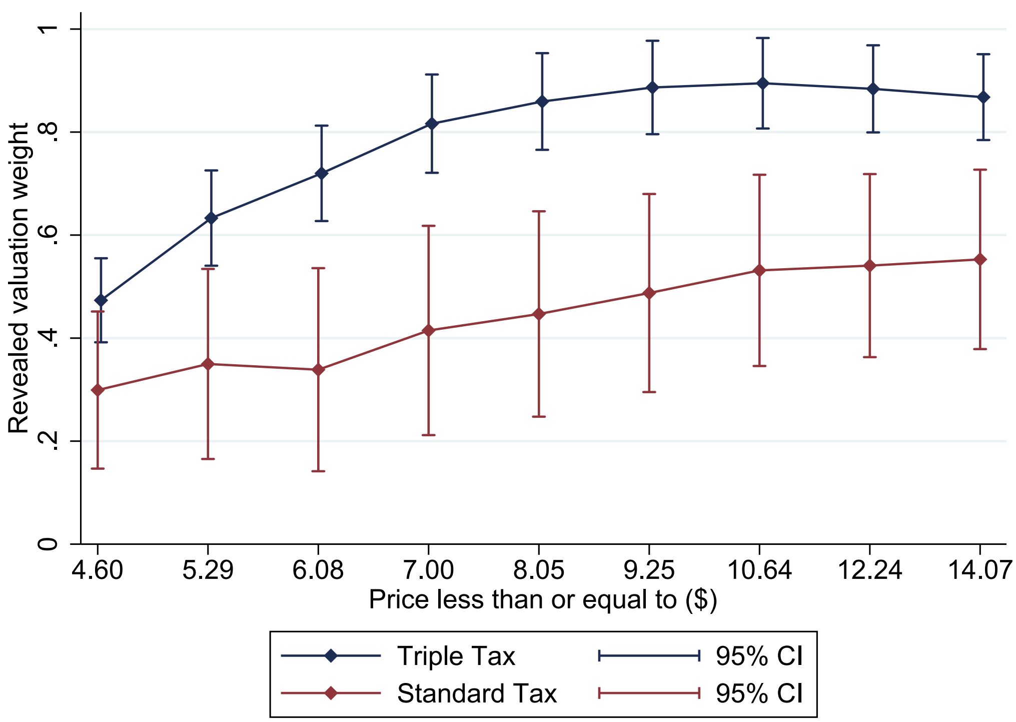


Figure E.1 recreates figure 3, restricting to the 70.3% of the main sample who could identify their local sales tax rate within 0.5 percentage points.

Figure E.2: Average revealed valuation weight for posted prices at or below a cutoff: restricting to participants with nearly-accurate beliefs and strong computational ability

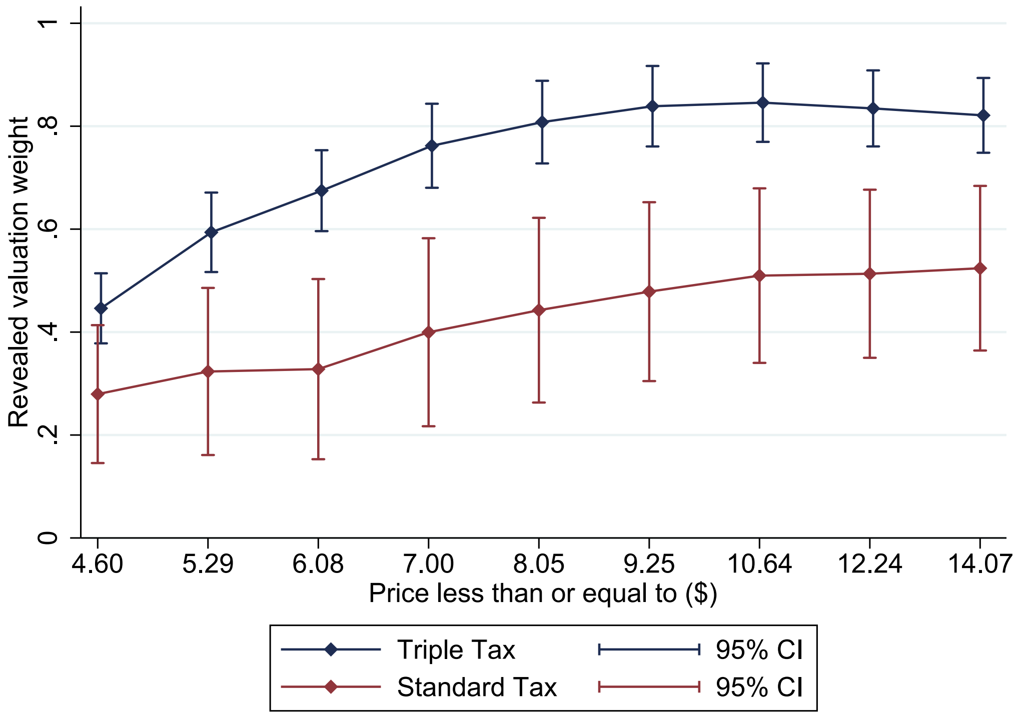


Figure E.2 recreates figure 3, restricting to the 62.9% of the main sample who could identify their tax burden within $0.05 on an $8.00 item purchased in their city of residence.

Table E.1: 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	0.20
	[0.85, 1.37]	[1.18, 1.45]	[-0.01, 0.42]
(2): Low valuation wgt.	0.34	0.73	0.40
	[0.13, 0.55]	[0.64, 0.83]	[0.24, 0.55]
(3): (1) − (2)	0.77	0.58	-0.19
	[0.47, 1.08]	[0.44, 0.73]	[-0.44, 0.06]

Table E.1 repeats table 1, restricting to the 60.0% 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 E.2: 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	0.95	0.04
	[0.64, 1.17]	[0.82, 1.07]	[-0.16, 0.24]
(2): High Adj.	0.43	0.86	0.43
	[0.23, 0.63]	[0.76, 0.96]	[0.28, 0.58]
(3): (1) − (2)	0.48	0.08	-0.39
	[0.20, 0.76]	[-0.04, 0.21]	[-0.62, -0.17]

Table E.2 repeats table 2, restricting to the 60.0% 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 E.3: 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	0.75	0.84
	[0.41]	[0.60]	[0.24]
Supremum (Lower Bound)	1.84	1.74	0.93
	[1.31]	[1.56]	[0.05]

Table E.3 repeats table 3, restricting to the 60.0% 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.

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

In our primary analyses we exclude 256 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 F.1 repeats figure 3 including these 292 participants.[25] 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 higher than the estimate in the main sample (0.79, 95% CI 0.72-0.86).

Tables F.1-F.3 replicate tables 1-3 including the respondents who failed the comprehension checks. We still exclude participants with monotonicity violations, as our estimation procedure in Section 5.1 assumes monotonic preferences in estimating a willingness-to-pay.

[25]We continue to exclude one of these participants who reported being under age 18.

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 to -0.08). The adjustments and their difference are similar in magnitude to our main sample results (0.39 vs. 0.16; 95% CI for difference -0.43 to -0.03).

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 (0.52, 95% CI 0.28-0.75) and the triple tax regime (0.10, 95% CI -0.01-0.20).

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 F.1: Average revealed valuation weight for posted prices at or below a cutoff: including participants who fail comprehension checks

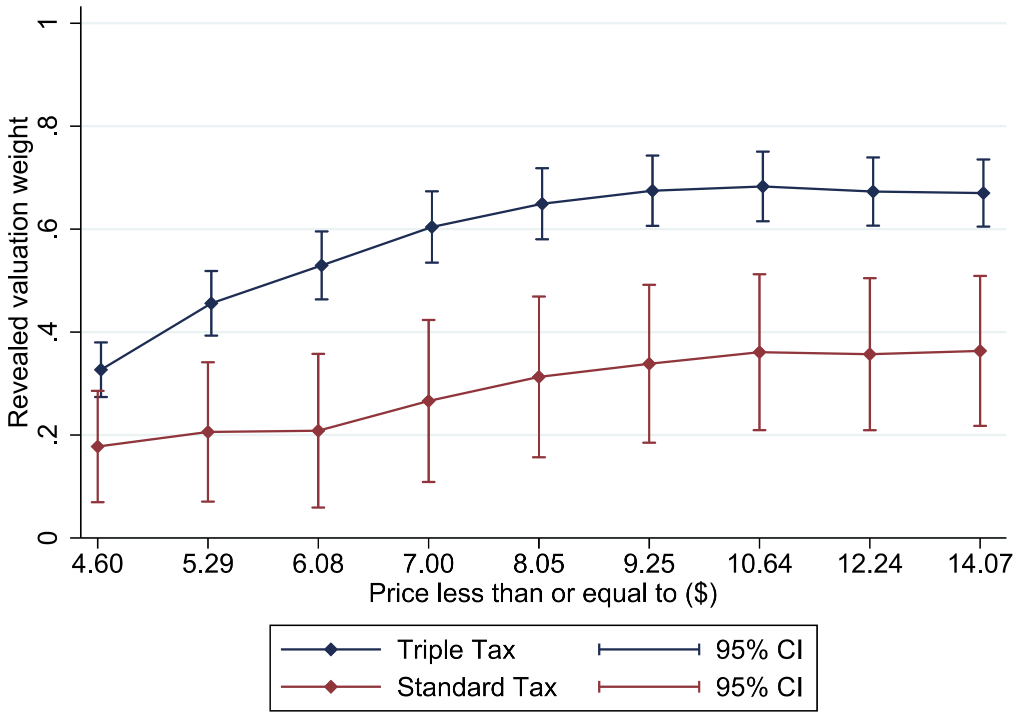


Figure F.1 recreates figure 3, including the 292 participants who failed comprehension checks or had monotonicity violations in purchase decisions.

Table F.1: Average revealed valuation weights by group: including participants who fail comprehension checks

	Standard	Triple	Triple − Standard
(1): High valuation wgt.	0.96	1.09	0.13
	[0.76, 1.16]	[0.99, 1.19]	[-0.02, 0.29]
(2): Low valuation wgt.	0.16	0.54	0.39
	[0.00, 0.32]	[0.48, 0.61]	[0.27, 0.50]
(3): (1) − (2)	0.80	0.55	-0.25
	[0.57, 1.03]	[0.44, 0.65]	[-0.42, -0.08]

Table F.1 repeats table 1 including 223 participants who were excluded from our main sample solely for failing our comprehension check.

Table F.2: Average revealed valuation weights by adjustment group: including participants who fail comprehension checks

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

Table F.2 repeats table 2 including 223 participants who were excluded from our main sample solely for failing our comprehension check.

Table F.3: Bounds on the dispersion of revealed valuation weights: including participants who fail comprehension checks

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

Table F.3 repeats table 3 including 223 participants who were excluded from our main sample solely for failing our comprehension check.

G Order Effects

One potential concern with our experiment design is that purchase decisions could be influenced by the order in which the nine purchase decisions are presented to consumers. In this appendix we test four potential order effects, and report the results in table G.1. First we examine whether the tax environment first shown to consumers impacts their buy probability. We test for this effect via the following model:

$$1 - Pr(buy_{ijk}|p) = \Phi\Bigg(\frac{\alpha \ln(p) + \beta^B \ln(1+\tau_i) \cdot I(\tau_{ik} = \tau_i) + \beta^C \ln(1+3\tau_i) \cdot I(\tau_{ik} = 3\tau_i) - \mu_j}{\sigma_j} + \frac{\gamma^B First_i^B + \gamma^C First_i^C}{\sigma_j}\Bigg) \tag{23}$$

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 (23), using $First_{ij}^k$ instead of $First_i^k$:

$$1 - Pr(buy_{ijk}|p) = \Phi\Bigg(\frac{\alpha \ln(p) + \beta^B \ln(1+\tau_i) \cdot I(\tau_{ik} = \tau_i) + \beta^C \ln(1+3\tau_i) \cdot I(\tau_{ik} = 3\tau_i) - \mu_j}{\sigma_j} + \frac{\gamma^B First_{ij}^B + \gamma^C First_{ij}^C}{\sigma_j}\Bigg)$$

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 (23), using $Last_{ij}^k$ instead of $First_i^k$:

$$1 - Pr(buy_{ijk}|p) = \Phi\Bigg(\frac{\alpha \ln(p) + \beta^B \ln(1+\tau_i) \cdot I(\tau_{ik} = \tau_i) + \beta^C \ln(1+3\tau_i) \cdot I(\tau_{ik} = 3\tau_i) - \mu_j}{\sigma_j} + \frac{\gamma^B Last_{ij}^B + \gamma^C Last_{ij}^C}{\sigma_j}\Bigg)$$

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

For our last 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 $\kappa_1, \kappa_2, \kappa_3 = \{A, B, C\}$:[26]

$$1 - Pr(buy_{ijk}|p) = \Phi\Bigg(\frac{\alpha \ln(p) + \beta^B \ln(1+\tau_i) \cdot I(\tau_{ik} = \tau_i) + \beta^C \ln(1+3\tau_i) \cdot I(\tau_{ik} = 3\tau_i) - \mu_j}{\sigma_j} + \frac{\sum_{\kappa_2 \neq \kappa_1; \kappa_3 \neq \kappa_2, \kappa_1} \gamma^{\kappa_1 \kappa_2 \kappa_3} I(First_{ij} = \kappa_1, Second_{ij} = \kappa_2, Third_{ij} = \kappa_3)}{\sigma_j}\Bigg)$$

We compute the Wald statistic for $\gamma^{ACB} = \gamma^{BCA} = ... = 0$, which has a corresponding p-value of 0.28.

Together, these tests provide evidence that the ordering in which questions are presented to consumers does not significantly impact purchase decisions.

[26] We omit the store ordering A, B, C due to collinearity.

Table G.1: 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 G.1 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).

H Covariates of attention

H.1 Local tax rate variation

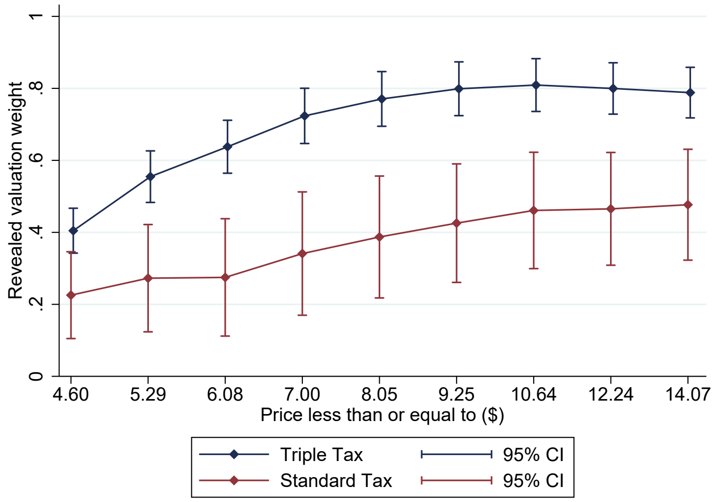
We first divide the sample into those whose local tax rate is above 7.00%, the median in our sample ("high tax group"), and those below 7.00% ("low tax group"). We then run the regression in equation (4) separately for the above-median and below-median tax groups to create figures analogous to figure 3.

Figure H.1 presents the results. Panel (a) uses the main sample and is identical to figure 3. 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 evidence that participants in high sales tax locations have lower revealed valuation weights than those from low sales tax locations.

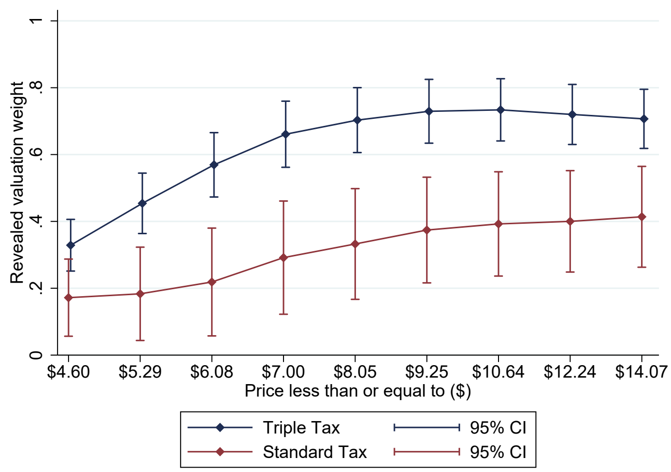
Table H.1 presents estimates of average θ by tax group using all prices. These estimates match the rightmost points of the series in figure H.1. The third column presents the difference, which is statistically significant for store C.

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

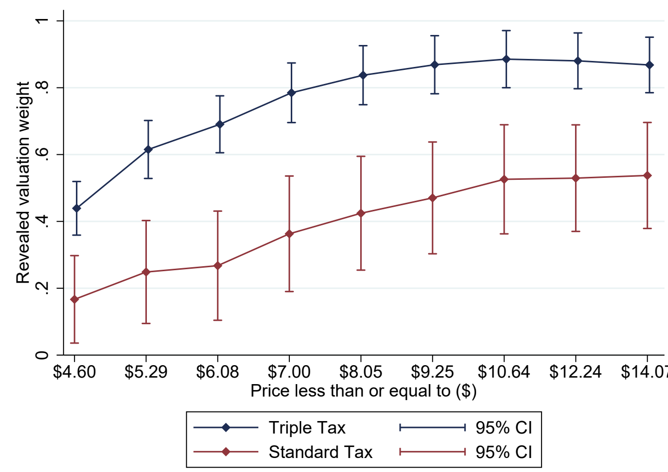
(a) Main sample



(b) Local sales tax rate above 7.00%



(c) Local sales tax rate at or below 7.00%



Panels (b) and (c) of figure H.1 recreate figure 3, restricting to participants above and below the median local sales tax rate respectively.

Table H.1: Average revealed valuation weights by tax group

	Standard	Triple
High tax group	0.41	0.70
	[0.26, 0.56]	[0.61, 0.79]
Low tax group	0.54	0.87
	[0.38, 0.70]	[0.79, 0.96]
Difference	-0.13	-0.17
	[-0.35, 0.09]	[-0.29, -0.05]

Table H.1 presents estimates of store-specific estimates $E[\theta]$ by tax group. θ 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.

H.2 Demographics

In this section we analyze how revealed valuation weights vary according to observed demographics. We separately analyze the effects of political party, education, income, and age.

Political party: Table H.2 presents average θ estimates for self-identified Republicans (28.5% of our sample), Democrats (32.1% of our sample), and individuals with independent or other political beliefs (39.6% of our sample).[27] Republicans and Democrats have an average θ_B of 0.52 and 0.51 respectively (95% CI for difference: -0.39-0.39). Republicans have a slightly larger θ_C than do Democrats in our sample (0.86 vs. 0.74), but the difference is not statistically significant (95% CI for difference: -0.06-0.30).

Education: Table H.2 compares the average θ 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 θ_B (0.51 vs. 0.46), but the difference is not statistically significant (95% CI for difference -0.27-0.38). Both education groups have the same estimate for θ_C (0.79, 95% CI for difference -0.14-0.15).

Income: Table H.4 presents average θ estimates for each income quartile. Individuals in the top income quartile have self-reported annual income above \$80,000, in the second quartile from \$49,000-\$80,000, in the third quartile from \$25,000-\$49,000, and in the bottom quartile below \$25,000.

All quartiles have average θ_C point estimates in the 0.77-0.84 range. Individuals in the top 3 quartiles have average θ_B estimates in the 0.50-0.52 range, while individuals in the bottom quartile have average an average θ_B of 0.39 (95% CI: 0.07-0.71). A test of equivalence between the θ 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.

[27] 30.6% of participants self-identify as independent and 8.9% of participants self-identify as other.

Table H.2: Average revealed valuation weights by political party

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

Table H.2 presents store-specific estimates of $E[\theta]$ by political party affiliation. θ 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 H.3: Average revealed valuation weights by education: college graduates versus not college graduates

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

Table H.3 presents store-specific estimates of $E[\theta]$ by education level. θ 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 H.4: Average revealed valuation weights by income quartile

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

Table H.4 presents store-specific estimates of $E[\theta]$ by income quartile. θ 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.

I Alternative construction of proxies for valuation weights

Figure I.1: Average revealed valuation weight by posted price

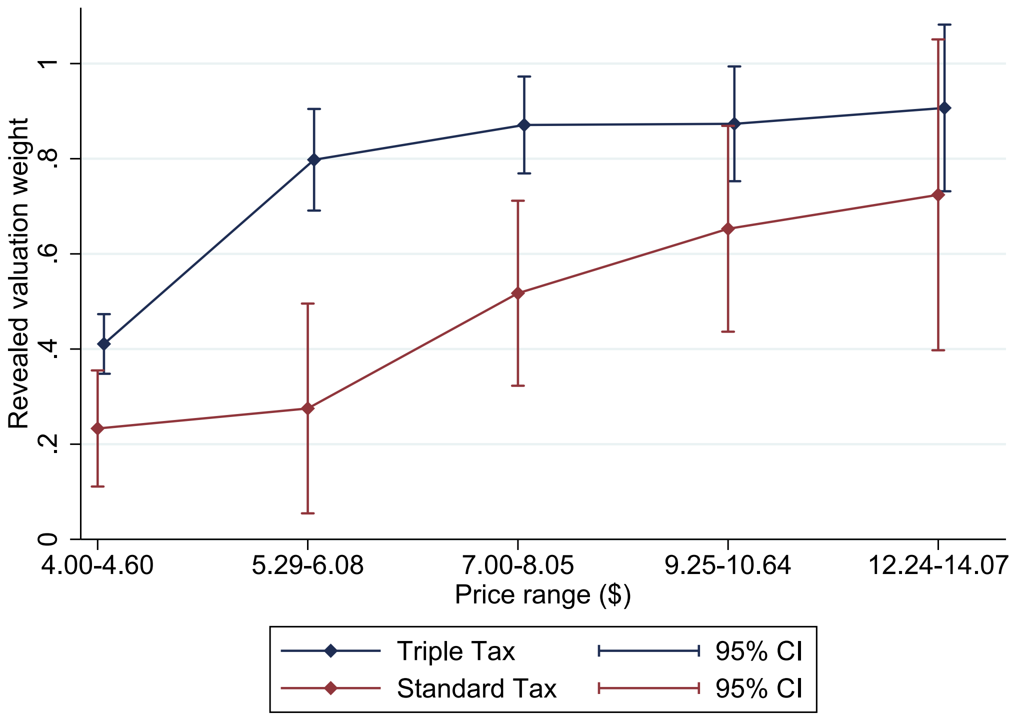


Figure I.1 presents store-specific estimates $E[\theta]$ by the posted price. θ is defined as the revealed valuation weight that consumers place on the sales tax, with $\theta = 0$ corresponding to complete neglect of the tax and $\theta = 1$ corresponding to the equal weight of the tax and salient price.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

Table I.1: Average revealed valuation weights by valuation weight group: instrumenting with an 80th percentile cutoff

	Standard	Triple	Triple − Standard
(1): High valuation wgt	1.24	1.32	0.08
	[0.99, 1.48]	[1.19, 1.45]	[-0.11, 0.27]
(2): Low valuation wgt	0.29	0.67	0.38
	[0.12, 0.45]	[0.60, 0.74]	[0.26, 0.50]
(3): (1) − (2)	0.95	0.65	-0.30
	[0.67, 1.23]	[0.52, 0.79]	[-0.51, -0.08]

Table I.1 repeats table 1 with an alternative instrument for 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 I.2: Average revealed valuation weights by valuation weight group: instrumenting with an 85th percentile cutoff

	Standard	Triple	Triple − Standard
(1): High valuation wgt	1.23	1.31	0.08
	[0.90, 1.55]	[1.17, 1.45]	[-0.17, 0.34]
(2): Low valuation wgt	0.35	0.70	0.35
	[0.22, 0.50]	[0.63, 0.77]	[0.25, 0.46]
(3): (1) − (2)	0.88	0.61	-0.27
	[0.55, 1.22]	[0.47, 0.75]	[-0.53, -0.00]

Table I.2 repeats table 1 with an alternative instrument for 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 I.3: Average revealed valuation weights by adjustment group: instrumenting with an 20th percentile cutoff

	Standard	Triple	Triple − Standard
(1): Low Adj.	0.85	0.85	0.01
	[0.62, 1.08]	[0.74, 0.97]	[-0.16, 0.17]
(2): High Adj.	0.38	0.77	0.39
	[0.23, 0.53]	[0.71, 0.84]	[0.27, 0.52]
(3): (1) − (2)	0.47	0.08	-0.39
	[0.24, 0.69]	[-0.03, 0.19]	[-0.57, -0.21]

Table I.3 repeats table 2 with an alternative instrument for high adjustment groups 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 I.4: Average revealed valuation weights by adjustment group: instrumenting with an 15th percentile cutoff

	Standard	Triple	Triple − Standard
(1): Low Adj.	0.93	0.87	-0.06
	[0.66, 1.21]	[0.74, 1.01]	[-0.27, 0.14]
(2): High Adj.	0.39	0.77	0.38
	[0.24, 0.55]	[0.70, 0.85]	[0.26, 0.50]
(3): (1) − (2)	0.55	0.10	-0.45
	[0.25, 0.84]	[-0.05, 0.24]	[-0.66, -0.23]

Table I.4 repeats table 2 with an alternative instrument for high adjustment groups 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 $F(\hat{\theta}_{i1C} - \hat{\theta}_{i1B}) \leq 0.15$.

J Additional details of the experiment

Additional Screenshots

Figure J.1: 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.

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

Continue

Figure J.2: Instructions (top of screen)

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.

Note: The rest of the instructions screen displayed the multiple price list seen in figure 1. Subjects did not shop for the Oversize Golf Umbrella in the experiment.

Figure J.3: Pre-purchase Comprehension Questions

(a) Question 1

Question 1. How big of a shopping budget do you have for each purchase decision?

- $10
- $16
- $30

(b) Question 2

Question 2. If you are selected to receive the shopping budget, how many of your purchase decisions will the computer randomly choose to implement?

- 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

(c) Question 3

Question 3. At what prices do you see the products in this study?

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

Note: Subjects answered these questions before making purchase decisions, and were given unlimited tries to correctly answer all three questions. The correct answers are $16, one purchase decision, and "the prices vary" respectively.

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.

Product	Amazon.com price	Amazon.com product description
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.
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 repellant vinyl that easily wipes clean. With metal grommets along top of the liner to prevent tearing. Here in Monaco Blue, this liner is available in a variety of fashionable colors. With its wonderful features and fashionable colors, this liner could also make a great shower curtain.

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