Title
Essays on Field Experiments in Behavioral Economics

Author
Morrison, William

Publication Date
2023

Peer reviewed|Thesis/dissertation

Essays on Field Experiments in Behavioral Economics

By

William Morrison

A dissertation submitted in partial satisfaction of

the requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Comittee in charge:

Professor Dmitry Taubinsky, Chair

Professor Stefano DellaVigna

Professor Ben Handel

Spring 2023

Abstract

Essays on Field Experiments in Behavioral Economics

By

William Morrison

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Dmitry Taubinsky, Chair

This dissertation contributes to the literature of using field experiments to test and develop theories of consumer behavior. In Chapter One, my coauthors and I partner with the YMCA to analyze how public recognition for exercise affects consumer's utility. We find that while it motivates positive behavior change, it creates highly unequal payments, with low performers losing a lot of utility from having their exercise habits publicly shared. In Chapter Two, my coauthor and I use an online shopping experiment to study how consumers (mis)react to sales taxes. We find evidence consistent with the theory that consumers using heterogeneous rules of thumb to compute the opaque tax when the stakes are low, but using costly mental effort at higher stakes. The results allow us to differentiate between various economic theories of limited attention. In Chapter Three, my coauthor and I partner with an online apparel retailer to study the consequences of offering a one-time price discount to consumers, with a particular focus on consumer beliefs. We find that the net effect was no significant difference in revenue, order frequency or profit from the two groups in our experiment. We further find that price discounts do not change the perceived value of the brand or quality of the product, which contradicts many existing economic theories.

Contents

Introduction:

Field experiments are a crucial tool in the field of behavioral economics as they allow researchers to observe and test theories of human behavior in real-world settings rather than in a controlled laboratory environment. This dissertation contributes to the literature of using field experiments to test and develop theories of consumer behavior. In Chapter One, I develop a portable empirical methodology for measuring and monetizing social image utility, and deploy it in experiments on exercise and charitable behavior. Specifically, my coauthors and I partner with the YMCA to study the impact of having public recognition for exercise (or the lack of exercise) over the month and how that affects both exercise patterns.We find that public recognition motivates desirable behavior but creates highly unequal image payoffs. High-performing individuals enjoy significant utility gains, while low-performing individuals incur significant utility losses. We estimate structural models of social signaling, and use the models to explore the social efficiency of public recognition policies.

In Chapter Two, my coauthor and I create an online shopping experiment to study how consumers (mis)react to sales taxes. We have consumers shop for the same product in three different stores, where we exogenously vary the sales tax rate between stores. This setup allows us to test costly attention models of consumers' misreaction to opaque taxes. The field experiment involves shrouded sales taxes that are exogenously varied within consumer over time. Some consumers systematically underreact to sales taxes while others systematically overreact, but higher stakes decrease both under- and overreaction. This is consistent with consumers using heterogeneous rules of thumb to compute the opaque tax when the stakes are low, but using costly mental effort at higher stakes. The results allow us to differentiate between various theories of limited attention.

In Chapter Three, my coauthor and partner with an online retailer, with a market capitalization of around $50 million, to study the consequences of offering a one-time price discount to consumers, with a particular focus on consumer beliefs. We find that discount-eligible customers made more purchases and accounted for 30 percent more revenue during the two week period of coupon eligibility. However, the control group made substantially more purchases at full price in the period following discount eligibility. The net effect was no significant difference in revenue, order frequency or profit from the two groups in our experiment. Through an incentivized survey of customers, we find that discounts do not change the perceived value of the brand or quality of the product. Customers who received discounts report a higher likelihood of seeing discounts in the future. This contradicts much of the existing theoretical literature. Using this we propose a simple model that describes rational customers trading off whether to wait for discounts: paying a lower price at the cost of a time delay in receiving a product. Customer decisions depend on their Bayesian belief about the likelihood of a future discount. Understanding this, firms set the probability of a discount to maximize long-term profitability.

In short, all three chapters of this dissertation feature field experiments to enable analyzing human behavior in real-world settings. Two feature corporate partnerships, one with the YMCA and one with an anonymous apparel retailer with a market capitalization of $50 million, and in the third we develop our own online shopping platform. These allow us to test and develop theories related to the welfare impact of public recognition, consumers' (mis)reaction to shrouded prices, and how price discounts affect consumer beliefs.

Acknowledgments:

I would like to express my sincere gratitude to several individuals who have helped me completion this degree. First and foremost, I would like to thank my advisor, Dmitry Taubinsky, for his invaluable assistance throughout my academic journey. His insightful mentorship has helped shape the direction of this research and my career path, and have been crucial for my success.

I am grateful to my committee members, Stefano DellaVigna and Ben Handel, for their insightful comments, constructive feedback, and valuable suggestions, which have helped to strengthen this dissertation.

I would like to extend a special thanks to my classmates, Jakob Brounstein, Sree Kancherla, and John Wieselthier for their support and constant friendship during my time in graduate school.

To my partner, Nathan Kyaw, I am incredibly grateful for your unwavering support throughout my graduate studies. Your presence in my life has made this journey much more fulfilling.

Finally, I would like to thank my parents and siblings, Dan and Tim, for their unconditional love and encouragement. Their belief in me has been a constant source of motivation and inspiration throughout my time in graduate school.

For Lexi and Ori

Chapter 1

Measuring the Welfare Effects of Shame and Pride[1]

1.1 Introduction

The human desire to avoid negative social image and appear "good" is a powerful motivator (Loewenstein et al., 2014; Bursztyn and Jensen, 2017). For instance, 89% of businesses use some form of public recognition programs (WorldatWork, 2017), including examples like "employee of the month" (Kosfeld and Neckermann, 2011). Bloom and Van Reenen (2007) find that 60% of manufacturing companies publicly reveal and compare employees' performance data. Governments use public recognition programs to motivate citizens to pay their taxes (Bø et al., 2015; Perez-Truglia and Troiano, 2018), to motivate bureaucrats to do a better job (Gauri et al., 2018), and to encourage teachers, doctors, and managers in schools and hospitals to improve their performance.

Recent field studies confirm that public recognition of individuals' behavior has substantial effects in a number of economically important domains. Examples include charitable and political donations (Soetevent, 2005, 2011; Perez-Truglia and Cruces, 2017), tax compliance (Perez-Truglia and Troiano, 2018), education and career choices (Bursztyn and Jensen, 2015; Bursztyn et al., 2017b, 2019), employee productivity (Ashraf et al., 2014; Neckermann et al., 2014; Bradler et al., 2016; Kosfeld et al., 2017; Neckermann and Yang, 2017), voter turnout (Gerber et al., 2008), blood donation (Lacetera and Macis, 2010), childhood immunization (Karing, 2019), energy conservation (Yoeli et al., 2013), and credit card take-up (Bursztyn et al., 2017a).[2]

[1] Coauthored with Luigi Butera, Copenhagen Business School, Robert Metcalfe, USC, and Dmitry Taubinsky, UC Berkeley. This chapter includes previously published material from "Measuring the Welfare Effects of Shame and Pride." *American Economic Review*, 112 (1): 122-68. The YMCA experiment was approved by University of Chicago IRB, #IRB15-1647; the charitable contribution experiments were approved by Boston University IRB, #5473X (BU) and University of California Berkeley IRB, #2020-01-1288. AEA RCT ID: AEARCTR-0004004 and AEARCTR-0005737.

[2] Laboratory experiments also show that public recognition can enhance prosocial behavior. E.g., Andreoni and Petrie (2004), Rege and Telle (2004), Andreoni and Bernheim (2009), Ariely et al. (2009a), Jones and Linardi (2014), Bernheim and Exley (2015), Exley (2018), and Birke (2020).

The *financial* costs of utilizing public recognition to motivate behavior are typically low, but the *image* costs—such as the emotional costs of shame—may not be. Although behavioral scientists sometimes refer to social-influence-based interventions as light-touch, innocuous "nudges" (Halpern, 2015; Benartzi et al., 2017), it is well-understood that such a label would not be appropriate for a policy that leads to a significant number of individuals experiencing shame (see, e.g., Bernheim and Taubinsky, 2018a, for a review). Indeed, there is a vigorous debate about the appropriateness of public policies that generate feelings of shame, with some political and legal theorists arguing that such policies are an unjustifiable offense to human dignity and a form of mob-justice (Massaro, 1991; Nussbaum, 2009; see also Bénabou and Tirole, 2011 for formal analysis).[3] On the other hand, public recognition policies that mostly generate warm feelings of pride are arguably a "win-win." Developing quantitative methods for measuring the welfare effects of public recognition is therefore crucial for both positive and normative progress.

In this paper, we develop a portable approach for directly quantifying the image utility effects of public recognition. We deploy our approach in two different experimental designs conducted with four different subject pools. In each experiment, we address three research questions. First, do people have a significant willingness to pay to seek out or avoid public recognition of their behavior, implying that public recognition has a direct image utility effect? Second, how does utility from public recognition depend on people's realized behavior? In particular, are individuals choosing high levels of socially desirable behavior made better off (e.g., from experiencing pride), and are individuals choosing low levels of the desirable behavior made worse off (e.g., from experiencing shame)? Third, are the net image payoffs negative or positive? As we show, this third question relates to both the curvature of the public recognition utility function (PRU), and to the reference standard at which image payoffs transition from negative to positive.

Our first experiment was conducted in the field, in partnership with the YMCA of the USA and the YMCA of the Triangle Area (YOTA) in Raleigh, North Carolina.[4] We invited all members of YOTA to participate in a newly designed one-month program called "Grow & Thrive." This program encouraged members to attend their local YMCA more often by having an anonymous donor give $2 to the local YMCA for each day that an individual attended the YMCA. While this charity incentive was provided to everyone, participants could also be assigned to a public recognition program, which would reveal each participant's attendance and donation raised to all other participants in the program.

Our second set of experiments was conducted online and builds on the Ariely et al. (2009a) and DellaVigna and Pope (2018) "Click for Charity" task. The online experiments complement our field experiment by utilizing a design that gives us greater flexibility and control over the decision environment. In this real-effort task, participants raise money for the American Red Cross by repeatedly pressing two keys on a computer keyboard. The design was within subjects, and participants took part in three rounds. In the Anonymous

[3] Others promote such policies as instruments for the internalization of community norms (Etzioni, 1999; Kahan and Posner, 1999).

[4] The YMCA of the USA is a national, non-profit, charitable organization that supports local communities with a focus on youth development, healthy living, and social responsibility. The YMCA of the Triangle Area primarily serves the Raleigh-Durham, North Carolina, and surrounding communities. It is one of 850 member association YMCAs.

Effort Round, participants' scores were not shared with anyone. In the Anonymous and Paid Effort Round, participants additionally received pay for their effort. In the Publicly-Shared Effort Round, participants' contributions to the Red Cross were publicly shared with others in the experiment through a webpage that posted individuals' photos, amount raised, rank relative to other participants, and, for two of the subject pools, names.[5]

We administered the online protocol simultaneously to three different subject pools that differ in individuals' familiarity with each other: (i) the online panel Prolific Academic, where participants almost surely do not know each other (henceforth *Prolific sample*); (ii) UC Berkeley's pool of subjects for economics and psychology experiments, where some participants might know each other (henceforth *Berkeley sample*); and (iii) a section of Boston University's statistics class for second- and third-year undergraduate business majors, where students are likely to know each other (henceforth *BU sample*).

Our revealed-preferences approach to estimating the effects of shame and pride utilizes the incentive-compatible Becker-DeGroot-Marschak (BDM) mechanism to elicit participants' (possibly negative) willingness to pay (WTP) for public recognition at various possible realizations of their performance. An advantage of this "strategy method" approach is that it is robust to possible mis-forecasting of one's future behavior. In the YMCA experiment, participants' WTP to be publicly recognized was elicited in an initial online survey before the start of the month-long period during which incentives for attendance were provided. Participants were asked to state their WTP to be publicly recognized for all levels of attendance ranging from 0 to 30 days. To generate random assignment, as well as to minimize any negative inferences that could be drawn about participants who are not publicly recognized, the BDM responses were used to determine assignment to public recognition with only 10 percent chance. With 90 percent chance assignment was random.

In the charitable contribution experiments, we again used the BDM mechanism to elicit participants' WTP to have their contribution to the Red Cross publicly recognized, for different possible levels of performance. As before, participants' elicited preferences were implemented with 10 percent chance, while 90 percent chance participants were randomly assigned to have their outcome based on one of the three rounds. In the 10 percent of cases where participants' preferences were implemented, participants' contribution was based on a randomly chosen score from one of the three rounds, and participants with a preference to be recognized were listed alongside the participants randomly assigned to the Publicly-Shared Effort Round.

We present six sets of results. First, we find that public recognition substantially increased desirable behavior. In the YMCA experiment, it significantly increased attendance by 17 percent, and in the charitable contribution experiments, it significantly increased contributions by 13 percent, 14 percent, and 13 percent in the Prolific, Berkeley, and BU samples, respectively.

Second, we find that a majority of participants have a non-zero WTP for public recognition. The fraction of participants with positive WTP to either opt in or opt out of public recognition at some level of performance is 93 percent, 73 percent, 78 percent, and 89 percent in the YMCA, Prolific, Berkeley, and BU samples, respectively. Participants' eagerness to

[5] Birke (2020) utilizes a similar approach to public recognition of online participants. We thank him for his advice and for kindly sharing his code.

pay for social image is consistent with a long intellectual tradition of incorporating "psychic" or emotional effects into otherwise standard economic models using money metrics (starting with, e.g., Becker, 1968; Ehrlich, 1973).

Third, the WTP data allows us to examine how participants' image payoffs vary with performance. In all experiments, image payoffs are strictly increasing in performance, participants in the bottom quartile of performance receive negative payoffs, while participants in the top quartile of performance receive positive payoffs, on average. The robust presence of negative payoffs from public recognition is consistent with leading economics models of social signaling (e.g., Bénabou and Tirole, 2006; Andreoni and Bernheim, 2009), but it is not a robust implication of psychological theories of shame (Tangney et al., 1996, 2007). Psychologically, raising *any* amount of money for the Red Cross could have been perceived as commendable prosocial behavior.[6]

Fourth, we estimate structural models of social signaling. We consider "action-signaling" models in which individuals directly care about how their action compares to the population behavior (e.g., Becker, 1991; Besley and Coate, 1992; Blomquist, 1993; Lindbeck et al., 1999), and "characteristics-signaling" models in which individuals care about what their action reveals about their characteristics (e.g., Bénabou and Tirole, 2006; Andreoni and Bernheim, 2009; Ali and Bénabou, 2020). We provide a key out-of-sample test of the validity of our methodology and modeling framework by showing that data on (i) the treatment effect of public recognition and (ii) people's WTP for public recognition can be used to predict (iii) the effect of financial incentives on behavior. In the charitable contribution experiments the financial incentive was randomized, and thus we estimate its effects directly. In the YMCA experiment we compare our models' predictions to individuals' forecasts of how they would respond to a financial incentive. Across all four subject pools, we find that the models' predictions only slightly overestimate the effects of the financial incentives, and that the difference is not statistically significant at conventional levels. This suggests that our monetization of image payoffs is accurately capturing the (presumably nuanced) psychological effects of public recognition.

Fifth, we study the shape of the PRU. In our models, whether the net image payoffs are negative or positive depends on the degree of concavity and the reference standard for positive image. Intuitively, more concavity leads individuals to be more sensitive to negative image, while a higher standard increases the fraction of individuals who experience negative effects. For example, if people derive positive image if and only if they are "better than average," then, by Jensen's Inequality, a concave PRU makes public recognition negative-sum while a convex PRU would make public recognition positive-sum.

Both the reduced-form analyses and the structural estimates imply significant concavity in the YMCA and Prolific samples. We cannot reject linearity in the Berkeley and BU samples, although we also cannot reject that those samples feature as much concavity as the YMCA and Prolific samples. We also find that the standard for positive image payoffs is higher than the population average behavior in the YMCA and BU samples, is equal to the average in the Berkeley sample, and is lower than the average in the Prolific sample. Collectively, these results imply that public recognition is negative-sum in the YMCA and

[6] From a psychological perspective, shame is an emotion that accompanies moral transgressions (Tangney et al., 1996, 2007), and ex-ante it was unclear that any action in our experiments could be labeled as such.

BU samples, is approximately zero-sum in the Berkeley sample, and is positive-sum in the Prolific sample.

Sixth, we use our structural estimates to generate out-of-sample predictions about the welfare and behavior effects of scaling up the public recognition intervention in the YMCA experiment to all of YOTA. We find that at the parameters estimated for the YMCA sample, public recognition is substantially negative sum. However, if the PRU more closely resembled our estimate in the Prolific sample, then public recognition would be positive-sum.

Collectively, our results illustrate the importance of directly measuring the welfare effects of public recognition, and the potential benefits of our methodology. Our findings about the prevalence of negative image utility imply that the appropriateness of public recognition in settings such as ours could be legitimately debated. From a pure economic efficiency perspective, we find that public recognition could be a socially inefficient tool for behavior change in the YMCA field setting despite the low financial cost of the intervention and initial enthusiasm of our field partners. On the other hand, our results from the Prolific sample also illustrate that public recognition could be an efficient tool in other settings. This illustrates that it is inappropriate to judge the success of a public recognition policy solely by its effect on behavior, and how our methodology could help enrich the applied work on social signaling by helping researchers study both behavior and *welfare.*

The remainder of the paper is organized as follows. Section 1.2 further reviews the related literature. Section 2.2 introduces our theoretical framework. Section 1.4 describes the YMCA experiment and Section 1.5 reports the reduced-form results. Section 1.6 describes the charitable contribution experiments and Section 1.7 reports the reduced-form results. Section 1.8 presents our estimates of structural models of social signaling and welfare implications. Section 2.7 concludes.

1.2 Discussion of Related Literature

Our research is related to several literatures. The most closely related is the large and growing experimental literature studying the effects of public recognition on individual behavior, summarized above. However, this literature studies behavior, and does not assess welfare effects of positive or negative image. We build on this literature by developing a portable approach for measuring image utility, which can be productively incorporated into future experiments on public recognition.

Our work also relates to a recent literature that evaluates the welfare effects of scalable, non-financial policy instruments such as reminders (Damgaard and Gravert, 2018), energy-use social comparisons (Allcott and Kessler, 2019), calorie labeling (Thunstrom, 2019), and defaults (Carroll et al., 2009; Bernheim et al., 2015).[7] Our paper contributes to this literature by analyzing a different and highly popular non-financial policy instrument, and by providing new methods for testing and estimating models of social signaling. Unlike this prior work,

[7] Additionally, our work relates to the theoretical work of Kaplow and Shavell (2007), who derive conditions for when and how much to use policies that invoke shame or pride when the objective is to maximize social welfare.

our experiments utilize a new "strategy method" design technique that eliminates the need to rely on the assumption that individuals can correctly forecast their future behavior.[8]

Finally, our model-based design allows us to produce the first structural estimates of leading models of social signaling such as those of Bénabou and Tirole (2006).[9] We therefore also contribute to a recent and growing literature in structural behavioral economics (see DellaVigna, 2018 for a review). The work by DellaVigna et al. (2012) and DellaVigna et al. (2017) is closest in spirit to our paper in this literature, although they do not study the scalable lever of revealing people's behavior to others, nor do they estimate the leading social signaling models. These two papers quantify the social pressure effects of face-to-face interaction in charitable contributions and voting, respectively. They do this by using structural methods to infer the cost of social pressure from the degree to which individuals avoid interaction with others. In contrast, we use conceptually different, and more direct experimental techniques that leverage the richness of our action space and allow us to directly observe the shape of utility from the social motives. The richer data provided by our approach enables the estimation of structural models of social signaling.

1.3 Theoretical Framework for Analysis

1.3.1 The Models

We consider individuals who choose the level of intensity $a \in \mathcal{A} \subset \mathbb{R}^+$ to engage in some activity. Choosing a generates *material utility* $u(a;\theta) + y$, where y is the individual's income and $\theta \in \mathbb{R}$ is the type of the individual, which we typically interpret as the individual's intrinsic motivation to engage in socially desirable behavior.[10] We assume that $u(a;\theta)$ is single-peaked in a and that $\frac{d}{da}u(a;\theta)$ is increasing in θ and is bounded. Thus, each individual has some optimal intensity level $a^*(\theta)$, and higher types θ derive more benefit from choosing higher levels of a. In addition to material utility, individuals also derive public recognition utility S, which we define below.

Consistent with psychological theories, we recognize that people can derive image payoffs either directly from their behavior a or from their characteristics θ (see, e.g., Leary, 2007). We thus consider models of both of these mechanisms.

To simplify exposition, in the body of the paper we consider fully-revealing equilibria in which each individual's choice of action a is perfectly observed, and in which there is a one-to-one mapping between types θ and actions a. We present the models and solution concepts in full generality in Appendix A.

Formally, let S be an increasing function that satisfies $S(0) = 0$, and let $\nu \in \mathbb{R}^+$ be the "visibility parameter" (Ali and Bénabou, 2020), which might depend on the number of observers, or the extent to which the observers are paying attention to an individual's

[8] See Bernheim and Taubinsky (2018a) for a more detailed discussion of the the literature and potential confounds.

[9] Ariely et al. (2009a), Exley (2018), Bursztyn et al. (2019), and Karing (2019) test comparative statics of the Bénabou and Tirole (2006) model, and Karing (2019) quantifies the value of sending a positive (but not fully-revealing) signal. These papers do not estimate the underlying public recognition utility function.

[10] Assuming that utility is linear in income is a simplifying assumption that is not crucial for our theoretical exposition, but that is realistic given the relatively small financial stakes of our experimental setting.

behavior. The *action-signaling model* posits that when an individual's action is made public, the individual cares about how his action compares to a weighted average of behavior in the population (Becker, 1991; Besley and Coate, 1992; Blomquist, 1993; Lindbeck et al., 1999, 2003):

$$u(a;\theta) + y + \nu S\left(a - \rho\bar{a}\right) \tag{1.1}$$

where $\bar{a}$ is the average action in the population, and $\rho\bar{a}$ is the standard for what constitutes a positive versus negative image. The *characteristics-signaling model* posits that individuals derive utility from what their action reveals about their characteristics to the audience (e.g., Andreoni and Bernheim, 2009; Bénabou and Tirole, 2006; Ali and Bénabou, 2020):

$$u(a;\theta) + y + \nu S(\mathbb{E}[\theta|a] - \rho\bar{\theta}) \tag{1.2}$$

where $\mathbb{E}[\theta|a]$ is the inference about a person's type given their behavior, $\bar{\theta}$ is the average type in the population, and $\rho\bar{\theta}$ is the standard for what constitutes positive versus negative image.[11]

The parameter ρ determines how many individuals experience positive versus negative image. When $\rho = 0$, all individuals choosing $a > 0$ receive positive image payoffs from public recognition. When $\rho > 1$, the standard is particularly demanding, as individuals must perform better than average to receive positive image payoffs.

As the general model in Appendix A clarifies, the parameter ρ is a reduced-form parameter that is endogenous to the information structure. In our empirical estimates, the parameter should be regarded as a rough, not definitive, measure of whether individuals generally have high or low standard for positive image payoffs. In particular, in the case where (almost) nothing is revealed about individuals' behavior and characteristics, the general model makes the sensible prediction that individuals incur no image payoffs. Roughly speaking, the parameter ρ tends to 1 as the information structure coarsens. Additional parametric assumptions are necessary to use our estimates of ρ to make out-of-sample predictions about the impacts of other types of public recognition schemes.

1.3.2 Net Image Payoffs

Although theoretical work often makes the simplifying assumption that the net image payoff is zero by assuming that S is linear and that $\rho = 1$, it is well understood that both assumptions are not without loss of generality (e.g., Bénabou and Tirole, 2006, 2011). From a psychological perspective, because shame and pride are separate emotions of different valences (Tangney et al., 2007), people's well-being may not be equally sensitive to these two emotions, implying nonlinearity in S. And to the extent that shame is an emotion that accompanies moral transgressions (Tangney et al., 1996, 2007), it is also not clear that ρ might even be strictly positive for all behaviors. For example, raising *any* amount of money for charity might always lead to pride.

Both the curvature of S and the value of ρ determine the net image payoff. In particular, let $a^*(\theta)$ denote individuals' equilibrium strategies. Then the image payoffs in the two models

[11] Note that there always exists a separating equilibrium in the characteristics-signaling model when u is smooth and $\mathcal{A}$ is convex and compact (Mailath, 1987).

are, respectively, given by:

$$\mathbb{E}[S(a^*(\theta) - \rho\bar{a})] \tag{1.3}$$
$$\mathbb{E}[S(\mathbb{E}[\theta|a^*(\theta)] - \rho\bar{\theta})] \tag{1.4}$$

If S is concave and $\rho \geq 1$, then Jensen's Inequality implies that the net image payoffs in the two models are given by:

$$\mathbb{E}[S(a^*(\theta) - \rho\bar{a})] \leq S(\mathbb{E}[a^*(\theta) - \rho\bar{a}]) \leq 0$$
$$\mathbb{E}[S(\mathbb{E}[\theta|a^*(\theta)] - \rho\bar{\theta})] \leq S\left(\mathbb{E}[\mathbb{E}[\theta|a^*(\theta)] - \rho\bar{\theta}]\right) \leq 0$$

Thus, net image payoffs are negative when the function is concave and the standard for behavior/characteristics is at least as demanding as the average. Conversely, net image payoffs are positive when $\rho \leq 1$ and S is convex.[12] In general, the net image payoff decreases in ρ, decreases in the slope of $S(x)$ in the region $x < 0$, and increases in the slope of S in the region $x \geq 0$.

As we show in Appendix A, the relationship between $\mathbb{E}[S]$ and the shape of S holds more generally for any kind of public recognition scheme, such as two-tier public recognition schemes that publicize only the behavior of the top performers. Thus, if, for example, S is concave and people compare themselves to the average ($\rho = 1$), then the two-tier scheme will lead to a net negative image payoff as well. Intuitively, *not* being recognized as a top performer is worse than not having *any* information revealed about oneself, and thus the two-tier scheme cannot avoid inducing some amount of negative image payoff among those in the lower tier. Thus, our findings about the shape of S have implications beyond the fully-revealing public recognition schemes that we study in this paper.

In Appendix B we show that the net image payoff $\mathbb{E}[S]$ connects to a key economic question: whether public recognition is an efficient tool for behavior change relative to standard financial incentives. In addition to $\mathbb{E}[S]$, the other three key inputs to this question are (i) the cost of implementing the public recognition scheme (e.g., due to the need to set up monitoring and distribution of information), (ii) the shadow cost of public funds, and (iii) the extent to which public recognition or financial incentives are best targeted toward people with the highest social marginal value of behavior change.

1.3.3 Structural Versus Reduced-form Estimates of the PRU

Often, the economic questions of interest are about the effects of utilizing public recognition on a whole population, not just the experimental sample. Answering this question requires an additional step of analysis, because scaling up public recognition to more people can change the equilibrium.

To formalize, call $R : \mathcal{A} \to \mathbb{R}$ the *reduced-form public recognition function* which assigns, for each value a, a public recognition payoff $R(a)$. Let R_{exp} denote the function elicited for the experimental population during the experiment, and let R_{pop} denote the reduced-form

[12] In a similar vein, Corneo (1997) models trade union membership as a signaling game between workers, and shows that the reputation effect of trade union membership increases with union density if and only if S is concave.

public recognition function that would result if public recognition was applied to the whole population of interest. These two objects can be meaningfully different: when the public recognition lever is applied to the whole population, population behavior changes, and thus the benchmark for what is considered relatively good behavior may change as well.

As a simple example, suppose that $\rho = 1$ and suppose that in our YMCA setting, an individual is observed to have attended the YMCA four times during the month of the experiment, and that average population attendance is 3.5 attendances. In the context of the experiment, an individual attending four times would thus receive positive public recognition payoffs in the action-signaling model. However, suppose that after applying the public recognition intervention to the whole population, average attendance would increase to 4.5 attendances. Then an attendance of four would actually generate negative public recognition utility. Our reading of existing literature studying social comparisons and social pressure is that it often stops at R_{exp}.[13]

1.4 YMCA Field Experiment

1.4.1 Recruitment

The field experiment was conducted in collaboration with the YMCA of the USA and the YMCA of the Triangle Area in North Carolina (YOTA), and was publicly called "Grow & Thrive." YMCA members of two large YMCA facilities from YOTA were invited via email to sign up for this program by completing a survey. They were informed that for every day that they attended the YMCA during the program month, an anonymous donor would make a $2 donation to their YMCA branch.

The Grow & Thrive program ran from June 15, 2017 to July 15, 2017. On June 1, 2017, the 15,382 members of the two YOTA branches received an email from their local YMCA announcing the launch of a new pilot program aimed at helping YMCA members to stay active and support their community at the same time. The initial email informed participants about the Grow & Thrive program and included a link to an online survey. YMCA members

[13]For example, suppose that individuals' utility in Allcott and Kessler (2019) is a decreasing function of the difference between their energy use and the energy use of the neighbors they are shown. Then the utility that they receive from the information mailer depends on whether the mailer goes out to their neighbors as well. However, since not everyone received the mailer in the experiment, the reduced-form effects that they estimate cannot be used to directly evaluate the policy of sending out mailers to all households. To perform such an evaluation, it would be necessary to take a stand on the structural utility function for social comparisons, to estimate it using the experimental results, and to estimate the counterfactual equilibrium of sending the mailers to everyone in the population.

As another example, consider evaluating individuals' utility from encountering a surveyor who asks about voting behavior. DellaVigna et al. (2017) estimate the utility of doing so after votes have already been cast. But to evaluate the equilibrium impact of increasing the visibility of one's voting behavior, it is necessary to account for the fact that visibility also changes voting behavior, which changes the payoffs one receives from telling a surveyor if one has voted or not. Evaluating the equilibrium outcomes would thus require one to estimate the structural microfoundations of why individuals like to tell others that they voted.

were told that they could sign up for the program by completing the survey and agreeing to participate.[14]

1.4.2 Experimental Protocol

The survey began by explaining the nature of the incentives during the program.[15] Participants were told that an anonymous benefactor with an interest in promoting healthy living and supporting the broader community provided funds to incentivize YOTA members to attend their local YMCA more frequently. During the month of the Grow & Thrive program, a $2 donation was made on each participant's behalf for each day they visited the YMCA, up to a total donation of $60 per person (i.e., 30 visits).

Participants were then told that they might also be randomly selected to participate in the public recognition program. We explained that if a participant was selected into this program, they would receive an email at the end of Grow & Thrive, which would: (1) list the names of everyone in the program; (2) list their attendance during Grow & Thrive; and (3) list the total donations generated by them during Grow & Thrive. We explained that only participants in the public recognition program would receive and be listed in the email. Figure 1.1 provides a screenshot of what this public recognition email entailed.

We then utilized an incentive-compatible Becker-DeGroot-Marschak (BDM) mechanism to elicit participants' (possibly negative) willingness to pay (WTP) for public recognition for various possible realizations of their performance. The incentive-compatible method elicited WTP for public recognition the following possible contingencies of a person's performance: 0 visits, 1 visits, 2 visits, 3 visits, 4 visits, 5 or 6 visits, 7 or 8 visits, 9 to 12 visits, 13 to 17 visits, 18 to 22 visits, and 23 or more visits. For each of the eleven intervals, participants were first asked whether they would want to be publicly recognized if their attendance during Grow & Thrive fell in that interval. Participants were then asked how much they were willing to pay to guarantee that their choice was implemented.

Each of the eleven questions had the following structure: "*If you go to the YMCA [X times] during Grow & Thrive, do you want to participate in the personal recognition program?*" Participants were then asked to state, for each of the eleven levels of possible attendance, how much of an experimental budget of $8 they would be willing to give up to guarantee that their decision about public recognition was implemented. The question asked, "*You said you would rather [participate] [NOT participate] in the personal recognition program if you go [X times] to the Y. How much of the $8 reward would you give up to guarantee that you will indeed [participate] [NOT participate] in the personal recognition program?*"[16] The details were then explained in simple and plain language, and participants

[14] The "pilot" language was important for our field partner, but we recognize that in principle it could have affected people's perceptions about the longer-run consequences of their choices. However, recent work by and de Quidt et al. (2018) and DellaVigna and Pope (2019) suggests that framing effects of this sort seem to have muted effects on behavior. DellaVigna and Pope (2019) also suggest that academics seem to overestimate the extent to which such framing matters.

[15] The Experimental Instructions Appendix contains text and screenshots of the instructions and questions used in the experiment.

[16] Each of these eleven questions was presented to subjects on a separate screen. To make it clear which attendance level was relevant to their WTP elicitation, we highlighted it.

were told, in bold font, that *"it is in your interest to be honest about whether you want to participate in the personal recognition program, and how much of the $8 reward you would give up to ensure that you will or will not participate in the personal recognition program."* Figure 1.2 provides a screenshot from the survey of one of the pairs of questions.

To preserve random assignment, as well as to minimize any negative inferences that could be drawn about those not in the public recognition group, we informed participants that their responses would be used to determine assignment with 10 percent chance, and that with 90 percent chance their assignment would be determined randomly. For participants in the 10 percent, a computer would check their attendance during Grow & Thrive and match it with their answers. With 50% chance they would receive an $8 Amazon gift card and they would be assigned to the public recognition group if and only if they indicated a preference to be in that group. Otherwise, with 50% chance, the BDM mechanism was used to determine the participant's extra reward and assignment to the public recognition group.[17]

To obtain intuition for why truth-telling is incentive compatible with our mechanism, first note that a participant's chance of receiving public recognition is always higher if they indicate a preference for it in the first part of the elicitation. Second, after a participant commits their answer of whether or not they want public recognition, note that the bidding component of the elicitation is just a standard second-price sealed-bid auction against the computer. In summary, the procedure allowed participants to indicate a WTP for public recognition between -$8 and $8. For the 10 percent of participants whose decisions would be used to determine assignment, a bid of $8 guaranteed that the participant would be in the public recognition group, a bid of $0 generated a 50 percent chance of being in the public recognition group, and a bid of -$8 guaranteed that the participant would not be in the public recognition group.[18]

Because others' behavior plays a role in the models summarized in Section 2.2, it was important to help participants have accurate beliefs about others' behavior. Prior to making their decisions about being part of the public recognition program, participants were provided an estimate of the average YOTA monthly attendance in the past year.

In the last component of the survey we elicited participants' beliefs about their future attendance during Grow & Thrive with and without public recognition and under different levels of financial incentives. In this part we also elicited participants' preferences over different financial incentives, which we describe later in the analysis. Finally, we reminded participants that a computer would randomly determine whether they would be part of the

[17] Specifically, the computer generated a random number between 0 and 8, and a participant's preference for being in the public recognition program would be implemented if and only if the participant's WTP was higher than the random number. In this case, the computer's random number was subtracted form the participant's budget. If the computer chose a value greater than the participant's WTP to implement their choice, then the participant's preferred choice for being part of the public recognition program would NOT be implemented, and the participant would receive the $8.

[18] To formally see that this procedure is incentive-compatible, let v be denote a participant's preferences to be publicly recognized at a particular attendance level. Then if a participant indicated a preference for public recognition and bid a value b, their expected payoff would be $\pi_1(b) = \$8 + 0.5v + 0.5(v - b/2)(b/8)$. Conversely, if the participant indicated a preference for no public recognition and bids b to not get it, then the expected payoff is $\pi_2(b) = \$8 + 0.5v + 0.5(-v - b/2)(b/8)$. Clearly, $\pi_1 = \pi_2$ if and only if $b = 0$, with $\pi_1 \geq \pi_2$ if and only if $v \geq 0$. Conditional on $v \geq 0$, the bid that maximizes π_1 is $b = v$. Conditional on $v < 0$, the bid b that maximizes π_2 is $b = -v$.

public recognition group, and we asked them to explicitly agree to participate in Grow & Thrive.

All participants were notified via email about their treatment assignment on the morning of the first day of Grow & Thrive. Participants assigned to the public recognition treatment received a reminder summary of the public recognition treatment when they were notified of their assignment.

All communications with YMCA members took place via email. We prepared an FAQ document covering common questions YMCA members might have about the program. To guarantee the consistency of the responses, and to minimize the burden on YMCA employees, we instructed employees working at the front desk to encourage members to address their questions via email to a specific contact person at the YMCA; the contact person would then use the answers provided in the FAQ to respond.[19]

1.4.3 Attendance Data

We received administrative attendance records from May 1, 2016 to July 15, 2017 for YMCA members in the branches where we conducted the experiment, including those not in Grow & Thrive. Attendances were recorded whenever a member accessed the YMCA facilities by swiping their personal YMCA access card on a turnstile. Before a member could swipe in, a front desk employee verified that the access card belong to the member.[20] We utilize attendance data for non-experimental participants in the out-of-sample predictions in Section 1.8.

1.4.4 Discussion of the Design

What are individuals signaling?

Due to the nature of our setting and the wishes of the YMCA, we were not able to implement a treatment in which participants received public recognition without the Grow & Thrive incentive of raising $2 per attendance for YOTA. As such, we cannot fully differentiate between whether YMCA members were motivated by the desire to be recognized for being health-conscious, or for being charitable. However, like charitable giving, pursuing good health through exercise is also perceived by many as a social and moral obligation (Conrad, 1994; Whorton, 2014; Cederström and Spicer, 2015), and thus it is plausible that both motivations give rise to PRUs of similar structure.

Preferences for signaling versus preferences for information. Although all participants were given the average YOTA monthly attendance from the past year, only the public recognition group received information about others' behavior. To the extent that there was demand for this additional information, our WTP data is an upper bound on demand for public recognition alone. We chose to give any information to individuals only

[19]The YMCA contact reported that only one participant contacted him, asking if he could be added to the public recognition group. After the (negative) response, there were no further questions from the participant.

[20]While YMCA members have to swipe in to access the YMCA, they do not have to swipe out to leave. Therefore we do not have information about how much time participants spent at the YMCA. YMCA employees were told to track any unusual activities among YMCA members. YMCA employees did not report any unusual pattern of access to the facilities during the experiment.

in the public recognition group to better capture the reality of how such interventions are usually implemented. In practice, the counterfactual to a public recognition scheme is not anonymized information provision—it is nothing at all.

Anticipated versus realized image payoffs. Although our approach does not require people to correctly forecast their future attendance, it does rely on the assumption that people can anticipate the image payoffs of public recognition. Testing this assumption would require a design that elicits people's WTP for public recognition after their attendance is realized. This design is significantly less well-powered as it elicits only one data point per person, and thus is left for future work where larger samples can be acquired. However, because people experience shame and pride often, it is likely that they can accurately anticipate the intensity of these feelings, as is consistent with psychological evidence (Sznycer et al., 2016, 2017; Cohen et al., 2020).

1.5 Reduced-form Results From the YMCA Experiment

1.5.1 The Experimental Sample

A total of 428 YOTA members completed the survey and agreed to participate in Grow & Thrive. 192 participants were randomly assigned to participate in Grow & Thrive but not in the public recognition program and 193 participants were randomly assigned to participate in both Grow & Thrive and the public recognition program.[21] 43 participants were randomly assigned to receive the extra \$8 reward for themselves, which they were able to use to affect their likelihood of being publicly recognized. These 43 participants for whom participation in the public recognition program is endogenous are excluded from our empirical analysis.

Unless otherwise noted, from the remaining 385 participants we also exclude 15 participants who indicate a demand for public recognition that has no discernible relation to the number of attendances, and are thus likely confused or disengaged from the study.[22] The remaining *coherent sample* includes individuals whose WTP for public recognition is monotonically increasing in attendance, as well as individuals with preferences that are monotonically decreasing in attendance (i.e., a desire to be recognized as not wanting to attend the YMCA), or individuals with preferences that peak at intermediate levels of attendance (i.e., wanting to look "average"). In Appendix C.8, we also analyze the slightly smaller group of participants whose preferences for public recognition are monotonically increasing in YMCA visits.

Table 1.1 shows that all pre-experiment outcomes, as well as preferences elicited through our online component, are balanced by whether participants were randomly assigned to be in the public recognition group. One noteworthy property of our sample is the high average past attendance of 5.69, which is approximately twice as high as the past attendance of

[21] We randomized our 428 participants into the public recognition group by blocking and balancing over WTP survey responses and attendance in the twelve months preceding the experiment. All participants were notified by the YMCA of the Triangle via email about their treatment assignment the morning of the first day of Grow & Thrive.

[22] The results are qualitatively and quantitatively similar when using all participants.

3.02 of all YOTA members. However, we show below that past attendance does not vary meaningfully with people's preferences over public recognition.

1.5.2 The Effect of Public Recognition on Behavior

Figure 1.3 displays the cumulative distribution functions of attendance by treatment, showing that the impact of public recognition is positive across all levels of attendance. We quantify these results in Table 1.2. The table shows that public recognition increased attendance by approximately 1.2 visits. Given an average attendance of approximately 7 visits in the control group, this corresponds to an approximately 17 percent increase in attendance. This estimate is just outside the range of marginal statistical significance without controlling for participants' past attendance, but becomes highly statistically significant when controlling for participants' past attendance.

1.5.3 Willingness to Pay for Public Recognition

The significant effect of public recognition on behavior suggests that it constitutes a meaningful incentive to participants. Consistent with this, we find that 93 percent of participants have a strict preference to opt in or opt out of public recognition for at least one level of attendance.

Figure 1.4 plots the average WTP by the attendance level that would be publicized to other participants. These WTP profiles constitute model-free measures of the reduced-form PRU R_{exp} introduced in Section 1.3.3. We identify each set of possible visits from our elicitation with its midpoint, meaning that the first five sets $\{0\}, \{1\}, \dots, \{4\}$ are identified with $0, 1, ..., 4$, the "5 or 6 visits" set is identified with 5.5, the "9 to 12 visits" set is identified with 10.5, and so forth. Panel (a) presents data for participants with monotonic preferences for public recognition, panel (b) presents data from participants with coherent but non-monotonic preferences, and panel (c) presents data from the full coherent sample (the combination of panels (a) and (b)). In panels (a) and (c), we also plot the WTP of participants with above versus below median past attendance. The vertical dashed line in the panels corresponds to the average YOTA attendance of 3.14, which is a potential reference standard for positive versus negative image payoffs. As discussed in Section 2.2, the net effect image payoff is decreasing in the magnitude of the reference standard.

On average, as shown in panel (c), the WTP for public recognition is strictly increasing in the number of visits. It is negative at low numbers of visits and positive at high numbers of visits. This pattern is more pronounced in the monotonic panel, as shown in panel (a). Panel (b) shows that the remaining participants with non-monotonic preferences have a distinct WTP profile that peaks at approximately seven attendances and declines steeply afterward. Consistent with this non-monotonic profile, we find an essentially null (but noisy) effect of public recognition on the attendance of these 31 participants (0.39; 95 percent CI $[-2.59, 3.38]$).

Figure 1.4 also shows that participants' PRUs do not vary with their past attendance. We verify this formally in regression analysis in Appendix Table A1. This is important for two reasons. First, because participants in our study had a higher-than-average attendance, and thus a strong interaction between past attendance and WTP for public recognition could limit the external validity of our results. Second, this suggests that participants in our study did not self-select based on sensitivity to public recognition. If low attenders self-selected on being relatively insensitive to public recognition, while high attenders self-selected on being relatively sensitive to public recognition, then the WTP profiles for the above and below median groups in Figure 1.4 would diverge.

Table 1.3 quantifies the descriptive results in Figure 1.4 by presenting regressions of WTP for public recognition on the midpoint of the visits intervals. We present results both from OLS and Tobit regressions. Because some participants' WTPs were at the maximum possible amount of \$8 or the minimum possible amount of −\$8 for some of the elicitation intervals, some preferences were likely to be censored by our elicitation, and thus the Tobit models may give a more accurate assessment of how WTP for public recognition varies with the number of visits. We present linear regressions in odd-numbered columns, and we include a quadratic term for visits in even-numbered columns to study the curvature of the PRU. In this and all subsequent analyses of the WTP data, we cluster standard errors by participant.

All specifications in Table 1.3 generate two robust results, which are visually apparent in Figure 1.4. First, the WTP for public recognition is significantly increasing in the number of visits. Second, this relationship is significantly concave, as implied by the negative coefficient on visits squared.

The quadratic regression models allow us to quantify the curvature of the reduced-form PRU, R_{exp}. One measure of curvature is $-R''_{exp}/R'_{exp}(\bar{a}_{pop})$, where $\bar{a}_{pop}$ is the average attendance of YOTA participants, which is analogous to the coefficient of absolute risk aversion (ARA). Another measure of curvature is $-R''_{exp}/R'_{exp}(\bar{a}_{pop})$ multiplied by the standard deviation of attendance of YOTA participants. This second measure quantifies the percent decrease in R'_{exp} from a one standard deviation change in behavior, and is a unitless measure akin to the coefficient of relative risk aversion (RRA). The unitless property allows us to compare our estimates of curvature across both the YMCA and the charitable contribution experiments.

Table 1.3 shows that while the coefficients in the Tobit models are almost twice as large as the corresponding coefficients in the OLS models, our measure of curvature is very stable. This suggests that while the censoring likely lead to a linear rescaling of the PRU, it did not affect the *shape*.

In addition to censoring, another potential concern is that participants may have been less serious about the WTP elicitation when asked to evaluate public recognition for an attendance level that was outside the range of what they thought was likely. This could lead participants with low expectations of attendance to be relatively insensitive to variation at the upper range of potential visits, and participants with high expectations of attendance to be relatively insensitive to variation at the lower range of potential visits. We investigate this possibility in Figure 1.5 and Table 1.4.

Figure 1.5 presents the WTP data analogously to Figure 1.4, but restricts to data points that involve visits intervals whose midpoints are within 4 visits of individuals' forecasts of attendance in the event that they are randomized into the public recognition group. The standard deviation of the difference between participants' past attendance and their attendance during Grow & Thrive is 4.42, thus visits within 4 of individuals' forecasted attendance should not seem unlikely. Like Figure 1.4, Figure 1.5 shows that WTP for public recognition is strongly increasing and concave in the number of visits, and is close to zero at the YOTA average of 3.14 attendances. The key difference is that the WTP profile in Figure 1.5 is significantly steeper. While the profile in Figure 1.4 spans payoffs between approximately -$2 and $2, the profile in Figure 1.5 spans payoffs between approximately -$4 and $4. This difference is consistent with the possibility that the data reported in Figure 1.4 features some attenuation due to participants being less sensitive to variation in visits that are outside the range of what they consider plausible.

Table 1.4 quantifies the results suggested by Figure 1.5. Columns (1)-(4) present estimates that restrict to data points where the midpoints of the visits intervals are within 4 visits of participants' expected attendance if they are assigned to the public recognition group. Columns (5)-(8) restrict to data points where the visits interval contains participants' expected attendance. Relative to Table 1.3, the estimated coefficients in Table 1.4 are on net almost twice as large. The lack of a meaningful difference between the estimates in columns (1)-(4) versus columns (5)-(8) suggests that the attenuation is mostly due to considering visits that are very far from one's expectations. However, our estimates of curvature are very similar to the estimates in Table 1.3, which suggests that participants' reduced sensitivity to variation in unlikely attendance levels is affecting the scale, but not the shape of the WTP profile. Appendix C.3 shows that the results in Table 1.4 do not vary by past attendance, further reinforcing that past attendance is not a correlate of preferences for public recognition.

While a pure linear scaling bias cannot affect qualitative results about the welfare effects of public recognition, it does affect the magnitudes, as well as the out-of-sample predictions of our structural models. For this reason, our structural analysis in Section 1.8 restricts to data where the midpoint of visits intervals is within 4 of participants' expectations, and utilizes the parametric assumptions of Tobit models to address censoring in the WTP data.

1.5.4 Further Robustness Checks

Potential bias from high visits questions. Appendix Table A3 shows that excluding high visits intervals slightly increases our estimate of curvature. Thus, our estimates are not biased by WTP for attendance in intervals that might fall outside the range of people's expected attendance.[23]

Potential bias from visits intervals increasing in size. To equalize the number of participants whose attendance falls within each bin and to avoid overburdening participants with too many WTP elicitations, we made the possible visits intervals larger at higher attendance levels. One concern is that this could have created an experimenter demand effect by signaling to participants that we expect differences in WTP for public recognition to be approximately constant across the intervals. This, in turn, could lead us to overestimate concavity. To gauge this concern, in Appendix C.5 we index the 11 attendance intervals with the integers 0 through 10, and investigate how WTP for public recognition varies across these index values. We find that WTP for public recognition is significantly concave, and even slight larger, with respect to this recoding of the intervals.[24]

[23] 10 percent of participants expected to attend the YMCA as many as 23 times

[24] To see why the estimate of curvature could increase, recall that quadratic functions are *locally* linear. A quadratic function that has a moderately smaller derivative at say 20 visits than at say 0 visits should in fact have similar derivatives at 0 visits and 10 visits. The fact that we find moderately smaller derivatives

Demand for public recognition as commitment. Individuals with perceived self control problems could in theory try to use our WTP elicitation to motivate their future selves to attend the YMCA more. We argue that this is unlikely for three reasons. First, the method for creating a commitment device using our WTP elicitation is nuanced. This entails individuals lowering expected payoffs for low attendance levels to discourage those low attendance levels. To do so, an individual needs to deviate from "truth-telling" by placing a bid that is not equal to the image payoff at that attendance level. Thus, the bias, if it exists, is unsigned, because the individual can place a bid that is either higher or lower than their true expected image payoff. However, we think it is psychologically unrealistic that individuals would try to manipulate their future behavior in such subtle and sophisticated ways. For example, while individuals could in principle use incentivized belief elicitations as a form of a commitment device, Yaouanq and Schwardmann (2019) provide evidence against this.

Second, as shown by Carrera et al. (forthcoming) and others, demand for commitment is unlikely in environments featuring at least moderate uncertainty about future behavior, such as ours. In our sample, the standard deviation of the difference between attendance in two adjacent months is 4.74, which suggests a level of uncertainty that would likely make dominated incentive schemes costly. Third, in Appendix C.6, we use additional survey questions to analyze whether people's perception of their time inconsistency correlates with their profile of WTP for public recognition, and find no evidence of this. We do this by utilizing the behavior change premium measure developed by Carrera et al. (forthcoming) and Allcott et al. (forthcoming).

1.5.5 Realized Image Payoffs

We end our reduced-form analysis by estimating the realized image payoffs induced by public recognition. We used the reduced-form PRU obtained from our WTP data, together with participants' actual attendance levels, to compute participants' average payoffs by quartile of attendance. To address the potential scaling bias discussed in Section 1.5.3, we estimate payoffs for each level of attendance using the specification in column (4) of the two panels in Table 1.4: we use the Tobit model, and we restrict to WTP data that involves attendance intervals with midpoints within four visits of participants' expected attendance. To compute a participant's realized image payoff, we use the estimated regression to estimate the payoff associated with the participant's realized attendance during the month of the experiment. We present results using the raw WTP data in Appendix A.3.7.

Figure 1.6 presents the results, both for the monotonic and the coherent sample. On average, participants who were publicly recognized received a net-zero image payoff. Participants in the lowest quartile of attendance receive significantly negative payoffs, participants in the second quartile receive somewhat negative payoffs, and participants in the top two quartiles receive significantly positive payoffs.

Importantly, because participants in our experiment have significantly higher YMCA attendance than the average YOTA member, these reduced-form calculations constitute

at an index value of 10 than at an index value of 0 thus implies substantial curvature with respect to the rescaled interval values.

an upper bound on the net image payoff that would result from scaling up our public recognition intervention to the whole YOTA population. This suggests that scaling up the public recognition program to all of YOTA would generate a significantly negative average image payoff, consistent with our findings in Section 1.8.

1.5.6 Over-Optimism and the Benefits of the Strategy Method

A key feature of our design is that our elicitation of people's WTP for public recognition does not require them to form beliefs about their future attendance. In Appendix A.3.2 we assess the accuracy of individuals' beliefs, and find significant overestimation of attendance, consistent with other work (e.g., DellaVigna and Malmendier, 2006; Acland and Levy, 2015; Carrera et al., forthcoming).

Because the PRU is (on average) monotonically increasing in attendance, this misprediction implies that simply eliciting WTP for being in the public recognition program, without conditioning on attendance, would create upward bias in conclusions about the welfare effects of public recognition. Related considerations apply to other social-influence-based interventions, such as the social comparisons studied in Allcott and Kessler (2019).

1.6 Charitable Contribution Experiment

1.6.1 Recruitment

The charitable contribution experiments were administered online to three separate subject pools: (i) members of the online platform Prolific Academic, (ii) participants from UC Berkeley's Experimental Social Science Laboratory (Xlab), who are primarily undergraduate students, and (iii) undergraduate students from a mandatory statistics class, QM222, at Boston University's Questrom School of Business. We refer to these pools as the Prolific, Berkeley, and BU samples, respectively.

For all samples, the experiment ran for one week from April 18, 2020 to April 24, 2020.[25] For the Prolific sample, we recruited only participants who (i) reside in the U.S., (ii) had a 95 percent or higher approval rating, and (iii) had completed at least 15 prior studies on Prolific. For the Berkeley sample, we restricted to participants who had not taken any studies involving deception through Xlab. For the BU sample, all 350 students enrolled in QM222 received an email from their professor inviting them to participate in the experiment.[26] Participants from all subject pools were informed they could only complete the experiment on a laptop or personal computer with a working webcam.

[25]Before the experiment started, we preregistered our design and analysis plan on the AEA RCT Registry (AEARCTR-0005737). We had originally planned to also recruit from the QM221 statistics class for first-year students (who know each other less well than the QM222 students), but the response rate was too low to make use of this data. The results for the limited QM 221 data (N = 52) are in Appendix A.4.3.

[26]The course was broken up into nine classes taught by five professors. Coauthor Robert Metcalfe taught three of the classes.

1.6.2 Experimental Protocol

Except where noted below, the experimental protocol was identical for each of the three samples.[27] Perhaps the biggest implementation difference was the difference in incentive levels. Relative to the Prolific sample, we scaled up all incentives by a factor of 2.5 in the Berkeley and BU samples. This was done to reflect differences in payment norms across the samples. Prolific requires researchers to pay all participants at least \$6.50 per hour, Berkeley Xlab requires researchers to pay at least \$20 per hour, and BU requires researchers to pay at least \$15 per hour.

In the experiment, participants could raise money for the Red Cross by successively pressing the "a" and "b" keys on the computer. Each pair of button presses earned a point, which translated to money donated to the Red Cross by the experimenters, and in some cases also to additional payments to the participants.

After consenting to participate in the experiment, participants first reviewed instructions about the button-pressing task. Participants then practiced the task for up to 30 seconds.

Participants were then presented with an overview of the structure of the experiment. Participants were told that they would complete three rounds of the button-pressing task (presented in random order), and that each round would last up to 5 minutes. We gave participants the option to finish each round early, since this "extensive margin" option appears to lead to more elastic labor supply, as suggested by DellaVigna et al. (2019), DellaVigna and Pope (2019), and our own pilots.

In all rounds, participants in the Berkeley and BU samples raised 5 cents for the Red Cross for every 10 points that they scored, while participants in the Prolific sample raised 2 cents for every 10 points. In the Anonymous Effort Round, this was the only incentive, and participants' performance remained anonymous. In the Anonymous and Paid Effort Round, participants also earned financial compensation for themselves, which was identical to their Red Cross contribution (5 cents/10 points in the Berkeley and BU samples, and 2 cents/10 points in the Prolific sample). Participants' performance in this round also remained anonymous.

In the Publicly-Shared Effort Round, participants' performance would be revealed to all participants in their experimental group after the conclusion of the study. In this round, participants' effort only translated to Red Cross donations, not to their own earnings. Specifically, after the end of the study, all participants would receive a link to view the pictures and contributions raised for the Red Cross of all participants in their group who were assigned to have their effort publicly shared with others. The information shared would include participants' photos, their scores and donations in the button-pressing task, their ranks relative to other publicly-recognized participants and, for the Berkeley and BU samples, their names.[28] All participants were required to take a picture of themselves using their webcam, and they were given the option to upload an alternative picture or retake their picture. In summary, we included one baseline round where participation remained

[27]The Experimental Instructions Appendix contains text and screenshots of the instructions and questions used in the experiment. An online example of the experiment is available here: https://wharton.qualtrics.com/jfe/form/SV_2mImcVP4XP3Pmf3.

[28]We did not collect and reveal participants' names in the Prolific sample because this would violate the platform's privacy requirements.

anonymous, one round where participants earned performance-based financial compensation, and one round where effort was publicly recognized.

Each round had a 30 percent chance of being randomly chosen to determine a participant's outcome. With 10 percent chance, participants' preferences for public recognition would be used to determine whether their performance would be publicly recognized or remain anonymous—we called this the Choose Your Visibility option.

The Choose Your Visibility option involved an incentive-compatible elicitation procedure that was analogous to that of the YMCA experiment. We asked eighteen pairs of questions about WTP for public recognition, corresponding to eighteen possible intervals of performance. The eighteen intervals were 0-99 points, 100-199 points, ..., 1600-1699 points, and 1700 or more points. For each interval, we first asked participants if they wanted their effort to be publicly shared if it fell in one of those intervals, and we then asked them to state their WTP to have their preference implemented. Participants were given a $10 budget for this elicitation in the Prolific sample, and a $25 budget in the Berkeley and BU samples. As in the YMCA experiment, we told participants, in bold font, that "carefully and honestly answering the questions is in your best interest."

If the Choose Your Visibility option was randomly chosen to determine a participant's outcome, then the score from one of the three rounds was randomly chosen to determine the participant's contribution to the Red Cross. However, the webpage identifying participants' contributions did not differentiate between participants who were randomly assigned to be in the Publicly-Shared Effort Round and participants assigned to the Choose Your Visibility option—all recognized participants and their contributions were presented identically. Thus, the proper inference about any publicized participant is that their score was probably based on the Publicly-Shared Effort Round, and that the reason their contribution was publicized was likely due to random chance rather than because of the preferences elicited in the Choose Your Visibility option. This procedure also ensured that participants' performance in all three rounds carried equal importance and, by creating some uncertainty about the score used, broadened the range of scores that participants could consider relevant for the Choose Your Visibility elicitation.

Because others' behavior can play a role in social image utility, we first collected an initial round of data to provide participants with signals of others' performance in the Publicly-Shared Effort Round. Participants in the Prolific sample were presented with information from a 79-person pilot, and participants in the Berkeley and BU samples were given information from a 52-person pilot. Participants were informed of the average performance and the 25th, 50th, and 75th percentiles of performance from these samples. Participants were also informed of the sample size of the data, and were also provided a link to view a full CDF of past performance.

For the Berkeley and Prolific samples, participants were also informed about the size of their experimental group. In the Berkeley sample, participants were randomly divided into groups of approximately 75 participants, and they were told that approximately 25 participants in their group would have their effort publicly shared with all others in the group. In the Prolific sample, participants were randomly assigned to be in a group of 300, 75, or 15 participants, and were told that approximately 100, 25, or 5 participants in their respective group would have their effort publicly shared with all others in the group. We did not include language about group size in the BU sample because we did not have a sufficiently precise prediction about the response rate to provide truthful information. Importantly, the group assignment in the Prolific and Berkeley samples was completely random, which implies that standard errors need only be clustered at the participant level in all analyses.

The timing of the experiment was as follows. First, participants learned about the three rounds and the Choose Your Visibility option. Second, participants received information on past performance and their group size, and answered an attention check question that instructed them to leave the question blank and advance to the next screen. Third, participants indicated their preferences for public recognition in the Choose Your Visibility option. Fourth, participants completed the three button-pressing rounds. The order of the rounds was fully randomized. In each round, participants were reminded of the conditions of the round. In the Publicly-Shared Effort Round, participants were also shown the image that would be seen by other participants.

Participants were informed of what round was randomly selected to count as soon as they completed the study. Within three days of the end of the study, participants were randomly divided into groups and were sent a link to view the performance information of all participants in their group who were assigned to have their effort publicly shared with others. Participants had 72 hours to view this information, and could only access it by entering the Prolific ID or university email address they had entered when completing the study. If participants clicked to view the additional information, they would receive an additional \$0.50 if in the Prolific sample, or \$1 if in the Berkeley or BU samples. The experimenters did not match the identities and scores of any participant who was not selected to be publicly-recognized, and the participants were informed that they would be anonymous even from the experimenters if they were not assigned to be publicly recognized.

1.6.3 Discussion of the Design

Within-person variation We chose to have participants complete all three possible rounds for two reasons. First, this ensured that there would not be differential attrition. In a between-subjects design where each participant completed only one of the three rounds, a realistic possibility is that participants might be more likely to attrit from conditions in which they did not receive additional pay for their performance, or conditions in which they might incur negative image payoffs. Second, our design maximizes statistical power for comparisons of performance across the three rounds, and allows for some additional analyses of individual differences.

Relation to the YMCA experiment The charitable contribution experiments complement the YMCA experiments in five key ways.

First, the experiments explore a different domain, and one that is arguably a more common target of public recognition: giving time and effort to charity. This permits an initial investigation of the cross-domain stability of various aspects of people's preferences over public recognition.

Second, by simultaneously running the experiment on three different samples, we are able to explore cross-population stability. One notable difference between our three samples is people's familiarity with each other.

Third, the charitable contribution experimental design more directly eliminates the possibility that participants might use the WTP for public recognition elicitation as a type of commitment device. There is only a 5-15 minute gap between when participants complete the elicitation and when they begin the real-effort rounds, and thus all of these decisions are likely to be regarded as "now." Augenblick's (2018) estimates of discounting in real-effort tasks similar to ours strongly support this interpretation.[29]

Fourth, the large size of the Prolific sample allows us to analyze how group size might affect participants' preferences to be publicly recognized. This analysis is helpful for refining out-of-sample predictions that involve larger groups than those in the experiment. The possible effects of group size can be captured by the ν parameter in the structural models in Section 2.2, but the effects are ambiguous. On the one hand, larger group sizes imply larger audiences. On the other hand, larger group sizes imply that any recognized participant is likely to receive less attention.

Fifth, the charitable contribution experimental design has other features that make analysis and interpretation more straightforward: (i) the design provides subjects not just with the mean of past performance, but with the whole distribution, which could be important if people care about statistics other than average performance; (ii) the design has a significantly larger allowable range in the WTP elicitation, which essentially eliminates all censoring; (iii) the elicitation interface has evenly-sized performance intervals, which eliminates potential worries about what participants might infer from variable interval widths; (iv) all participants, not just those publicly recognized, see the performance of the publicly-recognized group, which implies that WTP for public recognition cannot be affected by a demand for additional information.

1.7 Reduced-form Results from the Charitable Contribution Experiment

1.7.1 The Experimental Samples

1017, 407, and 121 participants completed the Prolific, Berkeley, and BU experiments. We make two preregistered exclusions for our analysis. We exclude participants failing the attention check, and we exclude participants with "incoherent" preferences for public recognition, where "incoherent" is defined analogously to the YMCA analysis. This yields a final sample of 968, 384, and 118 participants in the Prolific, Berkeley, and BU experiments. Out of the remaining participants, almost all (all but 1.0, 1.8, and 1.7 percent of Prolific, Berkeley, and BU participants, respectively) had monotonically increasing preferences for public recognition, and our results are qualitatively and quantitatively unchanged if we

[29] Augenblick (2018) estimates discount factors for a real-effort task very similar to ours at time horizons varying between a few hours and seven days, using the Berkeley Xlab pool. The estimates imply no plausible discounting for time horizons that are shorter than 15 minutes. For example, while Augenblick (2018) estimates a discount factor of 0.87 for a 7-day horizon, he estimates discount factors of 0.91 and 0.94 for 24-hour and 3-hour horizons, respectively. Extrapolating with any reasonable parametrization of the discount factor to a horizon of 0.15 hours would imply virtually no discounting at that horizon.

restrict to this monotonic sample. Thus, to simplify the analysis, we present results only for the coherent sample.

In this final sample, Prolific participants were divided into 17 groups of 13-15 participants each, 6 groups of 71-79 participants each, and 1 group of 278 participants. All Berkeley participants were divided into 5 groups of 75-79 participants each, and all BU participants were in the same group.

There was minimal censoring in the WTP for public recognition elicitation. Prolific, Berkeley, and BU participants chose to use all of their budget in only 6, 4, and 6 percent of all cases, respectively.

Our 100-point intervals in the WTP elicitation generated nearly complete coverage of the distribution of effort. Only 1.1, 2.6, and 2.0 percent of scores in the Prolific, Berkeley, and BU samples, respectively, were 1700 points or higher.

The average age was 35, 21 and 20 for the Prolific, Berkeley, and BU samples, respectively. The percent of Prolific, Berkeley, and BU participants who identified as female was 50, 69, and 51 respectively.

The averages of the standard deviations of the difference in points scored between any two rounds were 390.9 points, 423.4 points, and 469.7 points in the Prolific, Berkeley, and BU samples, respectively. These scores suggest a fair amount of uncertainty about the score that would be used if selected for the Choose Your Visibility option.

1.7.2 The Effects of Public Recognition on Behavior

Figure 1.7 displays the cumulative distribution functions of points scored by treatment, showing that the impact of public recognition is positive across all levels of points scored in each of the three samples. The figure also suggests that the effect of public recognition is about half of the effect of financial incentives in the Prolific sample, and is only somewhat smaller than the effect of financial incentives in the Berkeley and BU samples.

Table 1.5 quantifies the effects depicted in Figure 1.7. The table reports results from OLS regressions of points scored on the experimental round. Column (1) presents results from the Prolific sample, column (2) presents results from the Berkeley sample, and column (3) presents results from the BU sample. Column (4) analyzes whether the effects of public recognition in the Prolific sample vary by group size. In all columns, we control for the order of the round by including dummies for whether the round appeared first, second, or third to a given participant, although the F-tests presented in Table 1.5 do not detect any fatigue or other order effects. We cluster standard errors at the participant level in this all and subsequent analyses.

As columns (1)-(3) of Table 1.5 show, public recognition increases participants' total effort by over 10 percent in all three rounds, which is highly statistically significant. The effects of the financial incentive are substantially larger in the Prolific sample, and modestly larger in the Berkeley and BU samples. Column (4) presents preliminary evidence that the three different group sizes considered in our Prolific experiment do not seem to moderate the effects of public recognition. Thus, the results suggest that the effect of a larger audience is offset by the decrease in attention any recognized individual receives.

Robustness We find no evidence that within-subject estimates differ from between-subject estimates. Table A.4.1 in Appendix A.4 analyzes pure between-subject variation by limiting to the first round the participants encountered. The effects of public recognition and financial incentives are virtually identical to the within-subject estimates in the Prolific and Berkeley samples. The effects of both public recognition and financial incentives are substantially smaller in the BU sample, although they are measured very imprecisely due to the small size of this sample, and the confidence bands include the within-subject estimates.

1.7.3 Willingness to Pay for Public Recognition

Consistent with the significant effect of public recognition on behavior in all three samples, we find that 73 percent, 78 percent, and 89 percent of participants in the Prolific, Berkeley, and BU experiments, respectively, have a non-zero WTP for public recognition at one or more levels of performance.

Figure 1.8 plots the WTP for public recognition by level of publicized effort to raise money for the Red Cross, measured in points. We identify each interval below 1700 with its midpoint, so that the first interval corresponds to 50 points, the second interval corresponds to 150 points, and so forth. The last point in the figure corresponds to the "1700 or more" points interval. Panel (a) presents data from the Prolific sample, panel (b) presents data from the Berkeley sample, and panel (c) presents data from the BU sample. In addition to the sample averages, each panel also summarizes the WTP for participants with above and below median performance in the Anonymous Effort round. In all three panels, the vertical dashed line corresponds to the average score in the Publicly-Shared Effort round, which is a potential reference standard for positive versus negative image payoffs. As discussed in Section 2.2, the net image payoff is decreasing in the magnitude of the reference standard.

On average, WTP for public recognition is strictly increasing in points scored in all three samples. In all samples, it is negative at low levels of points scored and positive at high levels of points scored. Figure 1.8 also shows that participants' PRUs do not vary meaningfully with their score in the Anonymous Effort Round, suggesting that preferences for public recognition do not vary meaningfully with their cost of effort or intrinsic motivation to help the Red Cross. Figure A.4.2 in Appendix A.4 presents confidences intervals for the average WTP in each interval.

Table 1.6 quantifies the descriptive results in Figure 1.8 by presenting results from regressions of WTP for public recognition on effort to raise money for the Red Cross, measured in points. Because very few participants' responses are censored at their full budget, we report results from OLS regressions only. The results are virtually identical in Tobit regressions. Columns (1) and (2) report results from the Prolific sample, columns (3) and (4) report results from the Berkeley sample, and columns (5) and (6) report results from the BU sample. We present linear regressions in odd-numbered columns, and we include a quadratic term for visits in even-numbered columns to study the curvature of the PRU. For this analysis, we exclude the ≥ 1700 points interval as it does not represent a narrow band of performance like the other intervals. We make this exclusion in other analyses unless otherwise noted.

Consistent with Figure 1.8, all regressions imply that the WTP for public recognition is strongly increasing in the level of publicized effort. The implications for curvature are more mixed. The regressions imply significant concavity in the Prolific experiment, and smaller but imprecisely estimated levels of curvature in the Berkeley and BU samples. In the Berkeley and BU samples, we cannot reject linearity, although the 95 percent confidence intervals for curvature, $-R''/R'(\bar{a}_{pop})$, also include the point estimate from the Prolific sample.

Appendix A.4.4 uses the Bayesian Information Criterion (BIC) to formally show that the linear and quadratic models in Table 1.6 are the best fit to the data presented in Figure 1.8. We show that the slight convexity visible around some multiples of 500 is best explained by moderate "round number bias." When dummies at multiples of 500 are included, higher-order terms beyond the quadratic specification are estimated to be near 0. Second, the round number bias is sufficiently small that the BIC-minimizing models are a quadratic polynomial (without dummies at multiples of 500) in the Prolific sample and a simple linear model (without dummies at multiples of 500) in the Berkeley and BU samples.

The slight uptick in the WTP at the $\geq$ 1700 interval is consistent with theory, as individuals should assign a particularly high WTP to that interval if they believe that a score in that interval is perceived to be substantially higher than 1750. The mean performance conditional on being in that interval is 1791.6 (SE 28.2), 1871.6 (SE 94.2), and 1884.7 (SE 86.4) in the Prolific, Berkeley, and BU samples. Appendix Figure A.4.3 plots a variation of Figure 1.8 where the location of the $\geq$ 1700 interval on the x-axis is set equal to the average score in that interval; the figure reveals no trend-break around that interval. Consistent with this, Appendix Table A.4.3 replicates Table 1.6 on the full data that includes the $\geq$ 1700 interval, and finds essentially identical regression estimates.

We can compare our unitless measures of curvature, $-R''/R'(\bar{a}_{pop})$ multiplied by the standard deviation of behavior, across the YMCA and charitable contribution experiments. In the charitable contribution experiments, we use the standard deviation of behavior in the anonymous round. Column (2) shows that our estimate of normalized curvature in the Prolific sample is strikingly similar to the estimates in Tables 1.3 and 1.4 for the YMCA sample. The analogous estimates for the Berkeley and BU samples in columns (4) and (6) are smaller in magnitude, although the 95 percent confidence intervals include all point estimates from Tables 1.3 and 1.4. Overall, in the Berkeley and BU samples we can neither reject linearity nor the degree of curvature estimated in the YMCA and Prolific samples.

Any potential differences in WTP data between the Prolific, Berkeley, and BU samples are unlikely to be explained by group size. Consistent with our results about the effects on behavior not being affected by group size, Appendix Table A.4.6 shows that there is no interaction between group size and WTP for public recognition in the Prolific sample. We estimate fairly precise null effects for all interactions, which supports the hypothesis that the effect of a larger audience is offset by the decrease in attention any recognized individual receives.

Robustness and Heterogeneity In the YMCA experiment, participants' elicited WTP for public recognition was less sensitive to variation in performance that was outside the range of what they construed as likely behavior for themselves. We investigate this possibility in the charitable contribution experiments in Appendix Table A.4.2, which presents results from regressions analogous to those in Table 1.6, but restricting to data where the intervals for which WTP is elicited are within 500 points of participants' average performance in the three rounds. The estimates in Appendix Table A.4.2 are almost identical to those in Table 1.6. This is perhaps due to the fact that participants who have experienced economics experiments are better at answering more hypothetical/abstract questions.

We find some evidence for heterogeneity in preferences for public recognition, but consistent with our YMCA results, we find that these preferences do not covary with intrinsic motivation to raise money for the Red Cross, as measured by performance in the Anonymous Effort round. Table A.4.4 in Appendix A.4 shows that participants with an above-median difference in scores between the public and anonymous rounds also have a steeper PRU—that is, their WTP for public recognition is more steeply increasing in performance. This interaction is significant in the Prolific and BU samples in linear regressions of WTP on performance, but is more noisily estimated in the smaller BU sample, and in regressions that include a quadratic performance term. On net, these results suggest some stable individual differences in preferences for public recognition: some participants have steeper PRUs, and thus their performance is more sensitive to public recognition. Appendix E uses random coefficient models to estimate heterogeneity in PRUs more directly, and shows that while there is indeed significant heterogeneity in sensitivity to public recognition, there is little heterogeneity in the curvature of PRUs. Appendix Table A.4.5 shows that there is no relationship between the PRU and participants' intrinsic motivation, consistent with the graphical evidence in Figure 1.8.

1.7.4 Realized Image Payoffs

Finally, we estimate the net image payoff induced by public recognition. We do this by assigning to each participant the average WTP for public recognition that corresponds to the interval containing the participant's score in the Publicly-Shared Effort Round. We use the sample average WTP, instead of the participant's own WTP, to maximize statistical power. As discussed above, the PRU does not vary with participants' intrinsic motivation or with their score in the public recognition round, and thus using average WTP for a given interval increases statistical power without creating bias.

Figure 1.9 presents the results. The net image payoff of public recognition is positive in the Prolific sample, statistically zero in the Berkeley sample, and is negative in the BU sample. The bottom quartile of participants experiences significantly negative payoffs in all three samples. In the Prolific and Berkeley samples, the top three quartiles of participants all experience positive payoffs, while in the BU sample no quartile of performers experiences positive payoffs.

Although there are many differences between the three samples, one key difference is the degree of familiarity among participants.Our results provide suggestive evidence that greater familiarity increases the prevalence of shame, which is consistent with hypotheses and results from psychological research (e.g., Tajfel, 1970; Hogg, 1992; Bicchieri et al., 2020).

1.7.5 Consistency with Financial Incentive Effects

Before turning to structural estimation, we provide back-of-the-envelope calculations to validate our money-metric approach to measuring the PRU. The fundamental assumption of our approach is that the effects of public recognition on behavior can be fully captured by the money-metric measures of the PRU in Table 1.6. For example, column (1) of the table implies that the motivating effects of public recognition are approximately equivalent to a financial incentive of 0.93 cents per 10 points in the Prolific sample. Thus, a key test of our approach is whether a financial incentive of 0.93 cents/10 points indeed has a similar effect on behavior in the Prolific sample as does public recognition.

Simple calculations suggest remarkable consistency. In the Prolific sample, column (1) of Table 1.5 shows that public recognition increases performance by 105 points. A linear extrapolation thus implies that a 2 cent/10 points incentive should increase performance by $105 \times (2/0.93) = 226$ points, which closely matches the 186-point effect estimated in column (1) of Table 1.5. Analogous arguments imply that our Table 1.6 estimates imply that the financial incentive should increase performance by 216 and 150 points in the Berkeley and BU samples, respectively. Empirically, Table 1.5 reveals only slightly smaller effect sizes of 178 and 118 points, respectively. Our structural estimates in the next section facilitate more formal tests of consistency.

1.8 Structural Estimates

Our results thus far provide estimates of the reduced-form public recognition function R_{exp}. In this section, we build on the reduced-form results in three ways. First, we estimate parametric forms of the models presented in Section 2.2. Second, we validate our experimental and structural methodology by more formally implementing the consistency tests from Section 1.7.5. Third, we study the welfare effects of scaling up the public recognition intervention. Our main focus is on scaling up in the YMCA setting because it constitutes an important domain of behavior where there is significant interest in behavior change, and where social influence interventions such as ours are of potential interest. Appendix A.6 contains the details of the structural models, their equilibrium predictions, and our approach to identifying these models.

1.8.1 Estimation Methodology

Functional form assumptions For tractability, we follow Bénabou and Tirole (2006) in assuming that in the absence of public recognition, people's material utility u is quadratic:

$$u(a;\theta) = \theta a - ca^2/2,$$

where $\theta \in \mathbb{R}^+$ is the intrinsic motivation, and ca is the marginal cost of increasing a. We also assume that the structural PRU in both the action-signaling and characteristics-signaling models in Section 2.2 is quadratic. Letting $\bar{a}$ denote the average action, and $\bar{\theta}$ denote the average type, we assume that

$$\nu S^a(a - \rho\bar{a}) = \gamma_1^a(a - \rho\bar{a}) + \gamma_2^a(a - \rho\bar{a})^2 \tag{1.5}$$

$$\nu S^\theta(\mathbb{E}[\theta|a] - \rho\bar{\theta}) = \gamma_1^\theta(\mathbb{E}[\theta|a] - \rho\bar{\theta}) + \gamma_2^\theta(\mathbb{E}[\theta|a] - \rho\bar{\theta})^2 \tag{1.6}$$

for the action-signaling and characteristics-signaling models, respectively.[30] As shown in Appendix A.6, the resulting reduced-form PRU, $R(a)$, will be quadratic with both micro-foundations.

To close the models, it is necessary to take a stand on the comparison sample that generates $\bar{a}$ and $\bar{\theta}$. In the YMCA setting, where participants were members far before the experimental period, and where they have the opportunity to observe and interact with many members outside of Grow & Thrive, the most natural assumption is that individuals care about how they are seen relative to the other YOTA members of their YMCA branch.[31] In our charitable contribution experiments, by contrast, participants did not have a previously-established connection to the task, as the task was only introduced to them in the experiment. We thus assume that participants' comparison populations are simply those individuals who also completed the task—our experimental samples.[32]

Estimation Let $R_{exp}(a) = r_0 + r_1 a + r_2 a^2$ be the reduced-form PRU that is revealed by our WTP elicitation. We estimated this directly in column (4) of Table 1.4b for the YMCA sample, and in columns (2), (4), (6) of Table 1.6 for the Prolific, Berkeley, and BU samples.[33] As shown in Appendix A.6, estimates of the structural parameters γ_i^j and ρ from the structural PRUs in (1.5) and (1.6) can be obtained as functions of the reduced-form parameters r_0, r_1, r_2.

Given estimates of R_{exp}, the treatment effect of public recognition on behavior identifies the cost parameter c. In the absence of public recognition, the marginal benefits of increasing a are θ, and the marginal costs of increasing a are ca. Thus, individuals choose $a^*(\theta) = \theta/c$, and average performance in the absence of public recognition is

$$\mathbb{E}[a|PR = 0] = \mathbb{E}[\theta]/c. \tag{1.7}$$

In the presence of public recognition, the marginal benefits of increasing a are $\theta + r_1 + 2r_2 a$. Thus, individuals choose $a^*(\theta) = (\theta + r_1)/(c - 2r_2)$, and average performance in the presence of public recognition is

$$\begin{aligned}\mathbb{E}[a|PR = 1] &= \mathbb{E}[\theta]/(c - 2r_2) + r_1/(c - 2r_2) \\ &= \mathbb{E}[a|PR = 0] \cdot c/(c - 2r_2) + r_1/(c - 2r_2)\end{aligned} \tag{1.8}$$

Given an estimated average treatment effect $\bar{\tau}$ of public recognition on performance, the cost parameter c is identified by setting the difference between (1.8) and (1.7) equal to $\bar{\tau}$. We use the treatment effect estimates from column 5 of Table 1.2 for the YMCA sample, and estimates from columns (1)-(3) of Table 1.5 for the Prolific, Berkeley, and BU samples.

[30] To ensure that S is increasing, we further assume that $a \in [0, \bar{a}]$ and that $\gamma_1^j + 2\gamma_2^j \bar{a} \geq 0$.

[31] Moreover, individuals had little reason to expect that participants in Grow & Thrive were different from other YMCA members since we only provided information about the broader base of YOTA members.

[32] An alternative benchmark might be the hypothetical performance of all Prolific, Berkeley Xlab, or BU Section QM222 members. This assumption is equivalent to ours if our experimental participants believed the participants in our experiment were representative of these larger pools.

[33] As discussed in the reduced-form results, the specification in column (4) of Table 1.4 for the YMCA sample addresses potential attenuation resulting from censoring, and from participants' relative insensitivity to variation of publicized attendance that they consider unlikely.

Consistency with financial incentive effects The calculations above show that the structural models are identified using only data on the treatment effects of public recognition and participants' WTP for public recognition. The estimated models can then be used to make predictions about the effects of financial incentives on behavior, which can be compared to direct estimates from our data. In the presence of a constant marginal incentive of p and no public recognition, the marginal benefits of increasing a are $\theta + p$, and the marginal costs are ca. This implies that individuals choose $a^*(\theta) = (\theta + p)/c$, and thus that the financial incentive increases average performance by p/c.

For the charitable contribution experiments, we benchmark the model predictions against the effects of financial incentives estimated in Table 1.5. For the YMCA experiment, we were not able to randomize a purely financial incentive, but we did elicit participants' forecasts of how much they would attend the YMCA under three different scenarios: (i) if assigned to the Grow & Thrive control group; (ii) if assigned to the Grow & Thrive public recognition treatment group; (iii) if assigned to the Grow & Thrive control group but given a financial incentive of \$1 per attendance. Although forecasted attendance may differ from actual attendance due to overoptimism, Carrera et al. (forthcoming) find that people accurately predict how their attendance will *vary* with incentives for attendance. Consistent with this, participants in our experiment predicted that public recognition would increase their attendance by 1.50 visits, which is similar to, and statistically indistinguishable from, our empirical estimate of 1.19 visits.

Note that the predictions about the effects of financial incentives on behavior in the experiment depend only on the reduced-form PRU R_{exp}, and thus are identical for both the action- and characteristics-signaling models.

Heterogeneity In Appendix A.6.3 we generalize the model to include heterogeneity in individuals' cost of effort functions and PRUs, and show that our estimation approach is robust to this.

Uncertainty Suppose that at the time of the WTP elicitation, individuals are unsure about their type θ or the marginal costs, and that they learn this only after the elicitation when they choose their performance a. For example, individuals might be unsure about how motivated they will feel to work hard in the Click for Charity task, and only accurately learn that when they begin the task. This does not affect our analysis because of the strategy-method nature of our elicitation. All of our computations pertain to the signaling game that is played once individuals learn their type. This signaling game leads to the reduced-form PRU R, and our WTP elicitation exactly elicits $R(a)$ for each a. This robustness rests on the key feature of our design that WTP for public recognition is elicited in a performance-contingent fashion.

1.8.2 Estimation Results

Table 1.7 presents the structural estimation results. Panel (a) presents estimates of the action-signaling model and panel (b) presents estimates of the characteristics-signaling model. Panel (c) presents results on consistency with the effects of financial incentives.

Although the model parameters γ_i^j in panels (a) and (b) are in different units and thus have different magnitudes, the two panels deliver a similar message, which is consistent with the reduced-form results. First, there is significant concavity of the structural PRU in the YMCA and Prolific samples, although the curvature estimates are more ambiguous in the Berkeley and BU samples. The concavity is particularly pronounced in the characteristics-signaling model in the Prolific sample. Second, the standard at which negative image payoffs transition to positive image payoffs varies across the samples. In the YMCA sample, ρ is above 1 in both models, although we cannot reject the hypothesis that participants simply care about the average ($\rho = 1$). In the Berkeley sample, we estimate ρ close to 1 in both models. In the Prolific sample, we estimate ρ significantly below 1 in both models, indicating a lower standard for pride-worthy behavior. In the BU sample we estimate ρ substantially above 1, indicating a high standard for pride-worthy behavior.

Panel (c) shows that in all four samples, the models' predictions about the effects of financial incentives closely match the directly estimated effects. On net, we find slight overestimation, although the last column in panel (c) shows that this overestimation is not statistically distinguishable from zero at conventional levels. Moreover, the slight overestimation could be explained by a number of realistic features not incorporated into our intentionally parsimonious models.[34]

1.8.3 Welfare Effects of Scaling up Public Recognition

We now use our structural estimates to assess the average image utility generated by public recognition. Motivated by our results on group size effects in the Prolific sample, we assume that increasing the number of exposed individuals would not change the visibility parameter ν.

Under the assumption that our Prolific, BU, and Berkeley samples are representative of those respective populations, and that individuals in those samples construct the reference point from how the samples performed in the public recognition round, the welfare effects are immediately given by our reduced-form results in Section 1.7, and are summarized in Table 1.8

For the YMCA sample, however, the natural assumption (discussed above) is that individuals evaluate their performance relative to the performance of all members of YOTA. This implies that our reduced-form estimates of welfare effects are only partial equilibrium, and necessitates the use of our structural model. This need is particularly pronounced because the YMCA sample is not representative of the broader YOTA population.

We present the results in Table 1.9. Column (1) shows the net image payoffs and column (2) presents the predicted change in behavior. Panel (a) presents results from the action-signaling model and panel (b) presents results from the characteristics-signaling model. Except in several special cases, these models have somewhat different equilibrium

[34] For example, our quadratic cost of effort function implies a unit elasticity and thus that behavior is linear in the magnitude of incentives. This assumption would cause us to overestimate the effects of financial incentives if instead behavior were a concave function of financial incentives, as would be the case for isoelastic cost functions with elasticities below one. Various forms of correlated heterogeneity could explain the underestimation as well.

implications for behavior and welfare, illustrating the importance of working out the consequences of microfounded models.

We explore the welfare effects across a range of different structural assumptions. Row (1) in both panels considers the baseline estimates for the YMCA sample. Rows (2)-(4) explore the importance of varying ρ by considering the point estimates from the Prolific, Berkeley, and BU samples. Rows (5)-(7) consider the importance of varying curvature by using the point estimates from the Prolific, Berkeley and BU samples. Rows (8)-(10) jointly set ρ and curvature equal to the point estimates from the Prolific, Berkeley, and BU samples.

The table reveals two main insights. First, the average image utility from scaling up public recognition to the full YOTA population is predicted to be substantially negative, particularly in the action-signaling model.

Second, as rows (2)-(10) illustrate, variation in the reference point parameter ρ has a larger effect on net images payoffs than variation in curvature. Decreasing ρ to the Berkeley sample estimate, while holding curvature fixed at the YMCA estimate, results in a net image payoff near 0. Further reducing ρ to the estimate in the Prolific sample results in a positive image payoff. However, rows (5)-(7) show that holding ρ constant at the YMCA estimate and varying the curvature to match the estimates in the online samples always results in negative image payoffs. The welfare estimates in rows (8)-(10) are much close to those in rows (2)-(4) than in rows (5)-(7). This implies that the large variation in social image payoffs between all four of our samples is largely due to variation in the ρ estimate.[35]

In Appendix A.2 we formalize how the estimates of image utility in this section can be combined with several other statistics to determine whether public recognition or financial incentives are a more efficient means of changing behavior.

1.9 Concluding Remarks

A recent and growing literature establishes that public recognition can meaningfully influence behavior in a number of economically consequential field settings. We build on this literature by developing an empirical methodology for directly quantifying individuals' utility from public recognition. Across two different experimental designs and four different samples, we find that image payoffs from public recognition are significant and highly unequal: some experience significantly negative payoffs, consistent with shame, while other experience significantly positive payoffs, consistent with pride. In the YMCA setting, our results suggest that motivating exercise with public recognition might be less socially efficient than utilizing financial incentives. Our work illustrates how the social costs or benefits of public recognition can be substantial, and provides a framework for measurement and welfare analysis.

Of course, our results come with many caveats and leave open many research questions. First, our methods quantify only the direct effects of public recognition on utility, and are not designed to measure other key inputs for a wholistic welfare analysis. Appendix A.2 provides a formal framework for welfare analysis, and in particular for answering whether

[35]Note that the impacts on behavior, in percentage terms, are predicted to be larger in rows (8)-(10) than in Table 1.8 because the distribution of individuals' types is different. In particular, the effects will be proportionally larger for more left-skewed distributions.

another policy lever, such as financial incentives, might be more efficient in creating the same behavior change.[36]

Second, while our methodology is easily imported into many of the domains where researchers have studied the effects of public recognition on behavior, our specific results constitute only an initial set of data points on the welfare effects of public recognition. Consequently, extrapolation to other populations or domains of behavior must be done with caution. Indeed, while our results suggest that the effects of public recognition are invariant to some factors such as group size, our estimates appear to be less stable with respect to other factors such as individuals' familiarity with each other.

Third, even within the specific contexts of our experiments, our *quantitative* welfare estimates cannot be immediately applied to public recognition schemes that produce different information structures such as ones that recognize only the top performers. Although standard economic models imply that coarsening the information structure cannot eliminate feelings of shame if such feelings are prevalent in fully-revealing schemes (see Appendix A.1), and although our estimates of structural models can be used to generate predictions about these alternative schemes, limited attention or failures of equilibrium thinking could weaken the predictive power of standard economic models. Our flexible online experimental protocol can be augmented to further study how the effects of public recognition vary with the signal structure.

More generally, we suggest that our online protocol can be fruitfully extended to facilitate further testing and refinement of social signaling models. Empirical tests of social signaling models typically revolve around comparative statics on behavior, although underlying these comparative statics are predictions about individuals' social image payoffs. By providing a direct estimate of social image payoffs, our methodology can thus enable more direct tests of phenomena such as the overjustification effect and motivation crowding (Gneezy and Rustichini, 2000; Bénabou and Tirole, 2006), predictions about the effects of social information on prosocial behavior (Bénabou and Tirole, 2011), or the evolution of stigma and redistributive norms (Alesina and Angeletos, 2005).

With some extension, our approach could also be applied more broadly to study other social influence levers. Although such non-financial policy instruments have become popular tools in governments around the world under the banner of "nudge" (OECD, 2017), most existing studies focus on how these instruments affect behavior, and have little to say about *welfare* (see, e.g., Bernheim and Taubinsky, 2018a, for a review).[37] We view this as a limitation of existing research methods, not a reflection of actual social goals. Indeed, in the case of social influence, an honest assessment of the psychological, political, philosophical, and literary studies of human motivation reveals that people's well-being is intensely sensitive to the experience of shame and pride.

[36]We note that while financial incentives motivate desirable behavior and have little interaction with public recognition in our domains, there are also important cases where financial incentives could crowd out motivation because they dampen the effects of both shame and pride (e.g., Bénabou and Tirole, 2006; Ariely et al., 2009a).

[37]See, e.g., Bénabou and Tirole (2006), Bénabou and Tirole (2011), and Ali and Bénabou (2020) for an example of welfare analysis with social image.

Figures and Tables

Figure 1.1: Illustration of public recognition information

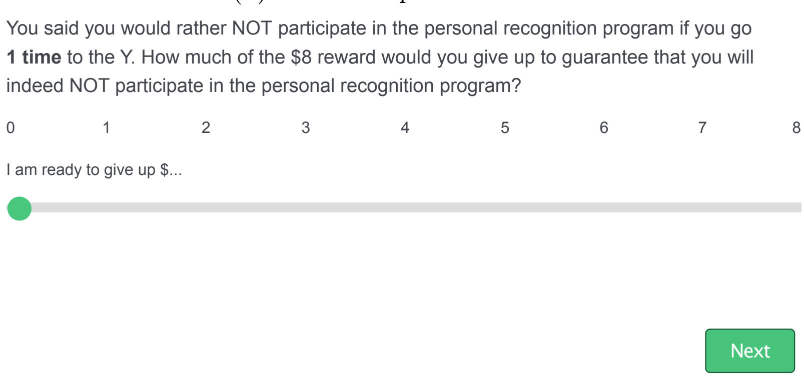
Thank you for joining Grow & Thrive from your friends at YMCA!		
	# of visits	Dollars Raised
1. John Doe	25	$50
2. Mary Adams	24	$48
..		
49. Jack Black	10	$20
..		

Notes: This figure shows an illustration of how individuals' attendance was publicized in the YMCA experiment.

Figure 1.2: An example of WTP for public recognition in the YMCA experiment

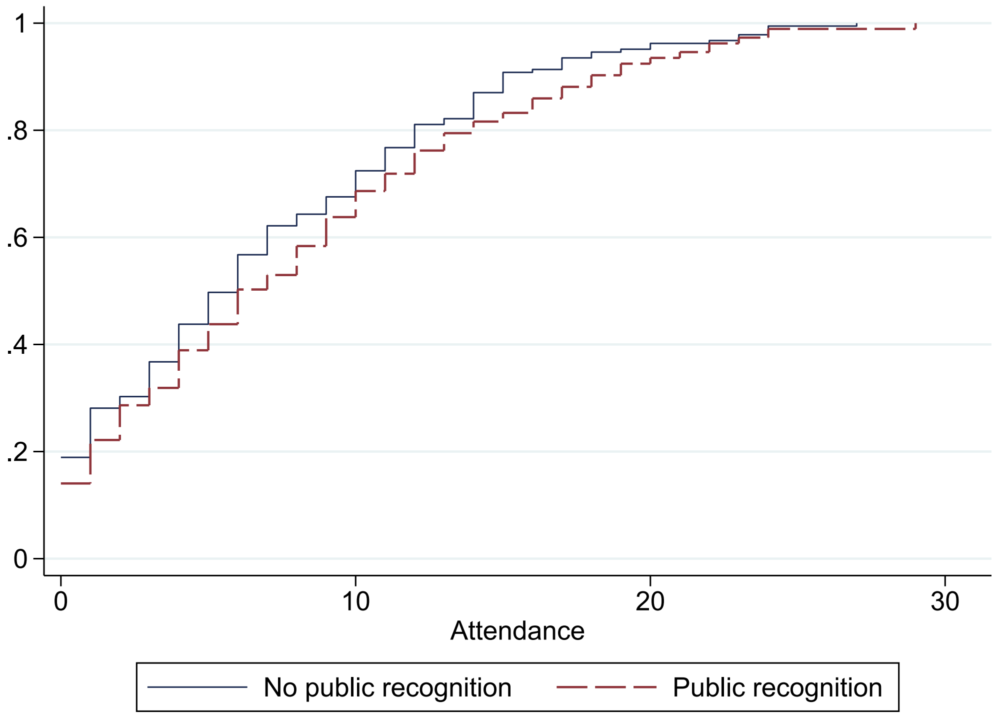
(a) First step of elicitation

(b) Second step of elicitation



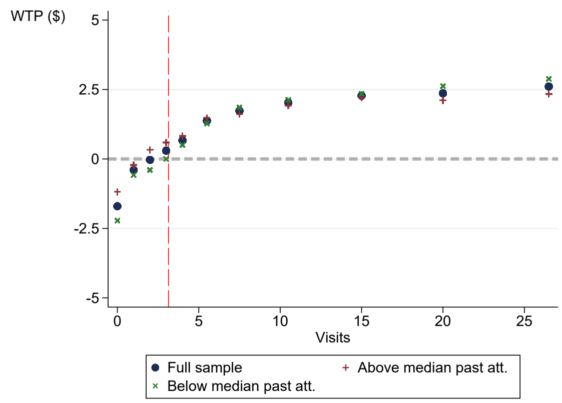
Notes: These figures present screenshots of the procedure for elicitation of WTP for public recognition. The example above shows the elicitation of WTP for attending the YMCA once during Grow & Thrive. The top panel presents the first step of the elicitation, where participants are asked whether they want to be publicly recognized. The bottom panel presents the second step, where participants are asked how much they are willing to pay (from \$0 to \$8) to guarantee that their preference from the first step is implemented. Participants choose the amount by moving the slider bar.

Figure 1.3: Cumulative distributions of attendance during the YMCA experiment, by treatment

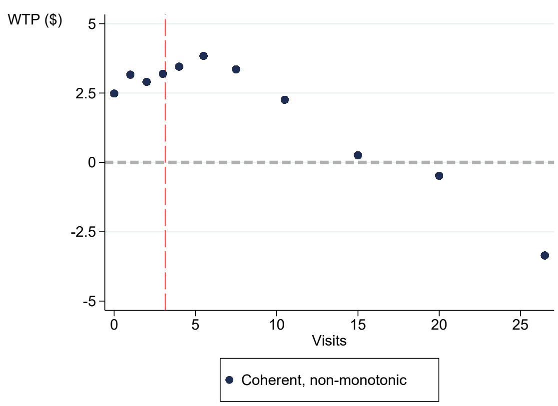


Notes: This figure plots the cumulative distribution functions of attendance during the experiment, by whether participants were in the public recognition group. The analysis excludes 15 participants with "incoherent" preferences for public recognition.

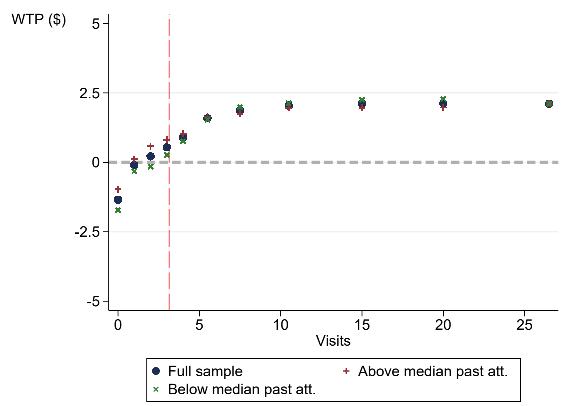
Figure 1.4: WTP for public recognition, by YMCA attendance



(a) Monotonic participants



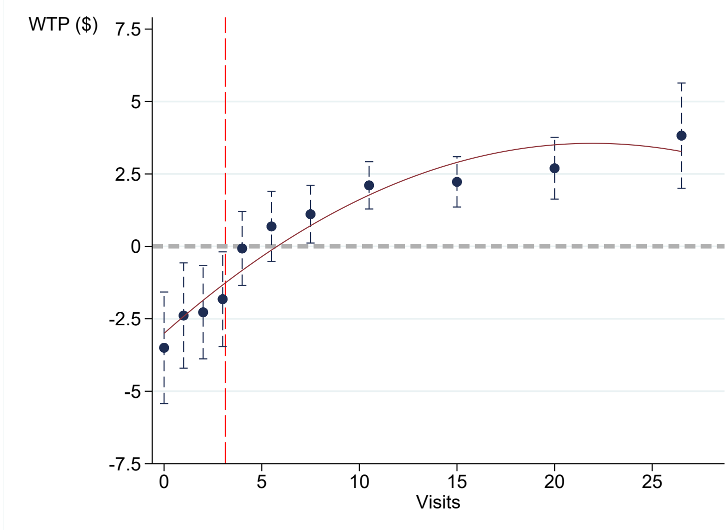
(b) Coherent but non-monotonic participants



(c) Main sample

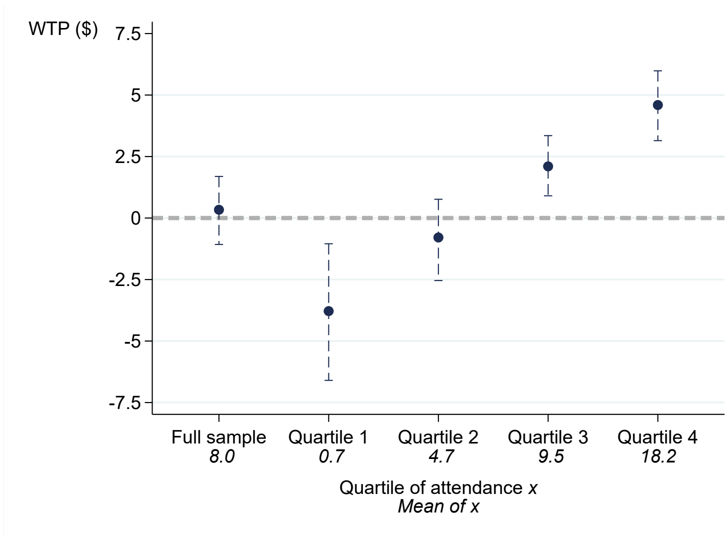
Notes: These figures plot the average WTP for public recognition by each of the eleven intervals of possible future attendance. For intervals including more than one value of visits (e.g., "5 or 6 visits"), the WTP is plotted at the midpoint the interval. Panel (a) reports the average WTP for participants with monotonic preferences for public recognition, as well as for this sample split by median past attendance. Panel (b) reports the average WTP for participants included in the main sample, but with non-monotonic preferences for public recognition. Panel (c) reports the average WTP for the full sample, as well as for this sample split by median past attendance. The average YOTA attendance during Grow & Thrive is indicated by the dashed red line.

Figure 1.5: WTP for public recognition by YMCA attendance, restricting to questions about visits close to participants' expectations



Notes: These figures plot the average WTP for public recognition by each of the eleven intervals of possible future attendance. For intervals including more than one value of visits (e.g., "5 or 6 visits"), the WTP is plotted at the midpoint the interval. The data in these figures is restricted to visits intervals with a midpoint within 4 of a participant's predicted attendance if assigned to the public recognition group. The analysis excludes 15 participants with "incoherent" preferences for public recognition. The average YOTA attendance is indicated by the dashed red line. 95 percent confidence intervals are constructed from standard errors clustered by participant. Quadratic fit curves are plotted in red.

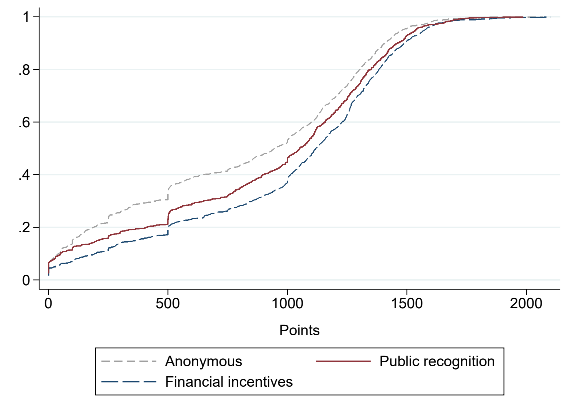
Figure 1.6: The net image payoffs in the YMCA experiment



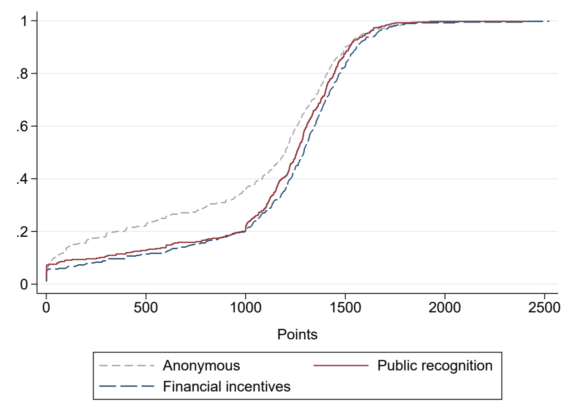
Notes: These figures plot the average realized public recognition payoff of participants assigned public recognition, for both the full sample and each quartile of actual attendance. The average attendance is reported below each subsample label. A participant's payoff is defined as the WTP predicted by the regression in column (4) of Table 1.4, given the participant's realized attendance. The analysis excludes 15 participants with "incoherent" preferences for public recognition. Bootstrapped percentile-based confidence intervals, sampled by participant with 1000 iterations, are displayed.

Figure 1.7: Cumulative distributions of points scored in each of the three rounds of the charitable contribution experiments

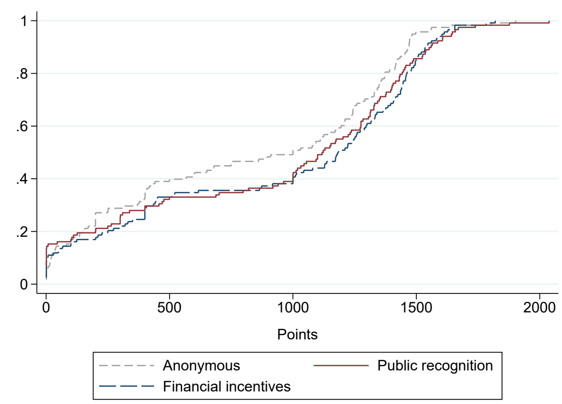
(a) Prolific



(b) Berkeley



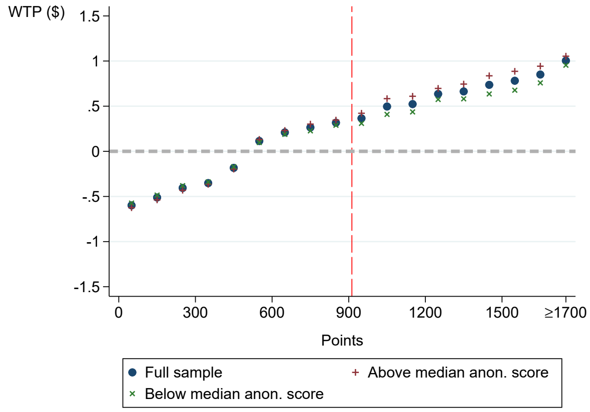
(c) BU



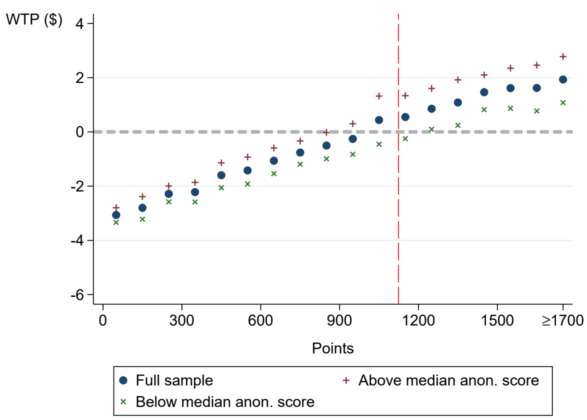
Notes: These figures plot the cumulative distribution functions of points scored in the Anonymous Effort Round, the Anonymous and Paid Effort Round, and the Publicly-Shared Effort Round. Panel (a) presents results for the Prolific sample, panel (b) presents results for the Berkeley sample, and panel (c) presents results for the BU sample. The analysis excludes 40 Prolific participants, 11 Berkeley participants, and 2 BU participants with "incoherent" preferences for public recognition.

Figure 1.8: Willingness to pay for public recognition by effort in the charitable contribution experiments

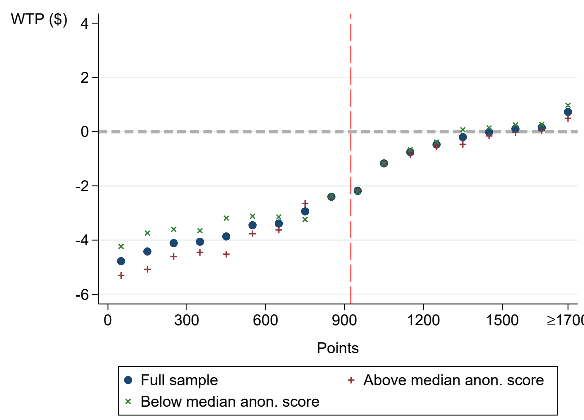
(a) Prolific



(b) Berkeley



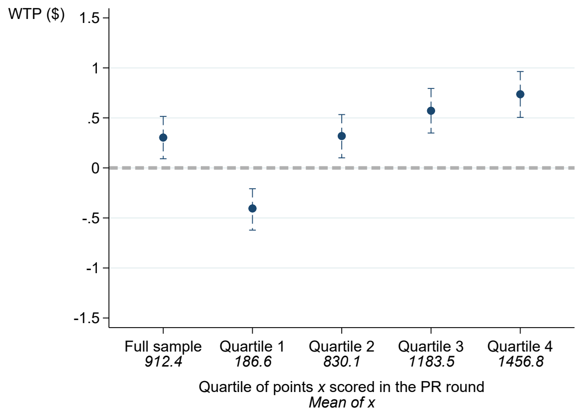
(c) BU



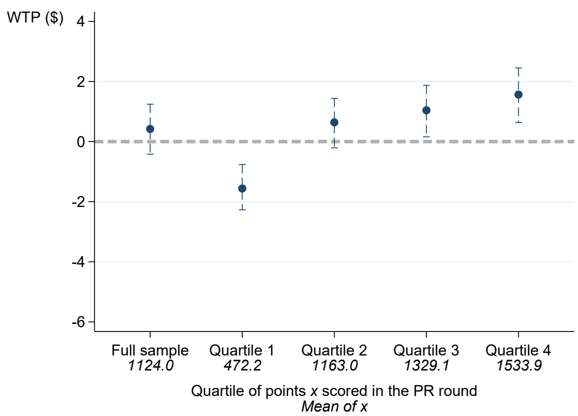
Notes: These figures plot the average WTP for public recognition by each of the 18 possible intervals of points scored. The WTP is plotted at the midpoint of each of the first seventeen intervals and at ≥1700 points for the 1700 or more points interval. Panel (a) presents results for the Prolific sample, panel (b) presents results for the Berkeley sample, and panel (c) presents results for the BU sample. The mean Publicly-Shared Effort Round scores are indicated by dashed red lines. The analysis excludes 40 Prolific participants, 11 Berkeley participants, and 2 BU participants with "incoherent" preferences for public recognition.

Figure 1.9: Image payoffs in the charitable contribution experiments

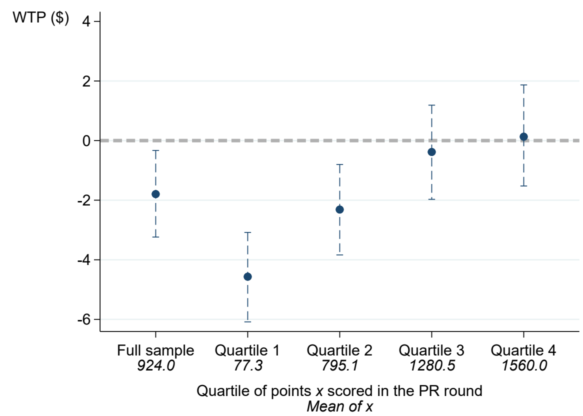
(a) Prolific



(b) Berkeley



(c) BU



Notes: These figures plot the average realized image payoff of participants assigned to public recognition, for both the full sample and each quartile of actual attendance. The average points scored in the public recognition round is reported below each subsample label. Panel (a) presents results for the Prolific sample, panel (b) presents results for the Berkeley sample, and panel (c) presents results for the BU sample. The analysis excludes 40 Prolific participants, 11 Berkeley participants, and 2 BU participants with "incoherent" preferences for public recognition. The average realized image payoff is defined as the average WTP reported across all participants for the points interval corresponding to the participant's score in the public recognition round. Bootstrapped percentile-based confidence intervals, sampled by participants with 1000 iterations, are displayed.

Table 1.1: Balance table for YMCA experiment

	No PR treatment	PR treatment	p-value
Average WTP (over all possible N. of visits)	1.10	1.09	0.98
	(5.13)	(5.03)	
Average monthly past attendance	5.75	5.64	0.86
	(5.64)	(5.67)	
Beliefs about attendance assuming public recognition	13.90	13.41	0.44
	(5.88)	(6.18)	
Beliefs about attendance assuming no public recognition	12.51	11.83	0.28
	(5.94)	(6.09)	
Gender (0=Male; 1=Female)	0.74	0.76	0.63
	(0.44)	(0.43)	
Age	44.24	43.70	0.65
	(11.19)	(11.60)	
N. Subjects	185	185	

Notes: This table reports summary statistics across all coherent participants, by assignment to the public recognition group. Variable "Average WTP (over all possible N. of visits)" is the average participant WTP across all possible intervals of future attendance. Variables "Beliefs about attendance assuming (no) public recognition" report the average forecast of future attendance conditional on (not) being part of the public recognition treatment. The last column reports two-sided p-values to test for balance across our experimental treatment. The analysis excludes 15 participants with "incoherent" preferences for public recognition. Standard deviations are reported in parentheses.

Table 1.2: The impact of public recognition on YMCA attendance

	(1)	(2)	(3)
Public recognition	1.10	1.19***	1.27***
	(0.69)	(0.46)	(0.45)
Avg. past att.		0.88***	0.77***
		(0.04)	(0.05)
Beliefs			0.19***
			(0.05)
Control mean	6.91	6.91	6.91
	(0.47)	(0.47)	(0.47)
N. Subjects	370	370	370

Notes: This table reports regression estimates of the effects of public recognition on attendance during the experiment. "Beliefs" reports the expectations YMCA members had about their attendance assuming that they would be part of the public recognition treatment. The analysis excludes 15 participants with "incoherent" preferences for public recognition. The control mean is the average attendance for participants in the experiment who are not in the public recognition program. Standard errors are clustered at the participant level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.3: WTP for public recognition by YMCA attendance

	(1)	(2)	(3)	(4)
Model	OLS	OLS	Tobit	Tobit
Dependent var.	WTP	WTP	WTP	WTP
N. visits	0.10***	0.36***	0.19***	0.62***
	(0.01)	(0.04)	(0.03)	(0.07)
N. visits sq.		-0.01***		-0.02***
		(0.00)		(0.00)
Constant	0.20	-0.57*	-0.03	-1.35**
	(0.30)	(0.32)	(0.59)	(0.63)
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})$	–	0.069	–	0.068
95% CI	–	[0.064, 0.075]	–	[0.062, 0.074]
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})\times$ SD	–	0.337	–	0.329
95% CI	–	[0.310, 0.364]	–	[0.299, 0.358]
Observations	4070	4070	4070	4070
N. Subjects	370	370	370	370

Notes: This table reports regression estimates from linear and quadratic models of willingness to pay for public recognition by attendance. Measures of the curvature of the estimated reduced-form public recognition function are $-R''_{exp}/R'_{exp}(\bar{a}_{pop})$ and $-R''_{exp}/R'_{exp}(\bar{a}_{pop}) \times SD$, where $\bar{a}_{pop}$ and $SD = 4.86$ are the average attendance and standard deviation of attendance for the general YOTA population, respectively. The analysis excludes 15 participants with "incoherent" preferences for public recognition. Standard errors are clustered at the participant level and reported in parentheses. 95 percent confidence intervals for the curvature statistics are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.4: WTP for public recognition by YMCA attendance, restricting to questions about visits close to participants' expectations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model	OLS	OLS	Tobit	Tobit	OLS	OLS	Tobit	Tobit
Dependent var.	WTP	WTP	WTP	WTP	WTP	WTP	WTP	WTP
N. visits	0.23***	0.56***	0.40***	0.88***	0.21***	0.59***	0.39***	1.03***
	(0.04)	(0.13)	(0.08)	(0.25)	(0.05)	(0.18)	(0.09)	(0.35)
N. visits sq.		-0.01***		-0.02**		-0.01**		-0.02*
		(0.00)		(0.01)		(0.01)		(0.01)
Constant	-1.27*	-2.60***	-2.47**	-4.40***	-0.69	-3.02**	-1.90	-5.71**
	(0.65)	(0.89)	(1.16)	(1.62)	(0.69)	(1.23)	(1.29)	(2.38)
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})$	–	0.057	–	0.053	–	0.051	–	0.048
95% CI	–	[0.039, 0.076]	–	[0.029, 0.078]	–	[0.031, 0.071]	–	[0.024, 0.073]
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})\times$ SD	–	0.279	–	0.260	–	0.247	–	0.236
95% CI	–	[0.188, 0.371]	–	[0.143, 0.377]	–	[0.149, 0.345]	–	[0.114, 0.357]
Restriction	≤ 4	≤ 4	≤ 4	≤ 4	Exact	Exact	Exact	Exact
Observations	923	923	923	923	370	370	370	370
N. Subjects	370	370	370	370	370	370	370	370

Notes: These tables report regression estimates from linear and quadratic models of willingness to pay for public recognition by attendance. Columns (1)-(4) restrict to visits intervals with a midpoint within 4 of a participant's predicted attendance if assigned to the public recognition group. Columns (5)-(8) restrict to intervals that contain the participant's predicted attendance if assigned to the public recognition group. Measures of the curvature of the estimated reduced-form public recognition function are $-R''_{exp}/R'_{exp}(\bar{a}_{pop})$ and $-R''_{exp}/R'_{exp}(\bar{a}_{pop}) \times SD$, where $\bar{a}_{pop}$ and $SD = 4.86$ are the average attendance and standard deviation of attendance for the general YOTA population, respectively. The analysis excludes 15 participants with "incoherent" preferences for public recognition. Standard errors are clustered at the participant level and reported in parentheses. 95 percent confidence intervals for the curvature statistics are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.5: The effect of public recognition and financial incentives on performance in the charitable contribution experiments

	(1)	(2)	(3)	(4)
Model	OLS	OLS	OLS	OLS
Dependent var.	Points	Points	Points	Points
Public recognition	105.01***	134.41***	103.61**	106.70***
	(12.25)	(22.56)	(45.25)	(18.72)
Financial incentives	185.74***	177.76***	118.33***	191.96***
	(12.56)	(22.04)	(39.62)	(18.98)
Group of 300				20.61
				(39.85)
Group of 300 × Public recognition				-3.12
				(28.43)
Group of 300 × Financial incentives				-18.85
				(29.05)
Group of 15				17.70
				(41.13)
Group of 15 × Public recognition				-3.21
				(31.13)
Group of 15 × Financial incentives				-3.27
				(31.90)
Control mean	807.9	989.8	815.9	
	(16.7)	(27.2)	(52.8)	
Round order dummies	Yes	Yes	Yes	Yes
Order dummies F-test	0.180	0.497	0.116	0.178
Sample	Prolific	Berkeley	BU	Prolific
Observations	2904	1152	354	2904
N. Subjects	968	384	118	968

Notes: This table reports regression estimates of the effects of public recognition and financial incentives on points scored. Column (1), (2), and (3) report estimates for the Prolific, Berkeley, and BU samples, respectively. Column (4) includes interactions with group size variables in the Prolific sample, which indicate the approximate number of individuals in the participant's randomly assigned public recognition group. The control mean is the mean points scored in the Anonymous Effort Round. Dummy variables for the order in which the round appeared (first, second, or third) are included, and the p-value from a test of their joint significance is reported. The analysis excludes 40 Prolific participants, 11 Berkeley participants, and 2 BU participants with "incoherent" preferences for public recognition. Standard errors are clustered at the participant level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.6: WTP for public recognition by effort in the charitable contribution experiments

	(1)	(2)	(3)	(4)	(5)	(6)
Model	OLS	OLS	OLS	OLS	OLS	OLS
Dependent var.	WTP	WTP	WTP	WTP	WTP	WTP
Points (00s)	0.093***	0.155***	0.310***	0.379***	0.347***	0.309***
	(0.007)	(0.018)	(0.033)	(0.070)	(0.060)	(0.116)
Points (00s) sq.		-0.004***		-0.004		0.002
		(0.001)		(0.004)		(0.006)
Constant	-0.557***	-0.733***	-3.130***	-3.325***	-5.186***	-5.076***
	(0.113)	(0.121)	(0.400)	(0.420)	(0.791)	(0.810)
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})$	–	0.076	–	0.027	–	-0.013
95% CI	–	[0.047, 0.106]	–	[-0.021, 0.075]	–	[-0.079, 0.052]
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}}) \times \mathrm{SD}$	–	0.245	–	0.114	–	-0.085
95% CI	–	[0.186, 0.303]	–	[-0.047, 0.275]	–	[-0.559, 0.388]
Sample	Prolific	Prolific	Berkeley	Berkeley	BU	BU
Observations	16456	16456	6528	6528	2006	2006
N. Subjects	968	968	384	384	118	118

Notes: This table reports regression estimates from linear and quadratic models of willingness to pay for public recognition by the level of publicized effort. Effort is measured in 100s of points scored. The regressions exclude the ≥1700 points interval. Measures of the curvature of the estimated reduced-form public recognition function are $-R''_{exp}/R'_{exp}(\bar{a}_{pop})$ and $-R''_{exp}/R'_{exp}(\bar{a}_{pop}) \times SD$, where $\bar{a}_{pop}$ and $SD = 4.86$ are the average and standard deviation of points scored in the anonymous round (in units of hundreds of points), respectively. The analysis excludes 40 Prolific participants, 11 Berkeley participants, and 2 BU participants with "incoherent" preferences for public recognition. Standard errors are clustered at the participant level and reported in parentheses. 95 percent confidence intervals for the curvature statistics are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.7: Structural estimates and tests of consistency

(a) Action-signaling model parameter estimates

Sample	$\hat{\gamma}_1^a$	$\hat{\gamma}_2^a$	$\hat{\rho}^a$	$\hat{c}$
YMCA	0.64	−0.020	1.85	0.46
	[0.37,0.92]	[−0.038,−0.001]	[0.88,2.53]	[0.19,1.63]
Prolific	0.12	−0.004	0.58	0.08
	[0.09,0.14]	[−0.005,−0.002]	[0.40,0.80]	[0.06,0.11]
Berkeley	0.30	−0.004	0.87	0.21
	[0.22,0.38]	[−0.011,0.003]	[0.63,1.15]	[0.14,0.33]
BU	0.38	0.002	1.61	0.34
	[0.19,0.53]	[−0.009,0.013]	[1.14,2.33]	[0.16,1.47]

(b) Characteristics-signaling model parameter estimates

Sample	$\hat{\gamma}_1^\theta$	$\hat{\gamma}_2^\theta$	$\hat{\rho}^\theta$	$\hat{c}$
YMCA	1.28	−0.079	1.40	0.46
	[0.32,2.26]	[−0.254,−0.001]	[0.36,2.22]	[0.19,1.63]
Prolific	1.30	−0.458	0.49	0.08
	[0.98,1.65]	[−0.765,−0.241]	[0.25,0.76]	[0.06,0.11]
Berkeley	1.35	−0.082	0.85	0.21
	[0.90,1.81]	[−0.330,0.046]	[0.55,1.17]	[0.14,0.33]
BU	1.13	0.021	1.68	0.34
	[0.11,2.40]	[−0.101,0.252]	[1.16,2.40]	[0.16,1.47]

(c) Predicted and actual effects of financial incentives (on attendance or points (00s))

Sample	Model prediction	Actual	Pred.− Act.
YMCA	2.16	1.77†	0.39
	[0.51,4.70]	[1.29,2.22]	[−1.31,2.97]
Prolific	2.41	1.82	0.60
	[1.81,3.12]	[1.56,2.07]	[0.07,1.25]
Berkeley	2.33	1.78	0.55
	[1.49,3.53]	[1.35,2.24]	[−0.18,1.63]
BU	1.48	1.18	0.29
	[0.21,2.89]	[0.47,1.94]	[−0.89,1.56]

†: Based on individuals' forecasted rather than realized behavior.

Notes: These tables report parameter estimates of the action-signaling and characteristics-signaling models described in Section 1.8.1, equations (1.5) and (1.6). For panel (c), the financial incentive is \$1/attendance for the YMCA sample, 2 cents/10 points for the Prolific sample, and 5 cents/10 points for the Berkeley and BU samples. The analysis excludes participants with "incoherent" preferences for public recognition (15 in YMCA participants, 40 Prolific participants, 11 Berkeley participants, and 2 BU participants). Bootstrapped percentile-based confidence intervals from 1000 replications, clustered at the participant level, are reported in brackets.

Table 1.8: Welfare estimates of public recognition in the charitable contribution experiments

Row	Sample	(1) Image payoffs	(2) Change in points scored
1.	Prolific	0.30	13.00%
2.	Berkeley	0.42	13.58%
3.	BU	−1.80	12.70%

Notes: This table reports the the average realized image payoff of participants assigned to public recognition. The analysis excludes 40 Prolific participants, 11 Berkeley participants, and 2 BU participants with "incoherent" preferences for public recognition. The average realized image payoff is defined as the average WTP reported across all participants for the points interval corresponding to the participant's score in the public recognition round. The estimates in Column (1) match the "full sample" estimates reported in Figure 1.9. Column (2) reports the change in points scored from public recognition as a percentage of the average points scored in the anonymous round, which are 808, 990, and 816 for the Prolific, Berkeley, and BU samples, respectively.

Table 1.9: Welfare estimates of scaling up public recognition at the YMCA

(a) Action-signaling model

		Parameter estimates			(1) Image	(2) Change
Row	Scenario	γ_1^a	γ_2^a	ρ^a	payoffs	in attendance
1.	Baseline (YMCA)	0.64	−0.020	1.85	−3.41	55.77%
2.	ρ from Prolific sample	0.64	−0.020	0.58	0.70	39.31%
3.	ρ from Berkeley sample	0.64	−0.020	0.87	−0.04	42.73%
4.	ρ from BU sample	0.64	−0.020	1.61	−2.46	52.41%
5.	Curvature from Prolific sample	0.64	−0.022	1.85	−3.51	57.06%
6.	Curvature from Berkeley sample	0.64	−0.010	1.85	−2.92	49.53%
7.	Curvature from BU sample	0.64	0.005	1.85	−2.25	41.93%
8.	ρ & curv. from Prolific sample	0.64	−0.022	0.58	0.66	38.81%
9.	ρ & curv. from Berkeley sample	0.64	−0.010	0.87	0.16	43.57%
10.	ρ & curv. from BU sample	0.64	0.005	1.61	−1.61	42.59%

(b) Characteristics-signaling model

		Parameter estimates			(1) Image	(2) Change
Row	Scenario	γ_1^θ	γ_2^θ	ρ^θ	payoffs	in attendance
1.	Baseline (YMCA)	1.28	−0.079	1.40	−1.18	47.55%
2.	ρ from Prolific sample	1.28	−0.079	0.49	0.51	40.23%
3.	ρ from Berkeley sample	1.28	−0.079	0.85	−0.12	43.12%
4.	ρ from BU sample	1.28	−0.079	1.68	−1.74	49.75%
5.	Curvature from Prolific sample	1.28	−0.077	1.40	−1.17	47.36%
6.	Curvature from Berkeley sample	1.28	−0.060	1.40	−1.07	45.72%
7.	Curvature from BU sample	1.28	0.022	1.40	−0.63	39.17%
8.	ρ & curv. from Prolific sample	1.28	−0.077	0.49	0.52	40.24%
9.	ρ & curv. from Berkeley sample	1.28	−0.060	0.85	−0.02	42.45%
10.	ρ & curv. from BU sample	1.28	0.022	1.68	−1.13	38.61%

Notes: These tables report welfare estimates based on the structural estimates of the action-signaling and characteristics-signaling models described in Section 1.8.1.

Chapter 2

Rules of Thumb and Attention Elasticities: Evidence from Under- and Overreaction to Taxes[1]

2.1 Introduction

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

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

In the first two of these example cases, the consumers *underreact* to sales taxes—they behave as if the taxes are smaller than they are. In the last case, the consumers *overreact*. When purchasing expensive electronics or an automobile, however, consumers may choose

[1] Coauthored with Dmitry Taubinsky, UC Berkeley. This chapter includes material forthcoming for publication from "Rules of Thumb and Attention Elasticities: Evidence from Under- and Overreaction to Taxes." *The Review of Economics and Statistics*. The experiment was approved by the Dartmouth College Committee for the Protection of Human Subjects (CPHS), # STUDY00029784.

to exert more cognitive effort to compute the actual price that they would end up paying, thereby reducing their propensity to both over- and underreact.

Prior literature on sales tax salience has convincingly shown that *on average*, individuals underreact to opaque sales taxes.[2] However, existing results about averages do not preclude that some individuals overreact, and provide little evidence about the degree to which individuals' misreaction to sales taxes is due to deliberate, and plausibly elastic, use of cognitive shortcuts. In this paper, we provide a series of tests, grounded in models of costly attention, that fill this gap. In doing so, we develop a methodology for testing models of costly attention that could be applied to other domains with opaque attributes that are imperfectly processed by consumers—energy prices (e.g., Allcott and Kessler, 2019), shipping and handling charges (Hossain and Morgan, 2006), various features of health insurance contracts (e.g., Handel and Kolstad, 2015; Bhargava et al., 2017), less significant digits (Lacetera et al., 2012), shrouded financial fees (Heidhues et al., 2017), and add-on charges (Gabaix and Laibson, 2006).

We begin our investigation in Section 2.2 by formalizing the economic environment and several types of costly attention models. Consumers must decide whether or not to buy a good or service that has both a transparent posted price and an opaque price. Consumers have a prior perception of the post-tax price that they can access costlessly, and which can vary between consumers, as in our example. We consider several formulations of the cognitive costs of updating: the Shannon cost function used in rational inattention models,[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. In other words, $\theta = 0.4$ means that consumers are 40 percent as responsive to taxes as fully attentive consumers would be.

The underlying costly attention models discipline the reduced-form valuation weights in 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

[2]See Chetty et al. (2009); Goldin and Homonoff (2013); Feldman and Ruffle (2015); Taubinsky and Rees-Jones (2018); Feldman et al. (2018); Bradley and Feldman (2020); Kroft et al. (2020)

[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.

sales tax rate increases). Moreover, the higher is the valuation weight at moderate stakes, the smaller is the degree by which it increases when stakes increase. In particular, the valuation weights should decrease for consumers who overreact and increase for consumers who underreact.

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

Our experiment features 1534 demographically diverse consumers from the forty-five U.S. states with positive sales taxes. The experiment utilizes an online shopping environment with nine different non-tax-exempt household products, such as cleaning supplies. Each consumer encounters three of the nine products in three different types of "stores" at random posted prices. The three different types of stores feature either 1) no sales taxes, 2) standard sales taxes identical to those in the consumer's city of residence, or 3) high sales taxes that are triple those in the consumer's city of residence. Each consumer thus encounters 3×3 product by store pairs, with each associated to a set of random prices. Decisions in the experiment are incentive-compatible: study participants receive a \$16 budget to potentially buy one of the randomly chosen products in one of the randomly chosen stores, and purchased products are shipped to their homes.

We begin our empirical analysis with a very simple test in Section 2.4: we estimate average underreaction to taxes of varying size, exploiting both the exogenous variation in prices and the exogenous variation in tax rates. Although this test has been implemented in several prior studies, our analysis is significantly better-powered and is unique in exploiting variation in both prices and tax rate sizes. We find striking evidence that misreaction depends on stakes. The average valuation weight is 0.23 for the smallest price at standard tax rates—meaning that at these stakes consumers are only 23 percent as responsive to taxes as fully attentive consumers would be. However, the average valuation weight is 0.79 for the largest price at triple tax rates—meaning that at these stakes consumers are 79 percent as responsive to taxes as fully attentive consumers would be. The average increases monotonically in the absolute size of the tax, and in a manner that is invariant to whether the absolute size of the tax is high because the tax rate is high or because the price is high.

In Section 2.5 we begin testing our novel predictions about individual differences in responses to stakes. A key challenge for tests of individual differences is that individual-level estimates of the valuation weights will necessarily involve significant measurement error. Thus, making inferences about individual-level differences merely from a distribution of individual-level point estimates would yield confounded conclusions if one took the point-estimates at face-value. This challenge is not unique to our setting, and poses problems for most within-subject experiments seeking to quantify individual differences (including ones where authors choose to report individual-level point estimates nonetheless). To overcome this challenge and generate simple reduced-form tests of our individual-level predictions, we leverage the multiple decisions feature of our design to examine whether people who seem to be most sensitive to taxes on one product react differently to taxes on the *other two* products.

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

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

Having established significant and persistent individual differences in valuation weights, as well as heterogeneous attention responses to higher stakes in linewith costly attention models, we ask two key questions in Section 2.6. First, are the individual differences large enough that some consumers overreact to standard taxes? If so, can we show that some consumers decrease their valuation weights when the stakes increase?

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

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

Our paper contributes to several literatures. First, our paper contributes to a recent literature that experimentally tests models of costly attention.[4] With the exception of Bartos et al. (2016), these papers utilize abstract information acquisition and problem-solving tasks to provide comprehensive tests of core assumptions of the models. Our paper complements this literature by focusing on a concrete empirical setting, and asking whether the “mistakes”

[4]See, e.g., Gabaix et al. (2006); Bartos et al. (2016); Martin (2016); Dean and Neligh (2019); Ambuehl et al. (2018); Caplin et al. (2020); Carvalho and Silverman (2019).

identified by reduced-form empirical work in that domain fit the patterns of costly attention models.

By focusing on the concrete setting of opaque sales taxes, and opaque prices more broadly, our paper also deepens the empirical work in those settings. Empirical work in these settings has not tested predictions about individual differences in attentional responses to stakes. Gabaix (2019) provides suggestive evidence from cross-study comparisons that average underreaction to opaque prices decreases with stakes. This paper continues this line of inquiry by providing rigorous experimental evidence of this from a series of more complete and theoretically-demanding tests.

While focusing mostly on normative implications of tax salience and using an experimental design importantly different from ours in key ways, Taubinsky and Rees-Jones (2018, henceforth TRJ) are closest to our work in that they also find that average underreaction decreases when tax rates are saliently increased. However, TRJ are not well-powered to estimatehow average underreaction varies by pre-tax price, and thus cannot rule out that their results are driven by other possibilities such as consumers overreacting to a surprising change that violates their shopping "norms" (Bordalo et al., 2020a). More importantly, the lack of within-consumer variation in tax rates in the TRJ data makes testing core predictions about individual differences in how attention responds to stakes infeasible. Out of the five key predictions that we test in this paper, the experimental design employed by TRJ allows only a partial test of the first prediction, and is infeasible for testing the other four mopre novel predictions.Unlike TRJ, we can reject models where no one overreacts to the taxes, which rejects the important class of costly attention models with common priors, as well as the bounded-rationality model in Chetty et al. (2007). Unlike TRJ, we can also reject theories in which individuals' priors about the opaque price are not persistent across stake size, which is crucial for inferring that the decrease in average underreaction after increases in tax rates is indeed due to costly mental adjustment from a heuristic rule-of-thumb, rather than people simply relying on different rule-of-thumb strategies. In Appendix B.6 we provide a detailed comparison to TRJ as well as other studies of tax salience.

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

2.2 Theoretical framework for hypothesis development

2.2.1 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 misprocessed. For example, a salient announcement that sales taxes will be tripled is likely to be fully noted by consumers, and corresponds to an increase in 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. For example, consumers who are very confident in their assessment will not bother to exert mental costs. We detail the link between mental effort and improvements to the prior perception in the subsections that follow.

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

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

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

We view the static costly attention models we work with to be "as if" models of this effortful thinking. Process-based models, such as sequential sampling (e.g., Fudenberg et al., 2018), could provide more complete accounts of how the allocation of costly attention to a decision improves accuracy.[5]

2.2.2 Simple example with binary attention strategies

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

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

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

$$\begin{aligned}\theta &= \frac{1}{\sigma t} \max\left(\hat{p}_o, \sigma t - \lambda/r\right) \\ &= \max\left(r + (1-r)\frac{\hat{t}}{t}, 1 - \frac{\lambda}{\sigma t r}\right) < 1 \end{aligned} \tag{2.1}$$

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

$$\theta = \min\left(r + (1-r)\frac{\hat{t}}{t}, 1 + \frac{\lambda}{\sigma t r}\right) > 1. \tag{2.2}$$

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

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

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 actually weakly increasing in σ.

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 and σ_2.

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

While the simple example above involved a binary choice of either paying attention or not, the key results and predictions generalize to models with more continuous choice of attention. In Appendix B.1.1 we consider the Shannon model (Sims, 2003; Matejka and McKay, 2015; Caplin et al., 2019) of attention, in which individuals are free to choose any signal structure they wish, and pay attention costs that are linear in the expected reduction of entropy. In Appendix B.1.2 we consider the Gabaix (2014) model, where individuals pay an attention cost that is a monotonic function of the the difference between the posterior mean and the prior mean. We show that both models have a reduced-form representation where the consumer buys the product if its value v to the consumer exceeds $p_s + \theta p_o$, where θ depends on p_o but not on p_s, and where θ is stochastic in the Shannon model but deterministic in the Gabaix (2014). We show that all the predictions about how the distribution of θ varies with stakes replicate in these two models of continuous attention.[6]

Finally, in Appendix B.2, we consider priors given by $\hat{t} + \varepsilon$, where $E[\varepsilon] = 0$ and $\hat{t}$ varies, and show that our main predictions hold under these more general assumptions as well.

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

2.2.4 Empirical tests of costly attention models

Following the intuition provided in the special case described in Section 2.2.2, and the more general results in the Appendices, our theoretical results provide five empirical tests. For concreteness, we focus on the case in which $E[\theta] < 1$, as our empirical application studies sales taxes, for which previous work has established underreaction. Consistent with our experiment, we consider a "standard stakes regime" ("standard" value of 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.*

Although mathematically straightforward, Prediction (4) is particularly demanding, including relative to Prediction (3). 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."

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

The predictions do not collectively hold for other mechanisms that generate misreaction to opaque prices. The predictions draw a sharp distinction between attention costs and other plausible mechanisms such as complete unawareness of the opaque price (e.g., Gabaix and Laibson, 2006), incorrect beliefs generated by systematic mislearning (e.g., Hanna et al., 2014) or misleading marketing (e.g., Anagol et al., 2017), forgetting (e.g., Bordalo et al., 2020a), or a systemic lack of financial literacy that prevents consumers from reaching the right answer (e.g., Hastings et al., 2013). We elaborate on other theories ruled out by our predictions in Appendix B.5, including attention models in which attention is exogenous to stakes but responds to non-pecuniary stimuli (see DellaVigna, 2009, for a summary), choice-set dependence models (Bordalo et al., 2013; Koszegi and Szeidl, 2013; Bushong et al., 2021), 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), attention models with homogeneous prior perceptions, and other theories of mental effort such as those proposed by Kahneman (2003) and Ariely et al. (2009b).

2.3 Experimental design and sample

Summary. Each consumer was randomly assigned three of nine household products utilized in the study and made purchase decisions for these three products in three different stores (nine total decision screens). Each store corresponded to a different sales tax rate. In store A, consumers made shopping decisions with a zero sales tax rate (no-tax store). In store B, consumers made shopping decisions with a standard tax rate identical to their city of residence (standard tax environment). In store C, consumers made shopping decisions with a sales tax rate equal to triple their standard tax rate (triple tax environment). The order of these nine sets of shopping decisions was randomized within-subject. Complete details of the experimental protocol are in Appendix B.19.

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

We asked ClearVoice to only recruit panel members from states with a positive sales tax. This excluded panel members from Alaska, Montana, Delaware, New Hampshire, and Oregon. The remaining forty-five states are all represented in our final sample. Prior to learning the details of the experiment, consumers were asked to report their state, county, and city of residence.[7] 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.[8]

Shopping decisions and environment. For each purchase decision, consumers first encountered a screen informing them of which store they were entering and for which product they were shopping. Consumers then clicked through to the next screen, which contained a product description and a picture, identical to how the product is presented on Amazon.com.

[7]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.

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

On this same screen consumers also saw a price list containing ten prices. These prices were chosen such that the minimum for all products was \$4.00, and then increased by a multiplicative factor of 15% up to \$14.07.[9] At each price, consumers were asked whether or not they would be willing to purchase the product. It was explained that the price shown excluded any applicable sales taxes.

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

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 B.19 contains screenshots of the instructions, as well as a list of the nine products, their Amazon.com prices, and their Amazon.com product descriptions.

Decisions in the experiment were incentive-compatible. All study participants who passed the necessary comprehension questions (described below) had a one-third chance of being selected to receive a \$16 budget.[10] 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 that they did not want to purchase at that randomly chosen price, then they would keep their \$16 budget and would not receive the product. If consumers indicated they would like to purchase at the randomly generated price, then the product was sold to the consumer at that salient price p_s—meaning that the consumer kept $16 - p_s(1 + \tau)$ dollars, where τ is the experimentally induced tax rate. The product was then shipped to the person by ClearVoice.

[9] 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.

[10] 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.

Participants received a full explanation of the payout scheme, including that each question, product, and price was equally likely to be chosen. Additionally, we explicitly informed participants that "it is in your best interest to answer each question honestly." Appendix B.19.1 contains a screenshot of the instructions shown to participants.

Ensuring comprehension. To ensure that study participants understood the environment and experimental tax rate, we had them answer six multiple-choice questions after showing them the instructions. Three of these questions concerned the payout, asking participants to identify their shopping budget, how many decisions will be randomly chosen to implement, and the prices at which they would be asked about purchasing the product (see Appendix B.19.2 for a list of these questions). The final three questions asked participants to identify the sales tax rate they would face for an item purchased in store A, store B, and store C—with the possible answers being "no sales tax," "standard sales tax in city of residence," and "triple the sales tax in city of residence." If participants answered a question incorrectly or left it blank, they were prompted to select the correct answer before they could begin. When answering these questions, participants could access the instructions which described the tax environments, provided a visual of the price list, and explained the payout structure.

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

Additional questions. After completing the purchase decisions and additional comprehension checks, participants received a short set of questions eliciting demographic information including household income, marital status, and political beliefs. Appendix B.19.2 contains a list of these questions. 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.

Sample. 1845 consumers completed the experiment. For our primary analyses, we exclude 255 respondents who incorrectly answered one or more of the comprehension questions and an additional 47 respondents who had monotonicity violations within a price list. Our main results in Section 2.4 do not require excluding participants with these monotonicity violations, but our analyses in Sections 2.5 and 2.6 require monotonic preferences to identify a willingness to pay for each product. In Appendix B.14 we replicate our analysis including those who failed our comprehension checks. We exclude nine additional participants with missing or zero sales tax rates in their city of residence. Our final sample includes 1534 respondents.

Experimental recruitment was targeted to generate a final sample approximating the income, age, and gender distribution of the U.S. adult population. Our sample has a median income of \$49000, an average income of \$60837, and an interquartile range of \$25000-\$80000. Our sample also has a median and mean age of 46 and an interquartile range of 32-59; all participants in the final sample are over the age of 18, and all but 56 participants are over the age of 21. The mean total sales tax rate charged in participants' city of residence is 7.24% (median 7.00%), with a standard deviation of 1.26%. For 90 percent of participants, the sales tax rate lies between 5.50% and 9.50%. The distribution of these basic demographics is broadly similar to the U.S. population, although selection on other unmeasured characteristics cannot be ruled out.

2.4 The average impact of stakes on inattention

2.4.1 Descriptive summary of behavior

Figure A.3.1 provides a summary of consumer behavior in our study. We begin with panel (a), where we summarize average demand in each tax environment. To construct average demand, we begin with product-specific demand curves $D_{jk}(p)$ where j indexes products and k indexes the store type—A, B, or C. We then construct average demand curves $D_{avg,k}(p) := \frac{1}{9}\sum_j D_{jk}(p)$. Panel (a) shows that consumers do react to sales taxes, as their willingness to buy at a given posted price is decreasing in the size of the sales tax.

However, panels (b) and (c) show that consumers on average underreact to taxes. In these panels, we construct the demand curves that would be expected if consumers reacted to the taxes fully. Since we only observe purchase decisions at finitely many prices, we construct the counterfactual Sidemand through linear interpolation, as described in Appendix B.7. Panel (b) reports the results for the standard tax environment, and panel (c) for the triple tax environment.

Comparing counterfactual to observed demand in the same store, we see evidence of underreaction. The underreaction is particularly sizeable at low posted prices.[11]

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

$$\ln v_{ijk} \geq \ln p + \theta_{ijk}\ln(1 + \tau_{ik}) \tag{2.3}$$

[11] 10.10 percent of participants never choose to buy a product in store C, compared to 4.89 percent and 5.87 percent of participants in stores A and B, respectively.

We then utilize condition (2.3) to estimate the average revealed valuation weights by estimating the following heteroskedastic probit model:

$$1 - Pr(buy_{ijk}|p) = \Phi\left(\frac{\alpha_j + \ln(p) + \bar{\theta}_B \ln(1+\tau_{ik}) \cdot I(k=B) + \bar{\theta}_C \ln(1+\tau_{ik}) \cdot I(k=C)}{\sigma_j}\right) \tag{2.4}$$

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.

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 B.8 we verify that this approximation error is negligible.

2.4.3 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.

Figure 2.2a plots $E[\theta|p \leq p^\dagger]$ against a price cutoff $p^\dagger$, such that all prices less than or equal to the cutoff value are included in calculating that average valuation weight. We estimate $E[\theta]$ at different posted prices using the empirical model in equation (2.4), 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 2.2a are reported in Appendix B.9.

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 2.2a, the average valuation weight is significantly higher in the triple tax condition than in the standard tax at each cutoff. At the \$4.60 cutoff, the difference in $E[\theta]$ between the triple tax and single tax environments is 0.18 (95% CI [0.09, 0.27]). This difference peaks at 0.38 (95% CI [0.26, 0.51]) at a price cutoff of \$8.05. When pooling over all prices, this difference is 0.31 (95% CI [0.20, 0.42]). These results are consistent with Prediction 1.

Third, the average valuation weights are increasing in the salient posted price p. In the standard tax environment, $E[\theta]$ more than doubles as we move from a cutoff of \$4.60 to pooling all prices: it increases from 0.23 to 0.48 (95% CI for difference [0.14, 0.37]). Similarly, in the triple tax environment, $E[\theta]$ approximately doubles as well, increasing from 0.40 to

0.79 (95% CI for difference [0.32, 0.44]) when moving from a price cutoff of \$4.60 to pooling over all prices. These results provide further evidence consistent with Prediction 1.

One potential concern in examining how the valuation weights vary by price is that the set of consumers on the margin at each price are mechanically different: the higher is the price, the higher is the product valuation of these marginal consumers. If valuation for the product is correlated with attention, this would confound our results about how average valuation weights covary with price. In Appendix B.10, we provide evidence against this concern, showing that consumers who are identified as having higher valuations for products in our experiments are not more attentive to taxes. Of course, the tax rate assignment is exogenous to these differences and is not subject to the same concern.

On the other hand, a concern with examining how valuation weights change in response to an increase in tax rates is that consumers see the triple tax store as a highly unusual environment, which affects their purchase decision beyond the pecuniary channel. Consumers might be significantly more responsive to higher tax rates simply because the increase triggers tax aversion, because the surprising and unusual environment simply draws more attention to itself (Bordalo et al., 2020a), or because of experimenter demand effects in a within-subject design. 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 behavior is say particularly responsive to a tripling of the tax rate because of experimenter demand effects, then we would expect to see that a tripling of the tax rate generates much larger changes in behavior than does an increase in price that leads to the same absolute change int the sales tax.

To do this, we divide the different prices in each store into five disjoint sets, for a total of $5 \times 2 = 10$ sets.[12] For each pair, we estimate the average valuation weight using an extension of model (2.4) with a separate $\bar{\theta}$ parameter for each pair. Specifically, we partition the prices into sets P_n, $n = 1, \ldots, 5$ and estimate $1 - Pr(buy_{ijk}|p)$ as:

$$\Phi\left(\frac{\alpha_j + \sum_{n=1}^{5}\left[\ln(p) + \bar{\theta}_{B,n}\ln(1+\tau_{ik})\cdot I(k=B) + \bar{\theta}_{C,n}\ln(1+\tau_{ik})\cdot I(k=C)\right] I(p \in P_n)}{\sigma_j}\right) \tag{2.5}$$

with σ_1 normalized to 1. We plot the average $\bar{\theta}_{k,n}$ against the average tax owed in the corresponding set: $E\left[\bar{p}_n \tau_{ik}|k\right]$, where $\bar{p}$ is the average price in set n.

Figure 2.2b presents the results. 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 experimenter demand effects generating an exaggerated response to a tripling of the taxes.

2.4.4 Robustness and correlates of misreaction

In Appendix B.14, we replicate our analyses including the 14% of our respondents who were not able to correctly answer the comprehension questions about the tax rates charged in stores A, B and C. Re-inclusion of these participants increases the estimate of average underreaction, but does not change any of the comparative statics. Second, in Appendix

[12] Since the MPLs for store C only contained nine prices, the final set contains only the highest price.

B.16, we analyze whether purchase decisions could be influenced by the order in which the nine purchase decisions are presented to consumers, a potential concern with our within-subject experimental design. We find no evidence of order effects.

In Appendix B.13, we analyze two potential mechanisms for misreaction: (i) inaccurate beliefs about local tax rates, and (ii) inability of participants to compute the sales tax they would need to pay for an item. Utilizing survey questions eliciting participants' knowledge of their local tax rate and their computational ability, we replicate our main results restricting to participants with nearly-accurate beliefs and high computational ability. In both cases, we find strong evidence for Prediction 1, and the estimates are also of very similar magnitude to the full sample results. The results suggest inaccurate beliefs and poor computational ability cannot explain the observed misreaction.

Appendix B.11 shows that we do not find variation in the average underreaction by income, education, or political party affiliation, although these correlations are not well-powered. In that appendix, we also analyze how average valuation weights vary by local tax rates, and find some evidence that participants in high sales tax locations have lower revealed valuation weights than those from low sales tax location, although this is likely a confounded test of costly attention models because local tax variation could be related to a number of differences in geography, including consumers' views and preferences about tax rates.

2.5 Reduced-form results on individual differences in attention

In Section 2.4 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.

2.5.1 Testing Predictions 2 and 3

Approach

The main idea of our reduced-form tests is to identify consumers who seem to be more sensitive to taxes on one product, and to examine their sensitivity to standard and triple taxes on *the other two* products. To construct proxies for sensitivity to taxes, we first construct estimates $\hat{\theta}_{ijk}$ of θ_{ijk} for each consumer i, product j, and store $k \in \{B, C\}$. To do so, we first approximate the maximum pre-tax price p^*_{ijk} at which consumer i is willing to buy product j in store $k \in \{A, B, C\}$ by $\ln p^*_{ijk} = 0.5 \left(\ln p^0_{ijk} + \ln p^1_{ijk}\right)$, 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), we use their non-incentivized answers about the maximum price at which they would be willing to buy the given product, and to reduce the impact of outliers, we set p^*_{ijk} to the the median of the self-reports for product j in store k. Using the buying condition in equation (2.3), we construct the estimates $\hat{\theta}_{ijk}$ for $k \in \{B, C\}$ as

$$\hat{\theta}_{ijk} = \frac{\ln(p^*_{ijA}) - \ln(p^*_{ijk})}{\ln(1 + \tau_{ik})} \tag{2.6}$$

However, we cannot directly use the $\hat{\theta}_{ijk}$ estimates 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.[13] 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.[14]

We instead use the $\hat{\theta}_{ijk}$ estimates to create proxies for high versus low valuation weight consumers. We use *one* product to divide consumers into two groups. The *low* group consists of those with low values of $\hat{\theta}_{ijk}$ and the *high* group consists of those with high values of $\hat{\theta}_{ijk}$. We then estimate our empirical model in (2.4) *on the other two products* to estimate average valuation weights for the low and high groups.

Concretely, the procedure is as follows. We index each of the three products for each person by $j \in \{1, 2, 3\}$. First, we start with $j = 1$ and we split the sample into two groups: those with $\hat{\theta}_{i1B}$ in the top 25% of the population and those with $\hat{\theta}_{i1B}$ in the bottom 75% of the population. We define $x^{75}_{i1} = I\left[F(\hat{\theta}_{i1B}) > 0.75\right]$ to be an indicator for the high group. We then use decisions regarding the *other two* products to estimate the average valuation weights $E[\theta_{ijk} | k = K, x^{75}_{i1}, j \neq 1]$ using equation (2.4), where $K \in \{B, C\}$. We repeat the procedure twice using products 2 and 3 to generate x^{75}_{i2} and x^{75}_{i3}, and estimate $E[\theta_{ijk} | k = K, x^{75}_{i2}, j \neq 2]$ and $E[\theta_{ijk} | k = K, x^{75}_{i3}, j \neq 3]$. Finally, we average the estimates from each of these three

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

[14] As a concrete example of patterns of behavior that are likely "measurement error" in $\hat{\theta}_{ijk}$, 33.8% of consumers are willing to buy at a higher pre-tax price for at least one product in the standard tax environment than in the no tax environment, 20.4% are willing to buy at a higher pre-tax price for at least one product in the triple tax environment than in the no tax environment, and 25.9% are willing to buy at a higher pre-tax price for at least one product in the triple tax environment than in the single tax environment. Attributing such patterns of behavior to consumers' valuation weights 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. (2019), it is prevalent in most experimental analyses of individual differences.

iterations to get an overall average estimate of θ_{ijk} for those in the high and low groups:

$$E[\theta_{ijk}|k=K,x_i^{75}] = \frac{1}{3}\sum_{\iota=1}^{3} E[\theta_{ijk}|k=K,x_{i\iota}^{75}, j\neq\iota]$$

We compute confidence intervals using percentile bootstrap, clustering by subject. In Appendix B.12 we summarize results that utilize alternative criteria for splitting the sample into high and low groups.

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

Assumption 1. For any j, $\left(\hat{\theta}_{ij'k} \perp \hat{\theta}_{ij''k'}\right)|\theta_{ijk}$ 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 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 for sensitivity to tax rates, but perform our actual estimation of $E[\theta_{ijk}|k=K]$ using only incentivized decisions.

Results

Table 2.1a presents presents results from the split-sample techniques described in Section 2.5.1. Rows (1) and (2) present estimates of $E[\theta_{ijk}|k=K]$ for the high and low valuation weight groups, respectively; row (3) presents estimates of the difference. For comparison, row (4) presents estimates of average θ for the full sample.

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

Consistent with Prediction 3, the low valuation weight group exhibits a significantly larger increase in their valuation weights than the high valuation weight group when tax rates are tripled (0.16 vs. 0.39; 95% CI for difference [−0.43, −0.04]). The high valuation group does adjust moderately upward by 0.39 (95% CI [0.26, 0.51]), but this is not inconsistent with theory when individual differences are taken into account.[15]

[15] As we will show, some individuals overreact significantly, and thus the high valuation group likely consists of individuals both with $\theta<1$ and with $\theta>1$. Thus, because underreacting individuals may adjust upward

2.5.2 Testing Prediction 4

Approach

We also use the strategy in Section 2.5.1 to analyze heterogeneity in how the valuation weights respond to stakes. This provides a test of Prediction 4. Specifically, we use one product to divide consumers according to how much they adjust their valuation weight when the tax rate increases, and we then examine how they respond to taxes on the other two products.

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 $j' \in \{1, 2, 3\}$. We then average to estimate $E[\Delta_{ij}|d_i^{25}]$ separately for the high and low adjustment groups.

Results

Table 2.1b 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 [0.30, 0.75]) *and* in the triple tax regime (0.86 vs. 0.76; 95% CI for difference [−0.01, 0.20]). The result for the triple tax regime is significant at the 10% level (p-value = 0.067).

As we discussed in Section 2.2.2, 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 idiosyncratic 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.

2.5.3 Robustness

In Appendix B.13, we replicate Tables 2.1a and 2.1b on the subsample of participants with nearly accurate beliefs about their sales tax rate and with strong computational ability. In Appendix B.14 we confirm that the results hold for the full sample of participants, including those failing comprehension checks. In Appendix B.15, we replicate results adding the condition that we exclude participants who always buy or never buy a product in any

by slightly more than the overreacting individuals adjust downward, the *average* response to stakes in this group is not guaranteed to be null (or negative).

store. In all three cases, the results conform with Predictions 2-4. In Appendix B.12, we replicate the results splitting at the median, 80th percentile, and 85th percentile.

2.6 Overreaction and heterogeneous attentional responses to stakes

While the evidence in Section 2.5 is consistent with at least moderate individual differences, it leaves open three key questions. First, are the individual differences large enough that some consumers overreact to standard taxes? Second, if we detect overreaction, can we show that some consumers decrease their valuation weights when the stakes are increased? Third, how big is the variance of the valuation weights, which Taubinsky and Rees-Jones (2018) and Farhi and Gabaix (2020) show is a key input into efficiency cost calculations? In this section, we develop econometric techniques for computing lower bounds on individual heterogeneity, which enable us to answer the three questions above.

A key moment that we have not exploited in the analysis in the previous section is how well correlated the binary proxies are with each other. In what follows, we will show how this moment, combined with the results in Section 2.5, helps generate a lower bound on the variance of θ. Intuitively, if our classification of consumers into high versus low θ groups is very imprecise, than the fact that average θ is still so different for consumers in the two groups we create must imply that the individual differences in θ are so high that even conditioning on poor proxies for θ yields large differences in conditional means. 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 2.6.1 below.

2.6.1 Methods for quantifying individual differences

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

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

$$Var[Y] \geq \frac{Cov[Y, X_1] \cdot Cov[Y, X_2]}{Cov[X_1, X_2]} \tag{2.7}$$

and

$$(\bar{Y} - E[Y])(E[Y] - \underline{Y}) \geq Var[Y]. \tag{2.8}$$

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

The result in (2.7) formalizes the intuition above: the less well-correlated the proxies X_i are with each other, the higher must be variance, given an estimate of the covariance between Y and X_i. We prove this result through an application of the Cauchy-Schwarz inequality.

The result in (2.8) is the Bhatia and Davis (2000) inequality. The intuition is that the variance of a random variable Y with a given mean $E[Y]$ and bounded support cannot be higher than the variance of a Bernoulli random variable with mean $E[Y]$ and all mass on the two endpoints of the support.

Proposition 1 enables us to use the types of binary proxies utilized in Section 2.5 to compute bounds on the variance and support of θ_{ij}. Define $x^q_{ijk} = I\left[F\left(\hat{\theta}_{ij} > 0.01q\right)\right]$ as an indicator for $\hat{\theta}_{ijk}$ being in the qth percentile or higher in store k. This is analogous to Section 2.5, 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 for $j \neq j' \neq j''$: the variance of θ_{ijk} in store K is*

$$Var[\theta_{ijk}|k = K] \geq \frac{Cov[\theta_{ijk}, x^q_{ij'k}|k = K] \cdot Cov[\theta_{ijk}, x^q_{ij''k}|k = K]}{Cov[x^q_{ij'k}, x^q_{ij''k}|k = K]} \tag{2.9}$$

and

$$\bar{\theta} \geq E[\theta_{ijk}|k = K] + \frac{Cov[\theta_{ijk}, x^q_{ij'k}|k = K] \cdot Cov[\theta_{ijk}, x^q_{ij''k}|k = K]}{E[\theta_{ij'k}|k = K] \cdot Cov[x^q_{ij'k}, x^q_{ij''k}|k = K]} \tag{2.10}$$

To derive the corollary, we set $x^q_{ij'k}$ and $x^q_{ij''k}$ to correspond to X_1 and X_2 and set θ_{ijk} to correspond to Y in Proposition 1. The assumptions of Proposition 1 are satisfied because $x^q_{ij'k}$ and $x^q_{ij''k}$ are identically distributed by assumption, and because Assumption 1 in Section 2.5.1 implies that $x^q_{ij'k}$ and $x^q_{ij''k}$ are independently distributed conditional θ_{ijk}. In particular, note that by working with $j \neq j' \neq j''$, we avoid any biases that result from conflating true individual differences with measurement error.

Analogously, we use Proposition 1 to derive bounds on adjustment $\Delta_{ij} = \theta_{ijC} - \theta_{ijB}$. We define $d^q_{ij} = I\left[F\left(\hat{\Delta}_{ij} < 0.01q\right)\right]$ as a binary indicator for $\hat{\Delta}_{ij}$ being in qth decile or lower. This is analogous to Section 2.5, 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):*

$$Var[\Delta_{ij}] \geq \frac{Cov[\Delta_{ij}, d^q_{ij'}] \cdot Cov[\Delta_i, d^q_{ij''}]}{Cov[d^q_{ij'}, d^q_{ij''}]} \tag{2.11}$$

and

$$\underline{\Delta} \leq E[\Delta_{ij}] + \frac{Cov[\Delta_{ij}, d^q_{ij'}] \cdot Cov[\Delta_{ij}, d^q_{ij'}]}{(E[\Delta_{ij}] - 1) \cdot Cov[d^q_{ij}, d^q_{ij'}]} \tag{2.12}$$

The assumption that $\Delta_i \leq 1$ is equivalent to assuming that when stakes increase, no consumers switch from being systematic underreactors to systematic overreactors (or that no overreacting consumers substantially *increase* their overreaction). This is consistent with the core of any costly attention model that could 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. To do so, we derive a bound for the fraction of overreactors,

$Pr(\theta_{ij} > 1)$, and for the fraction of consumers who adjust their valuation weight downwards, $Pr(\Delta_{ij} < 0)$. These results follow from a more general result proven in Appendix B.3.2, which can be seen as a converse of sorts to Chebyshev's inequality.

Proposition 2. *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*

$$Pr(\theta_i > 1) \geq \frac{Var[\theta_i] + E[\theta_i]^2 - E[\theta_i]}{(\bar{\theta} - 1)\bar{\theta}} \tag{2.13}$$

and

$$Pr(\Delta_i < 0) \geq \frac{Var[\Delta_i] + E[\Delta_i]^2 - E[\Delta_i]}{(\underline{\Delta})(\underline{\Delta} - 1)} \tag{2.14}$$

Both bounds are tight.

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

2.6.2 Bounds on the variance and the support

Estimation: In our empirical implementation of the bounds, we construct 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 (2.9) is that

$$Var[\theta_{ijk}|k = K] \geq \max_q \left\{ \frac{Cov[\theta_{ijk}, x^q_{ij'k}|k = K] \cdot Cov[\theta_{ijk}, x^q_{ij''k}|k = K]}{Cov[x^q_{ij'k}, x^q_{ij''k}|k = K]} \right\}. \tag{2.15}$$

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

Results: Table 2.2 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 2.2 also presents the supremum lower bound estimates. We estimate a lower bound of $\bar{\theta}_B$ to be 2.21 (5% confidence bound of 1.55) and for $\bar{\theta}_C$ to be 1.69 (5% confidence bound of 1.54). Both of these bounds are significantly above 1, indicating there are overreactors in the experimental population. Row (2) presents the estimate as a lower bound on $-\underline{\Delta}$. We estimate an upper bound on $\underline{\Delta}$ to be 0.94 (5% confidence bound of 0.16). This result is consistent with Prediction 5 of our theoretical results.

2.6.3 Bounds on propensity to overreact to taxes

By substituting the lower bound for $Var[\theta_{ijk}|k = K]$ from equation (2.15) into the lower bound from equation (2.13), we derive a bound for the fraction of overreactors as a function of $\bar{\theta}$.

Figure 2.3a plots $Pr(\theta_{ijk} > 1|k = K)$ as a function of $\bar{\theta}$, both for $K = B$ and for $K = C$, along with 5% confidence bounds computed by bootstrap clustered at the subject level. When $\bar{\theta} = 2.25$, we estimate that at least 20.5% (5% confidence bound of 9.5%) of the population is overreacting in store B, and at least 19.3% (5% confidence bound of 14.5%) is overreacting in store C. Both bounds are decreasing in $\bar{\theta}$: for $\bar{\theta} = 4.25$, we bound the fraction of overreactors at 4.2% (5% confidence bound of 2.0%) of the population is overreacting in store B, and at least 3.9% (5% confidence bound of 3.0%) in the triple tax environment.

Using equation (2.14), we derive an analogous bound for the fraction of consumers who adjust their valuation weight downward in response to higher stakes. Figure 2.3b plots $Pr(\Delta_{ij} < 0)$ as a function of $\underline{\Delta}$. For $\underline{\Delta} = -0.94$, the bound computed in table 2.2, we estimate that at least 35.5% (5% confidence bound of 6.1%) of consumers negatively adjust their revealed valuation weights when switching from the standard tax regime to the triple tax regime. This lower bound is decreasing in the magnitude of $\underline{\Delta}$, and at $\underline{\Delta}$ = -4.25 we estimate a lower bound on the fraction of participants with negative adjustment to be 4.1% (5% confidence bound of 0.7%).

Robustness In Appendix B.13, 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 B.14 we confirm that the results hold for the full sample of participants, including those failing comprehension checks.

2.7 Concluding remarks

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

Evidence of costly attention implies that shrouding taxes can generate deadweight loss by imposing cognitive costs on consumers. Evidence of significant heterogeneity in attention, generating both under- and overreaction to opaque prices, implies that there may be significant deadweight loss from misallocation of products to consumers, which we show in Appendix B.18 by utilizing the deadweight loss formula from Taubinsky and Rees-Jones (2018).

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

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

Yet we still think that providing evidence for costly attention in our setting can update researchers' priors about the likelihood that similar costly attention mechanisms have a potentially important qualitative role in other settings, including those with higher stakes. Empirical settings that in principle can be analyzed using quasi-experimental analogues of our methods would involve panel datasets where it is possible to observe individuals' behavior across multiple changes in the opaque attributes.

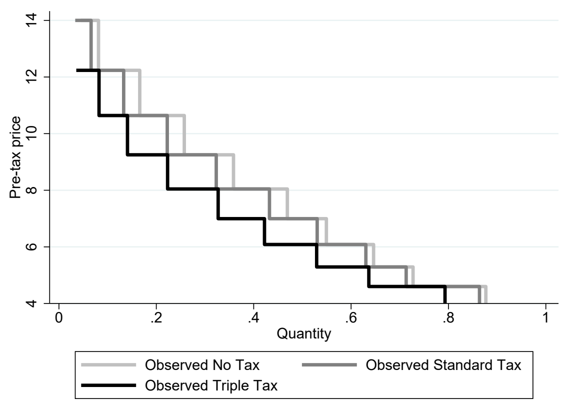
There may also be important interactions between costly attention models and other mechanisms. A key open question is where consumers' highly heterogeneous rules of thumb come from. Theories of memory and "experience effects" may serve as foundations (e.g., Mullainathan, 2002; Malmendier and Nagel, 2016;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

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

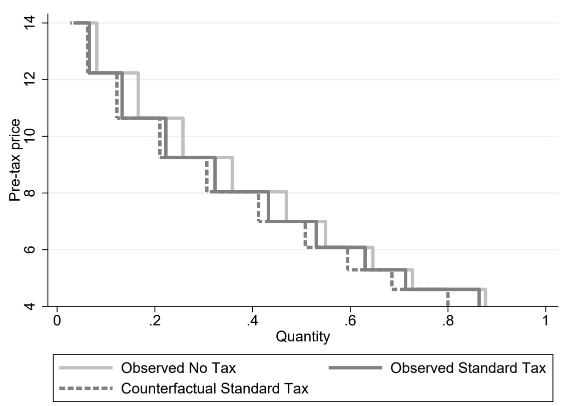
Figures and Tables

Figure 2.1: Demand curves

(a) Observed demand curves



(b) Observed vs. counterfactual demand: standard taxes



(c) Observed vs. counterfactual demand: triple taxes

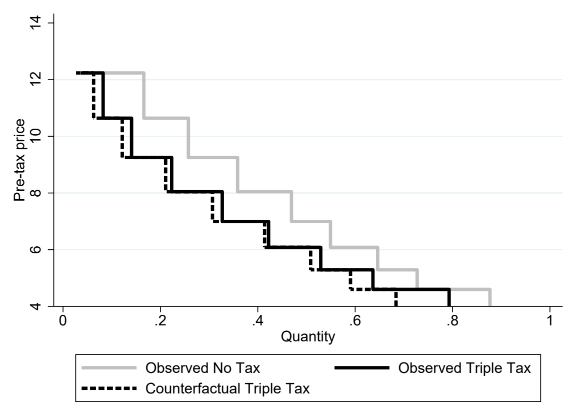
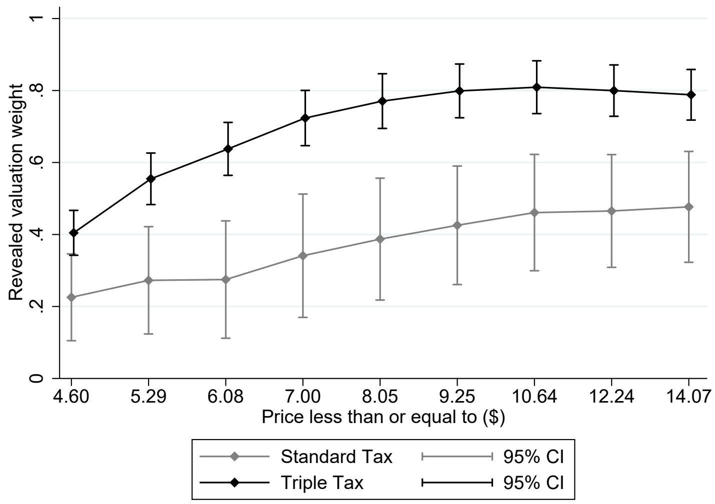


Figure A.3.1 presents demand curves, averaging across all nine products. We with demand curves $D_{jk}(p)$ where j indexes products and k indexes the store type, A (no tax), B (standard tax), or C (triple tax). Panel (a) presents the average demand curves $D_{avg,k}(p) := \frac{1}{9}\sum_j D_{jk}(p)$ 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 2.4.1.

Figure 2.2: Average revealed valuation weights by stakes

(a) Average revealed valuation weight (θ) by posted price



(b) Average revealed valuation weight (θ) by average tax owed

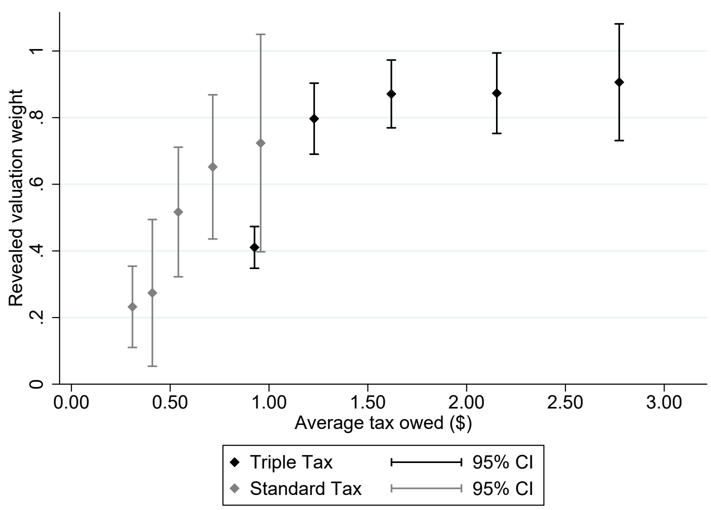
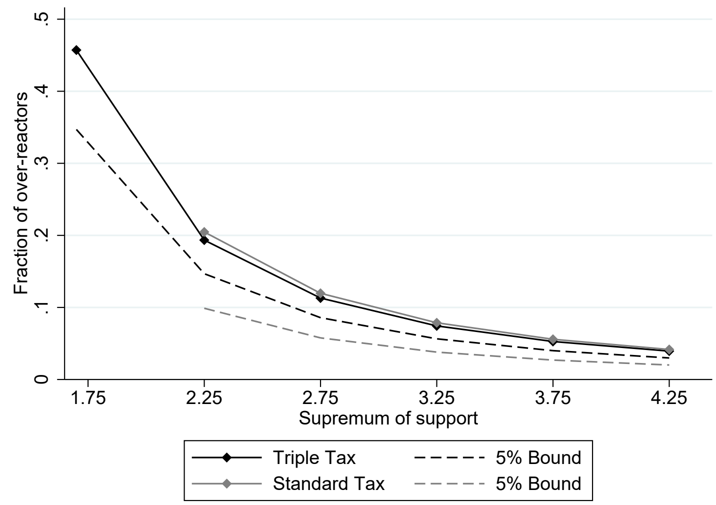


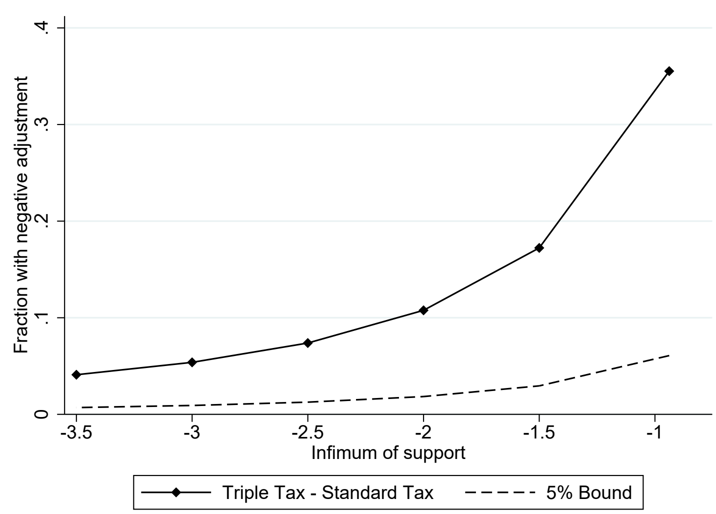
Figure 2.2a presents estimates of average θ, using estimating equation (2.4), for prices less than or equal to the cutoff specified on the x-axis. θ is defined as the revealed valuation weight of the sales tax (e.g., $\theta = 0$ is complete neglect, $\theta = 1$ is accurate processing, and $\theta > 1$ is overreaction). Figure 2.2b presents store-specific estimates of average θ by the average tax owed within each bin. For each tax environment—store B and store C—each bin is constructed by dividing the prices in the experiment into 5 ordered pairs. The average tax owed is constructed by taking the average of the two prices in each bin, and multiplying it by the average tax rate in stores B and C, respectively. The estimating equation is an extension of equation (2.4), described in equation (2.5). Standard errors are clustered at the subject level.

Figure 2.3: Bounds on overreaction

(a) Bounds on the fraction of individuals who overreact to taxes



(b) Bounds on the fraction of individuals whose valuation weight θ decreases with higher stakes



Panel (a) of this figure presents store-specific estimates for the lower bound on the fraction of consumers with revealed valuation weight $\theta > 1$, as a function of the supremum of the support of θ, while panel (b) presents estimates for the lower bound on $Pr(\theta_{iC} - \theta_{iB}) < 0$ as a function of the infimum of the support. The lower bound on $Pr(\theta_{ik}) > 1, k \in B, C$ is estimated from equation (2.13). The lower bound for the fraction of consumers who adjust their valuation weight downward in response to higher stakes is estimated using equation (2.14). The dashed lines present the 5% lower bound computed from a percentile-based bootstrap (1000 replications, clustered at the subject level).

Table 2.1: Average revealed valuation weights by group

(a) Average revealed valuation weights by valuation weight group

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

(b) 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, 0.85]	[0.30, 0.55]
(3): (1) − (2)	0.52	0.10	−0.42
	[0.30, 0.75]	[−0.01, 0.20]	[−0.60, −0.24]
(4): Full sample	0.48	0.79	0.31
	[0.32, 0.63]	[0.72, 0.86]	[0.20, 0.42]

Rows (1) and (2) of panel (a) present estimates for the high and low valuation weight groups, whose construction is described in Section 2.5.1. Rows (1) and (2) of panel (b) present estimates for the low and high adjustment groups, whose construction is described in Section 2.5.2. Row (3) presents the difference of the estimates in rows (1) and (2), for each column. Row (4) presents estimates using the full sample. The "Standard" column contains estimates of store B valuation weights in each of the two groups, as well as the differences between these groups. The "Triple" column contains estimates of store C valuation weights in each of the two groups, as well as the differences between these groups. The "Triple − Standard" column presents estimates of $E[\theta_{ijC}] - E[\theta_{ijB}]$ for each of the two groups in rows (1) and (2), and contains the differences in differences in row (3). Bootstrapped confidence intervals from 1000 replications, clustered at the subject level, are reported in brackets.

Table 2.2: 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 this table present store-specific estimates of the lower bound on $Var[\theta_{ijB}]$ and $Var[\theta_{ijC}]$,and on the supremum $\bar{\theta}$. Column (3) presents estimates of the lower bound of $Var[\theta_{ijB} - \theta_{ijC}]$ and the supremum of $\theta_{ijB} - \theta_{ijC}$. The methodology is described in Section 2.6.1 and the estimating equations are described in Section 2.6.2. Fifth percentile results from 1000 bootstrap replications, clustered by subject, are reported in brackets.

Chapter 3

The Effects of Price Discounts on Consumer Behavior and Beliefs: Evidence from a Field Experiment in the Apparel Industry[1]

3.1 Introduction

It is well known that many firms across different industries routinely offer price promotions. However, both academics and firm executives have radically different views surrounding the purpose and optimality of offering sales to customers. In this paper, we partner with a luxury online retailer to conduct experiments to clarify the trade-offs that companies face.

Proponents of frequent discount will argue that the practice allows firms to attract new customers Conlisk et al. (1984); Nakamura and Steinsson (2011); Narasimhan (1988), excite returning customers Slade (1999), improve brand loyalty de Oliveira Santini et al. (2016); Oyeniyi (2011) and boost short-term revenue Edelman et al. (2011); Gupta (1988). Others counter by claiming the practice reduces unit-profit margin Ailawadi et al. (2007); Edelman et al. (2011), antagonizes existing customers Anderson and Simester (2010); Dholakia (2006); Rotemberg (2005); Courty and Pagliero (2008), cheapens the perceived prestige of the brand Erdem et al. (2008); Anderson and Simester (2001a); Gneezy et al. (2014), and decreases customers' beliefs about the fair value of the products Bordalo et al. (2020b); DelVecchio et al. (2007); Anderson and Simester (2004). The argument against discounting is well summarized by Yves Carcelle, the former CEO of Louis Vuitton. In the context of describing why Louis Vuitton destroys excess inventory rather than offering discounts, he stated[2]: "We're never on sale. All the rest discount (sic). Us, never. When a customer invests in one of our products, they don't expect to see it discounted three weeks later, so we don't do it."

[1] Coauthored with Daniel Morrison, Princeton University. The survey and experiment were approved by Princeton University IRB# 13993. Princeton University served as the IRB of record for the study, and UC Berkeley relied it for review.

[2] Shearn, Michael. The Investment Checklist: The Art of In-Depth Research. Page 58.

In partnership with an online retailer, we conduct a randomized control trial in which we divide historical customers into two groups. The randomly selected treatment group received an email advertisement and a 20% coupon off select product categories, while the control group received just an email advertisement. We link treatment group assignment and coupon use to customers' order history both before and after the coupon's expiration.

Our main finding is that customers assigned a coupon did not spend more than the control group in the three months after the coupon was sent out, despite the fact that they spent 30.7 percent more in the two weeks during which the coupon was active. Consistent with our model, this suggests that rather than inducing more spending, the coupon induced customers to shift their purchases earlier.

To explore different mechanisms from the literature we follow Anderson and Simester (2001b); Shu and Gneezy (2010) in conducting an incentivized survey on existing customers at our partner company. We find that introducing the coupon did not affect consumers' average willingness to pay for the product, nor did it affect the perceived brand quality. However, customers in the treatment group expected more frequent and deeper discounts for all products in the future, including those ineligible under the current coupon.

To rationalize these findings, we introduce a simple model in which consumers trade off their desire to receive a product earlier with trying to minimize their purchase price. Consumers who observed past discounts assign a higher subjective probability of the likelihood of discounts happening again. This causes them to forego purchases during periods when products are listed at full price. Firms trade off higher current period profits when discounting with a decrease in long-term profitability due to a change in customer expectations about the likelihood of future discounts. Under plausible parameters, the model suggests that firms with customers primed for discounts can experience a reduction in long-term profitability by more than 10% compared with customers who are used to infrequent discounts. We also show that the optimal discount probability chosen by firms depends on customers' Bayesian priors about the likelihood of discounts being offered each period. When customers are primed to expect discounts, firms optimally choose to offer discounts regularly as even relatively impatient customers, acting on their beliefs, would rather wait for a likely future discount than pay the regular price today.

Our paper contributes to the literature in experimental economics involving the effects of discounting on firm profitability and customer behavior when products are semi-durable luxury goods. Several previous papers have utilized laboratory and field experiments to evaluate the consequences of discounting Simester (2017). For example, Yi and Yoo (2011) finds a decrease in laboratory participants' perceived brand quality of MP3 players after being subjected to repeated price promotions. In a field experiment involving luxury goods, Anderson and Simester (2001b) find that offering delayed payment options can decrease perceived brand quality. In our paper, we also find that frequent discounting can negatively affect a firm's long-term profitability, but we do not find any evidence supporting a brand quality channel. Instead, we take the novel approach of combining our field experiment with a survey that directly inquires about customers' beliefs about the likelihood and depth of future discounts. We find strong evidence that even a single discount can induce significant changes in customers perceptions about the likelihood of future discounts. We argue that it is this change in beliefs, rather than a change in perceived brand quality, that affects customers willingness to pay.

We also contribute to a large theoretical literature surrounding the optimal price promotion policy. Several previous studies have drawn attention to the importance of a customers internal reference prices and loss aversion when making decisions about purchasing products. In one example, Koszegi and Heidhues (2014) argues that, due to loss-averse customers, firms with market power optimally set a sticky regular price with randomly occurring sale prices. When studying the demand for storable goods Hendel and Nevo (2006) find that after a discount, customers wait longer to purchase the product again. In our model, we find that even when faced with risk-neutral customers and unit demand, a firm still may want to offer a non-trivial mix between a high and low price. Furthermore, a customer's prior beliefs about the likelihood of future discounts will greatly impact the optimal frequency of firms offering price promotions.

Section 3.2 provides both information about our partner organization and describes our experiment. In Section 3.3 we walk through the impacts of offering a discount code on customer purchases both before and after the period of discount eligibility. Section 3.4 discusses the design of the survey given to the firm's customers in the experiment, while Section 3.5 describes the results. We present a simple model in Section 3.6 to rationalize our findings. Finally, Section 3.7 concludes.

3.2 Description of experiment

We partner with an online retailer to conduct a randomized control trial to evaluate consumers' short-term and long-term responses to discounting.[3] Our corporate partner is a fast-growing luxury outdoor apparel store with annual revenue between $50 million and $100 million during the year of our partnership. Although they have recently opened a few brick-and-mortar stores, over 90% of their revenue derives from online orders. This company offers two clothing lines each year, one geared towards fall and winter ("Fall season") and the other towards spring and summer ("Spring season").

Throughout their history, this company has rotated through a variety of discounting regimes as seen in Figure 3.1. Prior to 2016, average discounts given to returning customers for in-season items were less than 10%. Since 2016, this company has occasionally offered deeper discounts, especially later in the season for products that are less popular or that have high inventory; however, discounts for products that are expected to sell out rarely occur when the product is in-season. Historically, this company predictably offers discounts ranging from 15-40% for remaining inventory when their products become out of season.

In our experiment, we conduct a randomized control trial to evaluate customer responses to a material change in discounting policy. They offered customers a 20% off coupon for specific product categories, all products listed as fleeces or insulation, in their order. By allowing the coupon to be used on multiple items, we sought to mimic a store-wide sale on the product category.

This coupon was individualized so that only one customer could use each coupon code. Fleeces and insulation were chosen as the eligible product category as these had relatively high sales volume and had not yet been offered at a price below their regular price for existing

[3]We received IRB approval from Princeton University, who did a joint review for Princeton and UC Berkeley. IRB #13993.

customers this season. The eligible products account for approximately half of the company's revenue. The remaining products were ineligible for the coupon. By restricting the coupon to select product categories, we are able to study the effect of the coupon not only on eligible products, but also on ineligible products. In particular, we can identify whether offering this coupon (i) induces more sales in the short-term for ineligible products by bringing customers to the website or (ii) alters sales in the long-term for ineligible products by priming customers to wait for sales, or otherwise reduces their internal reference price for these products due to changing brand perceptions.

We randomly selected two-thirds of the approximately 200,000 (check number) existing customers and sent an advertisement for the eligible products and the coupon to their email address on file. To isolate the effect of the coupon from the effect of receiving an email, the remaining one-third of customers also received the same email advertisement for the fleece and insulation line.

Customers had just over two weeks, from January 27 to February 14, to redeem the coupon. They received no reminder emails or additional correspondence about the sale.

3.3 Experiment results

In total, 51,995 past customers received the advertisement from the company, and two-thirds of them also received a coupon code.

Consistent with the nature and durability of the product, this company receives infrequent but high-priced orders from customers. The average historical customer has placed just over 2 orders with the company (median = 1) and received approximately 5 items (median = 2). On average, their first order was placed 21 months (median = 14 months) prior to the start of the experiment, and their most recent order was placed approximately 12 months (median = 10 months) prior to the experiment. Over their lifetime, customers have spent an average of approximately \$600 (median = \$327) with the company and an average of \$145 per item ordered (median = \$119).

Table 1.1 decomposes the variables into the treatment group-specific averages. As expected for a randomized control trial, there are no statistically significant differences between the two groups. Customers in our experiment have averaged 2.2 historical orders. The average customer has spent 606 dollars and received 4.9 items. On average, it has been about one year since the average customer placed their most recent order and about 21 months since they placed their first order.

Figure 3.1 presents a time-series of the historical average discount percentage obtained for returning customers on in-season products at the quarterly frequency. While the overall average discount obtained is approximately 10%, the company has offered larger discounts since 2017, with some decline in 2021. Importantly, these averages reflect the discount received by the customer, not the average depth or frequency of discount offered by the company. Since customers are more likely to purchase discounted items, this reported average is an overestimate of the average discount typically available to returning customers for in-season items. Nevertheless, our experiment offers a coupon of 20% which is significantly higher than these averages.

By linking coupon use to buyer's purchases in and after the experiment, we are able to measure the total purchases by the treatment and the control groups both during and after the period of coupon eligibility. Figure 3.2 presents the aggregate results, separately for products eligible and ineligible for the coupon code. Since the control group was half of the size of the treatment group, we doubled the purchases from it to make the two groups comparable.

This chart provides striking visual evidence in favor of our hypothesized mechanism. In particular, we find that the treatment group spent more during the period of coupon eligibility on coupon-eligible products. However, they spent less in the weeks after the coupon expired, such that the total revenue in these groups from January 27 to March 31 was approximately the same. This suggests that the coupon is not leading customers to spend more at the company, but rather shifting purchases to the period of coupon eligibility. We see similar trends in both profit and items ordered during this period, as reported in the Appendix.

In contrast to coupon-eligible products, we observe no aggregate differences in revenue, profit, or units ordered between the two groups during either the period of coupon eligibility or the period after it. This provides strong evidence that offering the coupon for eligible products did not affect customer's preferences over ineligible products.

3.4 Description of survey

Four days after participants received the email advertisement with or without the coupon, all participants were sent a second email with the subject line "help make [company] work for you" inviting them to take a 5-minute survey through which they could win prizes such as company products or Amazon gift cards. They were told the survey was done in partnership with UC Berkeley and Princeton researchers, but not given any information about the purpose of the research.

The survey began by explaining the nature of the incentives during the program. Participants were told that five participants would receive an Amazon gift card, four for \$100 and one for \$250, and that one participant would receive a company product valued at over \$100 based on their responses. Crucially, participants were told that their responses would not be seen by any company employees, but only by the UC Berkeley and Princeton researchers.

After giving consent, participants were asked if they preferred to receive men's apparel or women's apparel. Based on their responses, participants then saw a series of multiple price list (MPL) questions related to men's or women's products. Participants were first shown a company fleece that would have been eligible for the coupon, along with a similar product from competitors. We then informed customers that it was "in your best interest to answer honestly", as they had a random chance of being selected to receive a company product based on their responses to the MPL. Specifically, we told participants that a computer would randomly choose one row on the MPL and they would receive their choice. See Appendix XXX to view the MPL.

After answering the MPL for the company product, customers then completed an MPL for a comparable competitor product, whose value was presumably unaffected by the coupon.

They were then asked to complete an MPL about receiving the product now or in 6 months, to measure their time preferences and patience.

Participants were then asked a series of questions about their beliefs for the likelihood of future discounts. First, we asked participants for, over the 100 days from February 15 to May 26, how often do you think you will be able to purchase for certain a listed product for percentages off the listed price. Note, we started the 100 day window on February 15th as that date is after the expiration of the coupon that customers in the treatment group received. Specifically we elicited for the following intervals: 0%, 1-10%, 11-20%, 21-30%, 31-40%, 41-50%, and 51% or more. Participants could then fill out how many days they thought the product would be on sale at each discount range, and we forced the responses to add to 100.

Second, participants were asked what they thought was the lowest price, also displayed as a percent discount, that they would be able to purchase the product over the 100 days following the expiration of the coupon. We then repeated the question, asking participants what they thought the lowest price they thought they were "very confident" they would be able to receive the product over the 100 days following the expiration of the coupon.

Participants then repeated all these questions for a second product, which was not eligible for the original coupon. This allowed us to analyze whether a coupon on one specific product category affected the perception of discounts and valuation of the company's products more broadly.

Finally, participants answered a series of qualitative questions about the company and its competitors. First, participants were asked to rate the company and several competitors from 1 to 10 on how likely they are to recommend the company to their friends and family, to what extent they think the company cares about their customers, and if they recall a sale or coupon from the company in the past 30 days. We concluded by asking basic demographic information including their age, income, and gender.

3.5 Survey results

728 customers completed the incentivized survey. Of these, 490 had received the coupon while 238 had not received a coupon. We find that customers have some memory of receiving the coupon, but that many respondents have inaccurate beliefs. From the sub-sample that received the coupon, 49 percent said they had received a coupon in the last month, while 33 percent of those who did not receive the coupon recalled receiving the coupon in the past month. Given the experiment involved a single email advertisement, with no reminder emails, we did not expect perfect recall on this question. In our main sample, we exclude participants who inaccurately stated they did not receive a coupon.

Willingness to Pay (WTP) results: Using customer responses to the MPL questions, we can estimate aggregate customer valuations across our two treatment groups. Table 3.4 reports the results.

Across these questions, we do not find any statistically significant results for the differences between the coupon and no-coupon groups. For the corporate product that was eligible for the discount, customers assigned a coupon are, on average, willing to pay $6.19 more for the product compared to the control group ($127.28 versus $121.09, p-value 0.179). Likewise for the company product ineligible for the discount, there are no discernible differences between the willingness to pays of the two groups (\$87.95 versus \$85.71, p-value 0.581). For the competitor product, customers assigned a coupon are, on average willing to pay \$6.16 more for the product compared to the control group (\$112.97 versus \$106.81, p-value 0.183). Interestingly, the group that received the coupon is willing to pay more for both the coupon-eligible product and the competitor product than does the control group, which is the opposite of what we predicted and what is predicted by theory. However, since neither result is significant at the 10% level, both findings could simply be explained by noise.

Beliefs about Future Discounts results: Here we focus on the results which pertain to the largest discount that customers are "confident" or "very confident" they can receive in the next 100 days. Table 3.3 presents the results.

For the product eligible for the discount, we find that 73.9% of customers in the control group believe they will receive a discount of at least 20% in the 100 day period. This increases by 7.2 percentage points (p-value = 0.059) for customers who received the coupon. For the product ineligible for the discount, 73.9% of customers in the control group report that they believe they will receive a discount of at least 20% in the 100 day period. This increases by 6.8 percentage points (p-value = 0.076) for customers in the treatment group.

We find similar results when asking the same question with the "very confident" language. For the product eligible for the discount, 46.6% of customers who were not assigned a coupon state that they are very confident they can receive a discount of at least 20% off in the 100 day period. This percentage increases by 10.3 percentage points (p-value = 0.025) for the respondents who received a coupon. Likewise for the product ineligible for the discount in our experiment, 49.2% of customers in the control group report that they are "very confident" they will receive a discount of at least 20% off. This increases by 11.5 percentage points (p-value = 0.011) for the treatment group.

Finally, when we ask customers how many of the next 100 days they think they will be able to receive the product on discount, we find that the treatment group believes the eligible product will be discounted for an average of 3.61 (p-value = 0.200) more days than does the control group. For the ineligible product, this difference is 4.82 (p-value = 0.090).

Collectively, these survey questions provides strong evidence that customers who received a coupon in the past revise upward their beliefs about the likelihood of future discounts.

Other results: Analysis of our qualitative questions asked at the end of the survey yields two interesting findings. First, we find statistically significant differences in measures of brand quality between the two groups, but in a manner not predicted by most existing theories. Specifically, we find that customers who received a coupon perceived the company's brand quality to be roughly the same as the control group (9.15 versus 9.23 out of 10), but rated competitor's brand quality *lower* than the control group (5.73 versus 5.97 out of 10, p-value for difference: 0.15). Second, customers assigned a coupon reported a statistically higher belief in the company "caring about its customers" (9.03 versus 8.66 out of 10, p-value for difference: 0.012), but no differences for the same measure among competitors (6.38 versus 6.45 out of 10). Collectively, these results suggest that offering the coupon (i) does not damage the brand to existing customers[4], and (ii) increases perception among existing customers about the company caring about them.

Interpretation: Much of the literature[5] that describes negative effects of discounting focuses on how frequent discounting negatively psychologically affects a customer's experience or worsens a customers perception of the brand. These negative consequences of discounting are thought to affect customer's willingness to engage with the firm or reduce the internal valuation that customers place on the company's product. The claim is that these psychological factors reduce the likelihood that customers will pay full price in the future. In our survey, we find no evidence for these psychological explanations. Instead, we even find weak evidence in the opposite direction. Nevertheless, our experimental results match much of the past literature in finding that customers who receive discounts are less likely to buy products at full price in the future. Our survey results indicate one plausible explanation for this, which is that after receiving a discount, customers believe that discounts will come more frequently in the future. This incentivizes customers to wait to make a purchase until those discounts come. We explore this mechanism through a simple model in the next section.

3.6 Model

The empirical evidence displayed in the previous two sections contradicts many of the psychology-based arguments against discounting discussed in the literature. From our incentivized surveys, we find no evidence that customers perceive a lower brand value after a discount. We also find that, counter to the findings in Anderson and Simester (2010), customers who received a discount believe the company cares more about customers, rather than viewing the frequent price changes as antagonizing them. Despite this, we find no improvement in revenue or profit from the group that received the coupon.

In this section, we rationalize this finding by presenting a simple model in which customers trade off timeliness and expected price when deciding when to purchase a product. When customers believe discounts occur frequently, they disproportionately wait for the next discount to place an order. On the other hand, when customers believe temporary price reductions occur only rarely, their impatience makes them more likely to place orders at full price. Firms select the probability of offering a discount to maximize profit. We find that the firm-optimal discount probability depends heavily on borrower expectations about future discounts.

Firm Problem: A firm creates a single product with constant cost of production c. Each period, the firm makes their product available for a price p_t. The firm either offers its product at an undiscounted price p^H or at a discounted price p^L. For simplicity, these prices are exogenously determined. The firm chooses the probability, θ, that the discounted price will be offered to customers in each period. The firm's goal is to choose this to maximize their expected present discounted value of profits given a firm discount rate β. This is described mathematically in equation (1) below, where q_t is the number of units sold in period t:

[4] Furthermore, offering the coupon may even weaken the brand quality of competitors among customers, though we are under-powered to verify this claim (p-value: 0.15).

[5] Examples include Anderson and Simester (2001b, 2010); Dholakia (2006); Rotemberg (2005); Courty and Pagliero (2008); Gneezy et al. (2014); Erdem et al. (2008)

$$\theta^\star = \arg\max_\theta \sum_{t=1}^{\infty} \beta^t E[q_t(p_t - c)] \tag{3.1}$$

Customer Problem: Each period, a single new customer enters the market and stays in the market until making a purchase. Customer i, who enters the market in period i, values the first unit of the product at a level v_i, and places 0 utility on all subsequent units. Their valuation for the product does not change through time. For simplicity, assume that v_i has bounded support between v_L and v_H. If $v_L < p^L$,the customer will never buy the product, even if it is maximally discounted; we therefore consider only the interesting cases where $v_L \geq p^L$. At each period t, with probability $1-\theta$, the product is offered at full price, so that $p_t = p^H$. With probability θ, the product is discounted to a level $p_t = p^L$. At each period, customers choose whether to buy the product or wait for the next period. Once a customer buys the product, she exits the market. Upon making a purchase, customer i receives utility of $\delta^{t-i} * (v_i – p_t)$ where δ is a customer's discount rate. Customers do not know θ, but instead have prior beliefs that θ is distributed according to a Beta distribution with initial parameters a_0 and b_0. Thus, in expectation, customers in period t think there is an $\frac{a_t}{a_t+b_t}$ probability that the product will be discounted in the next period. Upon observing the discounting decision, a Bernoulli draw, customers update their beliefs about the distribution of theta according to Bayes Rule. Thus, $a_{t+1} = a_t + I(p_t = p^L)$ and $b_{t+1} = b_t + I(p_t = p^H)$.

First, note that any time a customer observes a price p^L,she will buy the product and exit the market. The nontrivial decision comes when the customer observes a price p^H. If $v_i \leq p^H$, the customer will wait for a discount in a future period. If $v_i > p^H$,the customer's utility from buying today is $v_i - p^H$. By declining to buy today, the customer's expected utility is given by the value function in equation (2), where the state variables define the customer's updated beliefs about the likelihood of future discounts:

$$U(a, b) = \delta \left(\frac{a}{a+b} \left(v_i - p^L\right) + \frac{b}{a+b} U(a, b+1) \right) \tag{3.2}$$

The customer chooses to buy the product today if $v_i - p^H > U(a, b)$. As we can see, several factors influence a customer's decision to buy now at full price or wait for a potential future discount. More patient customers, smaller δ, are more likely to wait for a discount. Customers who value the product relatively less, so that $(v_i - p^L)$ is large relative to $(v_i - p^H)$, are also more likely to wait. Finally, customers who believe a discounted price is likely to occur sooner, a_t is large relative to b_t, are more likely to wait. This last condition is particularly interesting as a_t and b_t are influenced by past firm discounting behavior.

Simulation:

Through a Monte Carlo simulation, we algorithmically compute the firm's optimal discount probability for different values of customers' prior beliefs about the likelihood of discounts. We see that these initial beliefs have a major impact on the firm's optimal choice and on the firm's expected profitability. When customers believe a firm discounts frequently, that firm is "forced" to continue to offer discount often or else face steep profit declines while resetting customer expectations.

In Figure 3.3, we compare the profitability and optimal choice of discounting probability in four different scenarios. In the baseline scenario, in black, customers start with a uniform prior over the firm's discount probability. In blue, customers initial beliefs are that discounts are relatively more rare occurring one time in six. In red, customers beliefs are that discounts are frequent, occurring five times out of six. We see that a firm optimally chooses to discount more when customers believe discounts are more frequent as many of these customers are primed to only buy when the product is on sale. Under the model specifications below, this results in upwards of a ten percent loss in the present discounted value of firm profitability.

Also shown in green in Figure 3.3 is what a firm's optimal discount policy would be in a world in which customers have dogmatic priors. For this simulation, we assume that customers believe there is a ten percent chance of a firm offering a discount in each period. This specification was set to match the firm's optimal policy in the baseline scenario. In this environment, a firm optimally chooses to discount more than twice as often as it does in the baseline scenario, resulting in a 4 percent increase in profits. The intuition for this result is the firm gets all the same benefits from discounting, but avoids paying the cost of incentivizing future customers to wait for future discounts.

Takeaway:

While this model is quite simple, it highlights the importance of a firm managing customer's expectations about future discounts. Offering a discount does not simply boost short-term profits at the expense of a firm's short-term profit margin, but it also changes customer expectations about future discounts and results in lower long-term profitability. This is exactly our finding from the experiment; when the company offered 20% discounts to customers, these customers bought more during the period of discount eligibility, but then diminished their purchases afterwards relative to the control group that did not receive a discount code. Survey results confirmed that customers did in fact have altered beliefs about the likelihood of future discounts after receiving a single discount code.

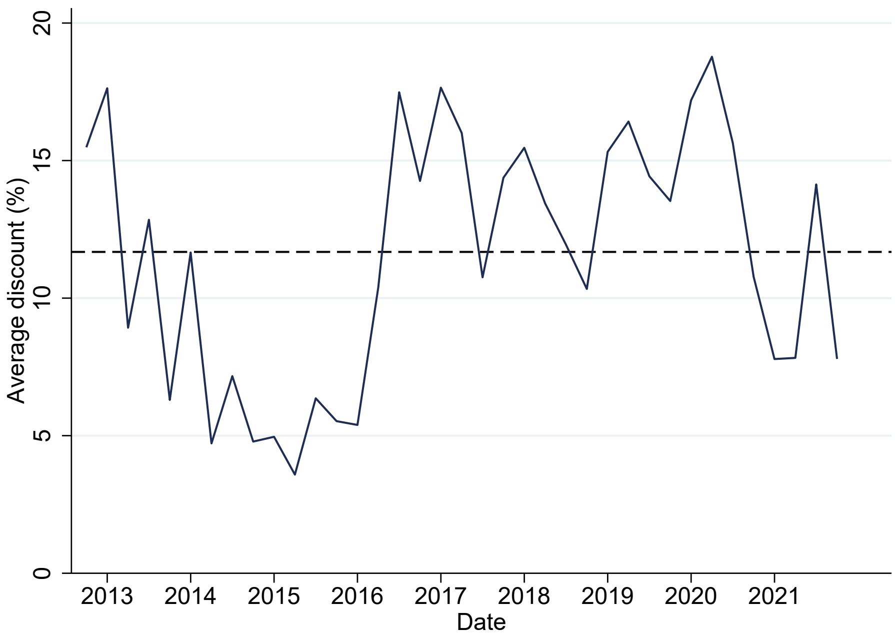
3.7 Concluding Remarks

We partner with an online apparel retailer to run a randomized control trial to evaluate the effects of discounting. We find that customers who received a coupon increased their purchases during the period of coupon eligibility, but that this increase was almost perfectly offset by a similar-in-magnitude decrease in purchases of full price products after the coupon expired. In addition, we conducted an incentized survey to try to uncover the mechanisms through which discounting alters customer behavior. We find no evidence for any psychological effects of discounting, such as antagonizing customers or diminishing the perceived value of the brand; however, we do find that customers who receive a discount revise upward their beliefs about the likelihood of future discounts. Through a simple model, we show that this mechanism alone can be large enough to provide firms with an incentive to discount less.

The external validity of our experiment is up for debate. Our experiment differs from much, but not all, of the prior literature in that ours is conducted on luxury goods that are semi-durable in an industry that is highly competitive. It is entirely plausible that frequent discounting has different effects on customers shopping for nondurable goods or on products that are not luxury products. We believe that examining the heterogeneity in the effect of discounting on different product types is a promising area for future research.

Figures and Tables

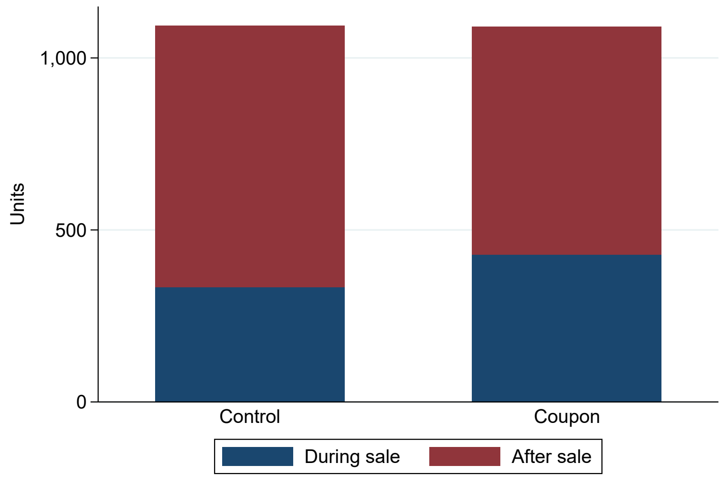
Figure 3.1: Historical average discounts used for return customers on in-season products

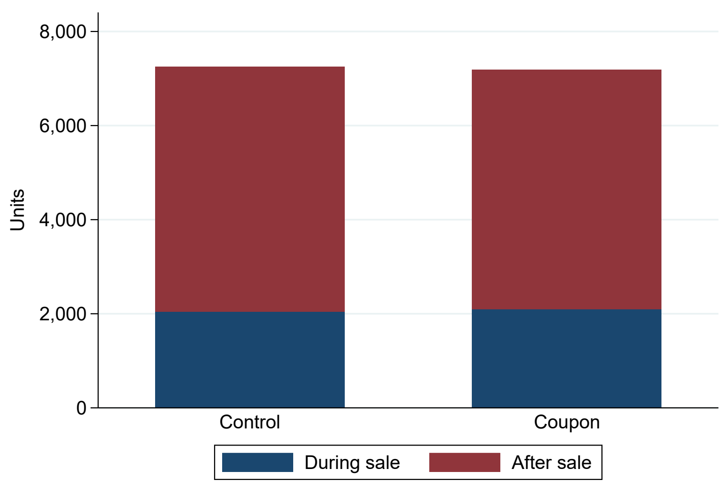


This figure shows shows the quarterly average discount at purchase for in-season products by return customers.

Figure 3.2: Revenue from the treatment and control groups during and after the sale

(a) Eligible Products

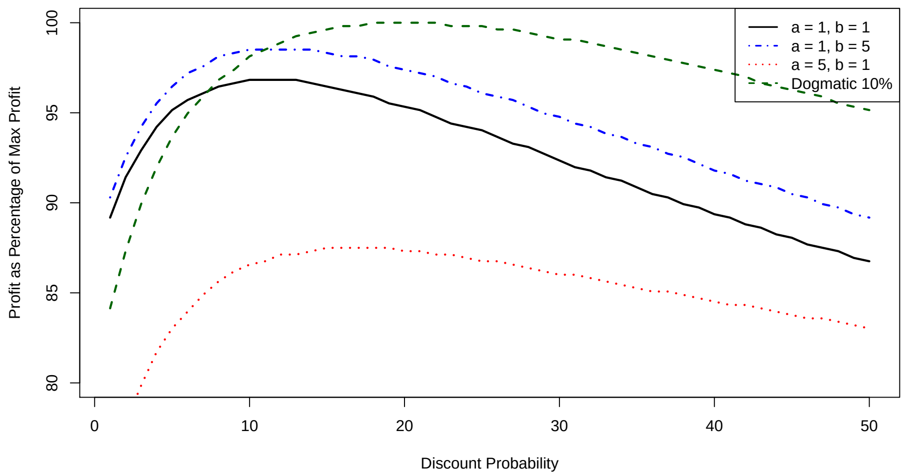




(b) Eligible Products

This figure shows the number of units sold to customers in the experiment before and after the sale. We double the quantities in the control group as there were twice as many participants assigned to the coupon group than the control group.

Figure 3.3: Effect of Discount Probability and Beliefs on Firm Profits



This plot shows a firms present value of profits as a function of their chosen discount probability. The vertical axis is rescaled so that 100% reflects the maximum profit across any of the scenarios. Four different customer priors are overlayed on this plot. The baseline scenario (black) represents customers with uniform priors over the likelihood of a discount. In blue (red), customers priors are that discounts are relatively less (more) frequent. Finally, the green dashed line considers customers with dogmatic priors. For the Monte Carlo simulation, we set $\beta = 0.95$, $\delta = 0.75$, $c = 50$, $v_L = 100$, $v_H = 200$, $p^L = 100$, and $p^H = 200$.

Table 3.1: Balance table

	No coupon	Coupon	p-value
N. past orders	2.256	2.241	0.533
	(2.507)	(2.482)	
N. past items ordered	4.948	4.889	0.391
	(7.739)	(7.593)	
Days since first order (as of Jan 1, 2022)	626.235	625.136	0.830
	(554.098)	(549.147)	
Days since last order (as of Jan 1, 2022)	357.376	359.186	0.522
	(301.782)	(304.218)	
Total money spent on past orders	613.381	602.851	0.200
	(897.556)	(876.169)	
N. Participants	17,319	34,676	

Notes: This table reports summary statistics across all participants, by assignment to coupon. Standard deviations are reported in parentheses.

Table 3.2: Effects of coupon on revenue, profit, and likelihood of purchase

(a) Eligible products, during coupon eligibility

Dependent var.	(1) Rev.	(2) Rev.	(3) Profit	(4) Profit	(5) Purchase	(6) Purchase
Coupon	0.285**	0.292**	0.173**	0.178**	0.002***	0.002***
	(0.123)	(0.123)	(0.086)	(0.086)	(0.001)	(0.001)
Constant	0.830***	0.951***	0.585***	0.671***	0.005***	0.005***
	(0.097)	(0.184)	(0.068)	(0.126)	(0.001)	(0.001)
Controls	N	Y	N	Y	N	Y
N.	51995	51995	51995	51995	51995	51995

(b) Eligible products, after coupon eligibility

Dependent var.	(1) Rev.	(2) Rev.	(3) Profit	(4) Profit	(5) Purchase	(6) Purchase
Coupon	-0.336*	-0.320*	-0.215*	-0.204*	-0.002*	-0.002*
	(0.180)	(0.179)	(0.117)	(0.117)	(0.001)	(0.001)
Constant	1.962***	1.632***	1.269***	1.036***	0.012***	0.009***
	(0.151)	(0.266)	(0.099)	(0.173)	(0.001)	(0.002)
Controls	N	Y	N	Y	N	Y

(c) Ineligible products, during coupon eligibility

Dependent var.	(1) Rev.	(2) Rev.	(3) Profit	(4) Profit	(5) Purchase	(6) Purchase
Coupon	-0.481	-0.395	-0.300	-0.240	-0.001	-0.000
	(0.518)	(0.515)	(0.364)	(0.362)	(0.001)	(0.001)
Constant	6.672***	4.940***	4.630***	3.414***	0.023***	0.017***
	(0.430)	(0.923)	(0.302)	(0.645)	(0.001)	(0.002)
Controls	N	Y	N	Y	N	Y

(d) Ineligible products, after coupon eligibility

Dependent var.	(1) Rev.	(2) Rev.	(3) Profit	(4) Profit	(5) Purchase	(6) Purchase
Coupon	-0.836	-0.639	-0.455	-0.322	-0.004*	-0.003
	(0.873)	(0.859)	(0.578)	(0.568)	(0.002)	(0.002)
Constant	16.543***	10.586***	10.928***	6.949***	0.055***	0.037***
	(0.719)	(1.470)	(0.473)	(0.970)	(0.002)	(0.003)
Controls	N	Y	N	Y	N	Y

Notes: All models are estimated using OLS with heteroskedasticity-robust standard errors.

Table 3.3: The effects of coupon assignment about beliefs for the largest future discount

	(1)	(2)	(3)	(4)
Model	OLS	OLS	OLS	OLS
Dependent var.	Conf.	Conf.	Very conf.	Very conf.
Coupon	0.072*	0.068*	0.103**	0.115**
	(0.038)	(0.038)	(0.046)	(0.045)
Constant	0.739***	0.739***	0.466***	0.492***
	(0.027)	(0.027)	(0.032)	(0.032)
N.	477	477	477	477

Notes: Columns (1) and (2) present results for the largest discount customers are confident customers can receive in the next 100 days, while columns (3) and (4) present analogous results for the largest discount customers are *very confident* they can receive in the next 100 days. Columns (1) and (3) present results for the product eligible for the discount, while columns (2) and (4) present analogous results for the product ineligible for the coupon.

Table 3.4: Willingness to pay for company and competitor products by treatment assignment

	(1)	(2)	(3)
Model	OLS	OLS	OLS
Dependent var.	WTP Eligible	WTP Ineligible	WTP Competitor
Coupon	6.188	2.236	6.164
	(4.596)	(4.053)	(4.623)
Constant	121.092***	85.714***	106.807***
	(3.253)	(2.869)	(3.272)
N.	477	477	477

Notes: Column (1) presents results for the WTP question concerning the product eligible for the coupon. Column (2) presents the same for the product ineligible for the coupon while product (3) presents the results for the competitor product.

Conclusion:

In conclusion, this dissertation makes significant contributions to the literature by using field experiments to test and develop theories of consumer behavior. In Chapter One, a portable empirical methodology was developed to measure and monetize social image utility, which was deployed in experiments on exercise and charitable behavior. The results showed that public recognition motivates desirable behavior but creates highly unequal image payoffs. In Chapter Two, we studied how consumers react to sales taxes using an online shopping experiment, which revealed that consumers can both over-react or under-react to sales taxes, consistent with models of costly attention. Finally, in Chapter Three we explored the consequences of offering a one-time price discount to consumers and discovered that price discounts do not change the perceived value of the brand or product quality, and did not impact the company's profits or revenue. In all three chapters, field experiments were used to analyze human behavior in real-world settings, and the results provide valuable insights for firms and policymakers.

Bibliography

Acland, Dan and Matthew R Levy, "Naiveté, projection bias, and habit formation in gym attendance," *Management Science*, 2015, *61* (1), 146–160.

Ailawadi, Kusum L., Bari A. Harlam, Jacques Cesar, and David Trounce, "Practice Prize Report: Quantifying and Improving Promotion Effectiveness at CVS," *Marketing Science*, 2007, *26* (4), 566–575.

Alesina, Alberto and George-Marios Angeletos, "Fairness and Redistribution," *American Economic Review*, 2005, *95* (4), 960–980.

Ali, S. Nageeb and Roland Bénabou, "Image versus information: Changing societal norms and optimal privacy," *American Economic Journal: Microeconomics*, 2020, *12* (3), 116–164.

Allcott, Hunt and Judd B Kessler, "The welfare effects of nudges: A case study of energy use social comparisons," *American Economic Journal: Applied Economics*, 2019, *11* (1), 236–176. Forthcoming.

_ , Joshua Kim, Dmitry Taubinsky, and Jonathan Zinman, "Are High Interest Loans Predatory? Theory and Evidence from Payday Lending," *Review of Economic Studies*, forthcoming.

Ambuehl, Sandro, Axel Ockenfels, and Colin Stewart, "Who Opts In," *Rotman School of Management Working Paper*, 2018, (3154197).

Anagol, Santosh, Shawn Cole, and Shayak Sarkar, "Understanding the Advice of Commissions-Motivated Agents: Evidence from the Indian Life Insurance Market," *The Review of Economics and Statistics*, 2017, *99* (1), 1–15.

Anderson, Eric T. and Duncan I. Simester, "Are sale signs less effective when more products have them?," *Marketing Science*, 2001, *20* (2), 121–142. Cited by: 42; All Open Access, Green Open Access.

_ and _ , "Research note: Price discrimination as an adverse signal: Why an offer to spread payments may hurt demand," *Marketing Science*, 2001, *20* (3), 315–327. Cited by: 25.

_ and _ , "Long-Run Effects of Promotion Depth on New Versus Established Customers: Three Field Studies," *Marketing Science*, 2004, *23* (1), 4–20.

_ and _ , "PRICE STICKINESS AND CUSTOMER ANTAGONISM," *The Quarterly Journal of Economics*, 2010, *125* (2), 729–765.

Andreoni, James and B. Douglas Bernheim, "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects," *Econometrica*, 2009, *77* (5), 1607–1636.

— **and Ragan Petrie**, "Public goods experiments without confidentiality: a glimpse into fund-raising," *Journal of Public Economics*, 2004, *88* (7), 1605–1623.

Ariely, Dan, Anat Bracha, and Stephan Meier, "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially," *American Economic Review*, 2009, *99* (1), 544–555.

— **, Uri Gneezy, George Loewenstein, and Nina Mazar**, "Large Stakes and Big Mistakes," *Review of Economic Studies*, 2009, *76*, 451–469.

Ashraf, Nava, Oriana Bandiera, and B Kelsey Jack, "No margin, no mission? A field experiment on incentives for public service delivery," *Journal of Public Economics*, 2014, *120*, 1–17.

Augenblick, Ned, "Short-Term Discounting of Unpleasant Tasks," 2018. Working Paper.

Bartos, Vojtêch, Michal Bauer, Julie Chytilová, and Filip Matêjka, "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *American Economic Review*, June 2016, *106* (6), 1437–75.

Becker, Gary S, "Crime and punishment: An economic approach," in "The economic dimensions of crime," Springer, 1968, pp. 13–68.

Becker, Gary S., "A Note on Restaurant Pricing and Other Examples of Social Influences on Price," *Journal of Political Economy*, 1991, *99* (5), 1109–1116.

Bénabou, Roland and Jean Tirole, "Incentives and Prosocial Behavior," *American Economic Review*, 2006, *96* (5), 1652–1678.

— **and** — , "Laws and Norms," Working Paper 17579, National Bureau of Economic Research 2011.

Benartzi, Shlomo, John Beshears, Katherine L Milkman, Cass R Sunstein, Richard H Thaler, Maya Shankar, Will Tucker-Ray, William J Congdon, and Steven Galing, "Should governments invest more in nudging?," *Psychological Science*, 2017, *28* (8), 1041–1055.

Bernheim, B Douglas and Christine L Exley, "Understanding Conformity: An Experimental Investigation," Technical Report, Harvard Business School 2015.

Bernheim, B. Douglas and Dmitry Taubinsky, *Behavioral public economics*, Vol. 1, Elsevier, 2018.

— **and** — , "Behavioral Public Economics," in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *The Handbook of Behavioral Economics*, Vol. 1, New York: Elsevier, 2018, pp. 381–516.

— **, Andrey Fradkin, and Igor Popov**, "The welfare economics of default options in 401 (k) plans," *American Economic Review*, 2015, *105* (9), 2798–2837.

Besley, Timothy and Stephen Coate, "Workfare versus welfare incentive arguments for work requirements in poverty-alleviation programs," *American Economic Review*, 1992, *82* (1), 249–61.

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor, "Choose to Lose: Health Plan Choices from a Menu with Dominated Options," *Quarterly Journal of Economics*, 2017, *132* (3), 1319–1372.

Bhatia, Rajendra and Chandler Davis, "A Better Bound on the Variance," *American Mathematical Monthly*, April 2000, *107* (4), 353–357.

Bicchieri, Cristina, Eugen Dimant, Simon Gächter et al., "Observability, social proximity, and the erosion of norm compliance," 2020.

Birke, David J., "Anti-Bunching: A New Test for Signaling Motives in Prosocial Behavior," *working paper*, 2020.

Blomquist, N. Soren, "Interdependent behavior and the effect of taxes," *Journal of Public Economics*, 1993, *51* (2), 211–218.

Bloom, Nicholas and John Van Reenen, "Measuring and explaining management practices across firms and countries," *Quarterly Journal of Economics*, 2007, *122* (4), 1351—1408.

Bø, Erlend E, Joel Slemrod, and Thor O Thoresen, "Taxes on the internet: Deterrence effects of public disclosure," *American Economic Journal: Economic Policy*, 2015, *7* (1), 36–62.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Salience and Consumer Choice," *Journal of Political Economy*, 2013, *121*, 803–843.

_ , _ , and _ , "Memory, Attention, and Choice," *Quarterly Journal of Economics*, 2020, *135* (3), 1399–1442.

_ , _ , and _ , "Memory, Attention, and Choice*," *The Quarterly Journal of Economics*, 04 2020, *135* (3), 1399–1442.

Bradler, Christiane, Robert Dur, Susanne Neckermann, and Arjan Non, "Employee recognition and performance: A field experiment," *Management Science*, 2016, *62* (11), 3085–3099.

Bradley, Sebastien and Naomi E. Feldman, "Hidden Baggage: Behavioral Responses to Changes in Airline Ticket Tax Disclosure," *American Economic Journal: Economic Policy*, November 2020, *12* (4), 58–87.

Bursztyn, Leonardo and Robert Jensen, "How Does Peer Pressure Affect Educational Investments?," *The Quarterly Journal of Economics*, 2015, *130* (3), 1329–1367.

_ and _ , "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure," *Annual Review of Economics*, 2017, *9* (1), 131–153.

_ , Bruno Ferman, Stefano Fiorin, Martin Kanz, and Gautam Rao, "Status goods: experimental evidence from platinum credit cards," *Quarterly Journal of Economics*, 2017, *133*, 1561–1595.

_ , Georgy Egorov, and Robert Jensen, "Cool to be smart or smart to be cool? Understanding peer pressure in education," *The Review of Economic Studies*, 2019, *86* (4), 1487–1526.

— , **Thomas Fujiwara, and Amanda Pallais**, "'Acting Wife': Marriage Market Incentives and Labor Market Investments," *American Economic Review*, 2017, *107* (11), 3288–3319.

Bushong, Benjamin, Joshua Schwartzstein, and Matthew Rabin, "A Model of Relative Thinking," *The Review of Economic Studies*, 2021, *88*, 162–191.

Caplin, Andrew, "Measuring and Modeling Attention," *Annual Review of Economics*, 2016, *8* (1), 379–403.

— , **Dániel Csaba, John Leahy, and Oded Nov**, "Rational Inattention, Competitive Supply, and Psychometrics," *Quarterly Journal of Economics*, 2020, *135*, 1681–1724.

— , **Mark Dean, and John Leahy**, "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," *Review of Economic Studies*, 2019, *86* (3), 1061–1094.

Carrera, Mariana, Heather Royer, Mark Stehr, Justin Sydnor, and Dmitry Taubinsky, "Who Chooses Commitment? Evidence and Welfare Implications," *Review of Economic Studies*, forthcoming.

Carroll, Gabriel D, James J Choi, David Laibson, Brigitte C Madrian, and Andrew Metrick, "Optimal defaults and active decisions," *Quarterly Journal of Economics*, 2009, *124* (4), 1639–1674.

Carvalho, Leandro and Dan Silverman, "Complexity and Sophistication," Working Paper 26036, National Bureau of Economic Research July 2019.

Cederström, Carl and André Spicer, *The wellness syndrome*, John Wiley & Sons, 2015.

Chetty, Raj, Adam Looney, and Kory Kroft, "Salience and Taxation: Theory and Evidence," Working Paper 13330, National Bureau of Economic Research 2007.

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

Cohen, Adam Scott, Rie Chun, and Daniel Sznycer, "Do pride and shame track the evaluative psychology of audiences? Preregistered replications of Sznycer et al.(2016, 2017)," *Royal Society Open Science*, 2020, *7* (5), 191922.

Conlisk, John, Eitan Gerstner, and Joel Sobel, "Cyclic Pricing by a Durable Goods Monopolist," *The Quarterly Journal of Economics*, 1984, *99* (3), 489–505.

Conrad, Peter, "Wellness as virtue: Morality and the pursuit of health," *Culture, medicine and psychiatry*, 1994, *18* (3), 385–401.

Corneo, Giacomo G., "The Theory of the Open Shop Trade Union Reconsidered," *Labour Economics*, 1997, *4* (1), 71–84.

Courty, Pascal and Mario Pagliero, "Do Consumers Care about How Prices Are Set?," 2008, (ECO2008/03).

Damgaard, Mette Trier and Christina Gravert, "The hidden costs of nudging: Experimental evidence from reminders in fundraising," *Journal of Public Economics*, 2018, *157*, 15–26.

de Oliveira Santini, Fernando, Valter Afonso Vieira, Claudio Hoffmann Sampaio, and Marcelo Gattermann Perin, "Meta-Analysis of the Long- and Short-Term Effects of Sales Promotions on Consumer Behavior," *Journal of Promotion Management*, 2016, *22* (3), 425–442.

de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth, "Measuring and Bounding Experimenter Demand," *American Economic Review*, 2018, *108* (11), 3266–3302.

Dean, Mark and Nathaniel Neligh, "Experimental Tests of Rational Inattention," Working Paper 2019.

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

_ , "Structual Behavioral Economics," in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics*, number 1, Elsevier, 2018.

_ **and Devin G. Pope**, "What Motivates Worker Effort? Evidence and Expert Forecasts," *Review of Economic Studies*, 2018, *126* (6), 2410–2456.

_ **and Devin Pope**, "Stability of Experimental Results," *working paper*, 2019.

_ **and Ulrike Malmendier**, "Paying not to go to the gym," *American Economic Review*, 2006, *96* (3), 694–719.

_ **, John A List, and Ulrike Malmendier**, "Testing for altruism and social pressure in charitable giving," *Quarterly Journal of Economics*, 2012, *127* (1), 1–56.

_ **,** _ **,** _ **, and Gautam Rao**, "Voting to tell others," *The Review of Economic Studies*, 2017, *84* (1), 143–181.

_ **, John List, Ulrike Malmendier, and Gautam Rao**, "Estimating Social Preferences and Gift Exchange with a Piece-Rate Design," *working paper*, 2019.

DelVecchio, Devon, H. Shanker Krishnan, and Daniel C. Smith, "Cents or Percent? The Effects of Promotion Framing on Price Expectations and Choice," *Journal of Marketing*, 2007, *71* (3), 158–170.

Dholakia, Utpal M., "How Customer Self-Determination Influences Relational Marketing Outcomes: Evidence from Longitudinal Field Studies," *Journal of Marketing Research*, 2006, *43* (1), 109–120.

Edelman, Benjamin, Sonia Jaffe, and Scott Kominers, "To Groupon or Not to Groupon: The Profitability of Deep Discounts," *Marketing Letters*, 10 2011, *27.*

Ehrlich, Isaac, "Participation in illegitimate activities: A theoretical and empirical investigation," *Journal of Political Economy*, 1973, *81* (3), 521–565.

Enke, Benjamin and Florian Zimmermann, "Correlation Neglect in Belief Formation," *Review of Economic Studies*, 2019, *86* (1), 313–332.

Erdem, Tulin, Michael P. Keane, and Baohong Sun, "A Dynamic Model of Brand Choice When Price and Advertising Signal Product Quality," *Marketing Science*, 2008, *27* (6), 1111–1125.

Etzioni, Amitai, "Back to the Pillory?," *The American Scholar*, 1999, *68* (3), 43–50.

Exley, Christine, "Incentives for Prosocial Behavior: The Role of Reputations,," *Management Science*, 2018, *64* (5), 2460–2471.

Farhi, Emmanuel and Xavier Gabaix, "Optimal Taxation with Behavioral Agents," *American Economic Review*, January 2020, *110* (1), 298–336.

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

Feldman, Naomi, Jacob Goldin, and Tatiana Homonoff, "Raising the Stakes: Experimental Evidence on the Endogeneity of Taxpayer Mistakes," *National Tax Journal*, 2018, *71* (2), 201–230.

Finkelstein, Amy, "Welfare Analysis Meets Causal Inference: A Suggested Interpretation of Hendren," *working paper*, 2019.

Fudenberg, Drew, Philipp Strack, and Tomasz Strzalecki, "Speed, Accuracy, and the Optimal Timing of Choices," *American Economic Review*, 2018, *108* (12), 3651–3684.

Gabaix, Xavier, "A Sparsity Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 2014, *129* (4), 1661–1710.

_ , "Behavioral Inattention," in Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics*, Vol. 2, Elsevier, 2019, pp. 261–343.

_ **and David Laibson**, "Shrouded attributes, consumer myopia, and information suppression in competitive markets," *Quarterly Journal of Economics*, 2006, *121* (2), 505–540.

_ , _ , Guillermo Moloche, and Stephen Weinberg, "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model," *American Economic Review*, 2006, *96*, 1043–1068.

Gagnon-Bartsch, Tristan, Mathew Rabin, and Joshua Schwartzstein, "Channeled Attention and Stable Errors," Working Paper 2020.

Gauri, Varun, Julian C Jamison, Nina Mazar, Owen Ozier, Shomikho Raha, and Karima Saleh, "Motivating bureaucrats through social recognition: evidence from simultaneous field experiments," 2018.

Gerber, Alan S, Donald P Green, and Christopher W Larimer, "Social pressure and voter turnout: Evidence from a large-scale field experiment," *American Political Science Review*, 2008, *102* (1), 33–48.

Gillen, Ben, Erik Snowberg, and Leeat Yariv, "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study," *Journal of Political Economy*, 2019, *127* (4), 1826–1863.

Gneezy, Ayelet, Uri Gneezy, and Dominique Olie Lauga, "A Reference-Dependent Model of the Price-Quality Heuristic," *Journal of Marketing Research*, 2014, *51* (2), 153–164.

Gneezy, Uri and Aldo Rustichini, "Pay Enough or Don't Pay at All," *Quarterly Journal of Economics*, 2000, *115* (3), 791–810.

Goldin, Jacob and Tatiana Homonoff, "Smoke Gets in Your Eyes: Cigarette Tax Salience and Regressivity," *American Economic Journal: Economic Policy*, 2013, *5* (1), 302–336.

Gupta, Sunil, "Impact of Sales Promotions on when, what, and how Much to Buy," *Journal of Marketing Research*, 1988, *25* (4), 342–355.

Halpern, David, *Inside the nudge unit: How small changes can make a big difference*, Random House, 2015.

Handel, Benjamin R and Jonathan T Kolstad, "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare," *American Economic Review*, 2015, *105* (8), 2449–2500.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein, "Learning Through Noticing: Theory and Evidence from a Field Experiment," *Quarterly Journal of Economics*, 2014, *129* (3), 1311–1353.

Hastings, Justine S., Brigitte C. Madrian, and William L. Skimmyhorn, "Financial Literacy, Financial Education, and Economic Outcomes," *Annual Review of Economics*, 2013, *5*, 347–373.

Hausman, Jerry, "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problemsfrom the Left," *Journal of Economic Perspectives*, 2001, *15* (4), 57–67.

Heidhues, Paul, Botond Kőszegi, and Takeshi Murooka, "Inferior Products and Profitable Deception," *Review of Economic Studies*, 2017, *84* (1), 323–356.

Hendel, Igal and Aviv Nevo, "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica*, 2006, *74* (6), 1637–1673.

Hogg, Michael A., *The Social Psychology of Group Cohesiveness: From Attraction to Social Identity*, Harverster Wheatsheaf, 1992.

Hossain, Tanjim and John Morgan, "...Plus Shipping and Handling: Revenue (Non)Equivalence in Field Experiments on eBay," *The B.E. Journal of Economic Analysis & Policy*, 2006, *5* (2), 1–30.

Jones, Daniel and Sera Linardi, "Wallflowers: Experimental evidence of an aversion to standing out," *Management Science*, 2014, *60* (7), 1757–1771.

Kahan, Dan M and Eric A Posner, "Shaming white-collar criminals: A proposal for reform of the federal sentencing guidelines," *The Journal of Law and Economics*, 1999, *42* (S1), 365–392.

Kahneman, Daniel, *Attention and Effort*, Prentice-Hall, 1973.

_ , "Maps of Bounded Rationality: Psychology for Behavioral Economics," *American Economic Review*, 2003, *93* (5), 1449–1475.

Kaplow, Louis and Steven Shavell, "Moral Rules, the Moral Sentiments, and Behavior: Toward a Theory of an Optimal Moral System," *Journal of Political Economy*, 2007, *115* (3), 494–514.

Karing, Anne, "Social signaling and childhood immunization: A field experiment in Sierra Leone," Working Paper 2019.

Kass, Robert E. and Adrien E. Raftery, "Bayes Factors," *Journal of the American Statistical Association*, 1995, *90*, 773–795.

Kim, Jae-Young, "Large Sample Properties of Posterior Densities, Bayesian Information Criterion and the Likelihood Principle in Nonstationary Time Series Models," *Econometrica*, March 1998, *66* (2), 359–380.

Kosfeld, Michael and Susanne Neckermann, "Getting more work for nothing? Symbolic awards and worker performance," *American Economic Journal: Microeconomics*, 2011, *3* (3), 86–99.

_ , _ , and Xiaolan Yang, "The effects of financial and recognition incentives across work contexts: The role of meaning," *Economic Inquiry*, 2017, *55* (1), 237–247.

Koszegi, Botond and Adam Szeidl, "A Model of Focusing in Economic Choice," *Quarterly Journal of Economics*, 2013, *128*, 53–104.

_ and Paul Heidhues, "Regular prices and sales," *Theoretical Economics*, 2014, *9* (1).

Kroft, Kory, Jean-William P. Laliberte, Rene Leal-Vizcaino, and Matthew J. Notowidigdo, "Salience and Taxation with Imperfect Competition," Working Paper 27409, National Bureau of Economic Research 2020.

Lacetera, Nicola and Mario Macis, "Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme," *Journal of Economic Behavior & Organization*, 2010, *76* (2), 225–237.

_ , Devin G. Pope, and Justin R. Sydnor, "Heuristic Thinking and Limited Attention in the Car Market," *American Economic Review*, May 2012, *102* (5), 2206–36.

Leary, Mark R, "Motivational and emotional aspects of the self," *Annu. Rev. Psychol.*, 2007, *58*, 317–344.

Lieder, Falk and Thomas L. Griffiths, "Resource-Rational Analysis: Understanding Human Cognition as the Optimal Use of Limited Computational Resources," *Behavioral and Brain Sciences*, 02 2019, pp. 1–85.

Lindbeck, Assar, Sten Nyberg, and Jorgen Weibull, "Social Norms And Economic Incentives In The Welfare State," *Quarterly Journal of Economics*, 1999, *114*, 1–35.

_ , _ , and Jörgen Weibull, "Social Norms and Welfare State Dynamics," *Journal of the European Economic Association*, 2003, pp. 533–542.

Loewenstein, George, Cass R Sunstein, and Russell Golman, "Disclosure: Psychology changes everything," 2014, *6*, 391–419.

Maćkowiak, Bartosz, Filip Matêjka, and Mirko Wiederholt, "Rational Inattention: A Disciplined Behavioral Model," Working Paper May 2018.

Mailath, George J, "Incentive compatibility in signaling games with a continuum of types," *Econometrica: Journal of the Econometric Society*, 1987, pp. 1349–1365.

Malmendier, Ulrike and Stephen Nagel, "Learning from Inflation Experiences," *Quarterly Journal of Economics*, 2016, *131* (1), 53–87.

Martin, Daniel, "Rational Inattention in Games: Experimental Evidence," Working Paper 2016.

Massaro, Toni, "Shame, Culture, and American Criminal Law," *Michigan Law Review*, 1991, *89*, 1880–1942.

Matejka, Filip and Alisdair McKay, "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, January 2015, *105* (1), 272–298.

Mueller, Andreas I., Johannes Spinnewijn, and Giorgio Topa, "Job Seekers' Perceptions and Employment Prospects: Heterogeneity, Duration Dependence, and Bias," *American Economic Review*, January 2021, *111* (1), 324–63.

Mullainathan, Sendhil, "A Memory-Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 2002, *117* (3), 735–774.

Nakamura, Emi and Jon Steinsson, "Price setting in forward-looking customer markets," *Journal of Monetary Economics*, 2011, *58* (3), 220–233.

Narasimhan, Chakravarthi, "Competitive Promotional Strategies," *The Journal of Business*, 1988, *61* (4), 427–49.

Neckermann, Susanne and Xiaolan Yang, "Understanding the (unexpected) consequences of unexpected recognition," *Journal of Economic Behavior & Organization*, 2017, *135*, 131–142.

_ , Reto Cueni, and Bruno S Frey, "Awards at work," *Labour Economics*, 2014, *31*, 205–217.

Nussbaum, Martha C, *Hiding from humanity: Disgust, shame, and the law*, Princeton University Press, 2009.

OECD, *Behavioural Insights and Public Policy: Lessons from Around the World*, OECD Publishing, 2017.

Oyeniyi, Omotayo, "Sales promotion and consumer loyalty: A study of nigerian tecommunication industry," *Journal of Competitiveness/*, 2011, pp. 66–77.

Perez-Truglia, Ricardo and Guillermo Cruces, "Partisan interactions: Evidence from a field experiment in the united states," *Journal of Political Economy*, 2017, *125* (4), 1208–1243.

_ and Ugo Troiano, "Shaming Tax Delinquents: Theory and Evidence from a Field Experiment in the United States," *Journal of Public Economics*, 2018, *167*, 120–137.

Raftery, Adrien E., "Bayesian Model Selection in Social Research," *Sociological Methodology*, 1995, *25*, 111–163.

Ratcliff, Roger, "A Theory of Memory Retreival," *Psychological Review*, 1978, *85* (2), 59–108.

Rege, Mari and Kjetil Telle, "The impact of social approval and framing on cooperation in public good situations," *Journal of Public Economics*, 2004, *88* (7), 1625–1644.

Rotemberg, Julio J., "Customer anger at price increases, changes in the frequency of price adjustment and monetary policy," *Journal of Monetary Economics*, 2005, *52* (4), 829–852. SNB.

Schwartzstein, Joshua, "Selective Attention and Learning," *Journal of the European Economic Association*, 2014, *12* (6), 1423–1452.

Schwarz, Gideon, "Estimating the Dimension of a Model," *The Annals of Statistics*, 1978, *6*, 461–464.

Shenhav, Amitai, Sebastian Musslick, Falk Lieder, Wouter Kool, Thomas L. Griffiths, Jonathan D. Cohen, and Matthew M. Botvinick, "Toward a Rational and Mechanistic Account of Mental Effort," *Annual Review of Neuroscience*, 2017, *40* (1), 99–124.

Shu, Suzanne B. and Ayelet Gneezy, "Procrastination of Enjoyable Experiences," *Journal of Marketing Research*, 2010, *47* (5), 933–944.

Simester, D., "Chapter 11 - Field Experiments in Marketing," in Abhijit Vinayak Banerjee and Esther Duflo, eds., *Handbook of Field Experiments*, Vol. 1 of *Handbook of Economic Field Experiments*, North-Holland, 2017, pp. 465–497.

Sims, Christopher A., "Implications of Rational Inattention," *Journal of Monetary Economics*, 2003, *50* (3), 665 – 690.

Slade, Margaret E., "Sticky prices in a dynamic oligopoly: An investigation of (s,S) thresholds," *International Journal of Industrial Organization*, 1999, *17* (4), 477–511.

Soetevent, Adriaan R, "Anonymity in giving in a natural context: a field experiment in 30 churches," *Journal of Public Economics*, 2005, *89* (11), 2301–2323.

_ , "Payment choice, image motivation and contributions to charity: evidence from a field experiment," *American Economic Journal: Economic Policy*, 2011, *3* (1), 180–205.

Steel, Mark F. J., "Model Averaging and Its Use in Economics," *Journal of Economic Literature*, September 2020, *58* (3), 644–719.

Sznycer, Daniel, John Tooby, Leda Cosmides, Roni Porat, Shaul Shalvi, and Eran Halperin, "Shame closely tracks the threat of devaluation by others, even across cultures," *Proceedings of the National Academy of Sciences*, 2016, *113* (10), 2625–2630.

_ **, Laith Al-Shawaf, Yoella Bereby-Meyer, Oliver Scott Curry, Delphine De Smet, Elsa Ermer, Sangin Kim, Sunhwa Kim, Norman P Li, Maria Florencia Lopez Seal et al.**, "Cross-cultural regularities in the cognitive architecture of pride," *Proceedings of the National Academy of Sciences*, 2017, *114* (8), 1874–1879.

Tajfel, Henri, "Experiments in intergroup discrimination," *Scientific american*, 1970, *223* (5), 96–103.

Tangney, June Price, Jeff Stuewig, and Debra J Mashek, "Moral emotions and moral behavior," *Annu. Rev. Psychol.*, 2007, *58*, 345–372.

_ **, Rowland S Miller, Laura Flicker, and Deborah Hill Barlow**, "Are shame, guilt, and embarrassment distinct emotions?," *Journal of personality and social psychology*, 1996, *70* (6), 1256.

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

Thunstrom, Linda, "Welfare effects of nudges: The emotional tax of calorie menu labeling," *Judgment and Decision Making*, 2019, *14* (1), 11–25.

Whorton, James C, *Crusaders for fitness: The history of American health reformers*, Princeton University Press, 2014.

WorldatWork, "Trends in Employee Recognition," 2017.

Yaouanq, Yves Le and Peter Schwardmann, "Learning about one's self," 2019. Working Paper.

Yi, Youjae and Jaemee Yoo, "The long-term effects of sales promotions on brand attitude across monetary and non-monetary promotions," *Psychology & Marketing*, 2011, *28* (9), 879–896.

Yoeli, Erez, Moshe Hoffman, David G Rand, and Martin A Nowak, "Powering up with indirect reciprocity in a large-scale field experiment," *Proceedings of the National Academy of Sciences*, 2013, *110* (Supplement 2), 10424–10429.

Appendix A

Measuring the Welfare Effects of Shame and Pride

A.1 General Formulation of Social Signaling Models

We now consider more general public recognition structures. We let $\mathcal{A}$ denote the action space, which is a subset of $\mathbb{R}$, and we let F denote the distribution of types. We let $G(\sigma|a)$ denote the distribution of signals σ conditional on an individual choosing action a. For example, two-tier schemes that recognize people who chose $a \geq a^\dagger$ can be represented as schemes where $\sigma = 1$ if $a \geq a^\dagger$ and $\sigma = 0$ otherwise. Schemes where people's performance is revealed with some probability q can be represented as $\sigma = a$ with probability q and $\sigma = \emptyset$ with probability $1 - q$. The signals are completely uninformative if $G(\sigma|a)$ does not depend on a.

We consider general formulations of the action-signaling and characteristics-signaling models that have the following three features. First, in a fully-revealing equilibrium, these models correspond to the models we introduced in Section 2.2 and, in particular, can be consistent with any non-negative value of ρ. Second, these models make the sensible prediction that when nothing is revealed about an individual's action and type, then the individual derives zero utility from public recognition. Third, individuals' utility from public recognition is continuous in the audience inference, and is continuous in the population distribution of behavior or types (in the weak topology).

To see how the second criterion can be limiting, suppose that for general signal structures, individuals' utility from public recognition is given by $\nu S(\mathbb{E}[\theta|\sigma] - \rho\bar{\theta})$, where $\mathbb{E}[\theta|\sigma]$ denotes the audience's expectation of the individual's action, and $\rho > 1$. If the signal is fully revealing, then this formulation is consistent with the signaling model we presented in Section 2.2. However, if the signals are completely uninformative—meaning that nothing is in fact learned about the individual's behavior and type—then this formulation makes the odd prediction that the individual's utility from public recognition is $\nu S(\bar{\theta} - \rho\bar{\theta}) < S(0) = 0$; that is, that the individual derives negative utility from public recognition when in fact nothing is learned about the individual.

To see how the third criterion can be limiting, consider a public recognition scheme that divides individuals into K tiers $[0, a_1), [a_1, a_2), ..., [a_{K-1}, a_K]$, and that in equilibrium the mean type in each tier is $\bar{\theta}_1, \bar{\theta}_2, \ldots, \bar{\theta}_K$. Suppose that individuals' utility is given by $\nu S(\mathbb{E}[\theta|\sigma] - r)$, where r is the largest value such that $Pr(\bar{\theta}_i \leq r) \leq 1/2$. In a separating equilibrium—where each tier in fact corresponds to a possible value of a—this corresponds to the intuitive-sounding formulation in which individuals compare their type to the median type. Note, however, that it is crucial to define r in terms of the tiers, rather than in terms of the underlying distribution of types: if r was always defined as the median of the distribution of θ, and if the mean of the distribution of θ was smaller than the median, then with a completely uninformative signal structure individuals would derive $\nu S(\bar{\theta} - r) < S(0) = 0$ utility from public recognition. The problem with defining r as the median of the tiers is that it leads to discontinuous payoffs from public recognition. For example, consider a two-tier system. If for $\epsilon > 0$, $0.5 + \epsilon$ individuals are in the bottom tier, then r would be defined as the average type in the bottom tier. But if $0.5 - \epsilon$ individuals are in the bottom tier, then r would be defined as the average type in the top tier. This would lead payoffs from public recognition to be sharply discontinuous in the distribution of types in the population, which is not only unintuitive, but also theoretically unattractive as it could lead to non-existence of (pure strategy) equilibria even with convex type spaces.

To satisfy the second and third criteria, we define the reference point against which the audience inference is compared to be a weighted average of the distribution of audience posteriors induced by the equilibrium distribution of behavior. E.g., in the context of the example above, the reference point would be the weighted average of $\bar{\theta}_j$—the mean type in each tier. This implies that when signals are completely uninformative, so that the distribution of audience posteriors places weight 1 on the average type, the reference point is just the average behavior or type in the population. Plainly, the weighted-average function is also a continuous function of the distribution of posteriors, and thus satisfies the third criterion.

A.1.1 Action Signaling

We let $\mathbb{E}[a|\sigma]$ denote the audience's expectation of the individual's action, given a realization σ of the signal. Let $\boldsymbol{a} : \Theta \to \mathcal{A}$ be the equilibrium action function, and let $G^*(\sigma)$ denote the unconditional distribution of signal values, induced by $\boldsymbol{a}$, F, and $G(\cdot|a)$, that results in equilibrium. We assume that the audience updates according to Bayes' Rule to form the inference $\mathbb{E}[a|\sigma]$, and we let H^* denote the unconditional distribution of audience posteriors, $\mathbb{E}[a|\sigma]$, induced by the distribution G^*.

To illustrate H^*, consider a public recognition scheme that divides individuals into K tiers $[a_0 = 0, a_1), [a_1, a_2), ..., [a_{K-1}, a_K]$. Suppose that in equilibrium, the mean action in each tier is $\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_K$, and that the fraction of people in tier $[a_{k-1}, a_k)$ is μ_k. Then H^* is simply the probability distribution that places weight μ_k on $\bar{a}_k$.

We define utility from public recognition, for an individual generating signal σ, to be

$$\nu S\left(\mathbb{E}[a|\sigma] - \frac{\int_{a\in\mathcal{A}} aw(a)dH^*(a)}{\int_{a\in\mathcal{A}} w(a)dH^*(a)}\right)$$

where ν is the visibility parameter, the weighting function w is a smooth function $w : \mathbb{R} \to \mathbb{R}$, and where S is a smooth function with $S(0) = 0$. The equilibrium action function is such that $\boldsymbol{a}(\theta) \in \mathcal{A}$ maximizes

$$u(a;\theta) + \nu \int_\sigma S\left(\mathbb{E}[a|\sigma] - \frac{\int_{a\in\mathcal{A}} aw(a)dH^*(a)}{\int_{a\in\mathcal{A}} w(a)dH^*(a)}\right) dG(\sigma|a).$$

for each θ, given the Bayesian inference function $\mathbb{E}[a|\sigma]$ and the induced distribution H^*.

Note that when the signals are completely uninformative, $\mathbb{E}[a|\sigma]$ is simply the average action in the population, $\bar{a}$, and H^* places mass 1 on $\bar{a}$. Thus,

$$\mathbb{E}[a|\sigma] - \frac{\int_{a\in\mathcal{A}} aw(a)dH^*(a)}{\int_{a\in\mathcal{A}} w(a)dH^*(a)} = \bar{a} - \bar{a} = 0$$

and individuals derive no utility from public recognition. Conversely, when the signals are fully informative, public recognition utility is given by

$$\nu S\left(a - \frac{\int_{a\in\mathcal{A}} aw(a)dH(a)}{\int_{a\in\mathcal{A}} w(a)dH(a)}\right)$$

where H is the probability distribution over actions. Note that

$$\frac{\int_{a\in\mathcal{A}} aw(a)dH(a)}{\int_{a\in\mathcal{A}} w(a)dH(a)}$$

is simply the weighted average of the population distribution of performance, and is equal to $\rho\bar{a}$ for an appropriately defined constant ρ. If $w(a)$ is constant in a, meaning that there is no reweighting, then $\rho = 1$ in all separating equilibria. If $w(a)$ is increasing (decreasing) in a, meaning that higher levels of performance receive more (less) weight, then $\rho > 1$ ($\rho < 1$) in all separating equilibria. If $w(a)$ places full weight on $a = 0$ (and some individuals choose $a = 0$ in equilibrium), then $\rho = 0$ in all equilibria.

A.1.2 Characteristics Signaling

We define this general version of characteristics-signaling models analogously to above.

We let $\mathbb{E}[\theta|\sigma]$ denote the audience's expectation of the individual's action, given a realization σ of the signal. Let $\boldsymbol{a} : \Theta \to \mathcal{A}$ be the equilibrium action function, and let $G^*(\sigma)$ denote the unconditional distribution of signal values, induced by $\boldsymbol{a}$, F, and $G(\cdot|a)$, that results in equilibrium. We assume that the audience updates according to Bayes' Rule to form the inference $\mathbb{E}[\theta|\sigma]$, and we let H^* denote the unconditional distribution of audience posteriors, $\mathbb{E}[\theta|\sigma]$, induced by the distribution G^*.

To illustrate H^*, consider a public recognition scheme that divides individuals' performance into K tiers $[0, a_1), [a_1, a_2), ..., [a_{K-1}, a_K]$. Suppose that in equilibrium, the mean type in each tier is $\bar{\theta}_1, \bar{\theta}_2, \ldots, \bar{\theta}_K$, and that the fraction of people in tier $[a_{k-1}, a_k)$ is μ_k. Then H^* is simply the probability distribution that places weight μ_k on $\bar{\theta}_k$.

We define utility from public recognition, for an individual generating signal σ, to be

$$\nu S\left(\mathbb{E}[\theta|\sigma] - \frac{\int_{x\in\Theta} xw(x)dH^*(x)}{\int_{x\in\Theta} w(x)dH^*(x)}\right)$$

where the weighting function w is a smooth function $w : \mathbb{R} \to \mathbb{R}$, and where S is a smooth function with $S(0) = 0$. The equilibrium action function is such that $\boldsymbol{a}(\theta) \in \mathcal{A}$ maximizes

$$u(a;\theta) + \nu \int_\sigma S\left(\mathbb{E}[\theta|\sigma] - \frac{\int_{x\in\Theta} xw(x)dH^*(x)}{\int_{x\in\Theta} w(x)dH^*(x)}\right) dG(\sigma|a).$$

for each θ, given the Bayesian inference function $\mathbb{E}[a|\sigma]$ and the induced distribution H^*.

Note that when the signals are completely uninformative, $\mathbb{E}[\theta|\sigma]$ is simply the average type in the population, $\bar{\theta}$, and H^* places mass 1 on $\bar{\theta}$. Thus,

$$\mathbb{E}[\theta|\sigma] - \frac{\int_{x\in\Theta} xw(x)dH^*(x)}{\int_{x\in\Theta} w(x)dH^*(x)} = \bar{\theta} - \bar{\theta} = 0$$

and individuals derive no utility from public recognition. Conversely, in a separating equilibrium, public recognition utility is given by

$$\nu S\left(\mathbb{E}[\theta|a] - \frac{\int_{x\in\Theta} xw(x)dF(x)}{\int_{x\in\Theta} w(x)dF(x)}\right)$$

where F is the probability distribution over types. Note that

$$\frac{\int_{x\in\Theta} xw(x)dF(x)}{\int_{x\in\Theta} w(x)dF(x)}$$

is simply the weighted average of the distribution of types, and is equal to $\rho\bar{\theta}$ for an appropriately defined constant ρ. If $w(\theta)$ is constant in θ, meaning that there is no reweighting, then $\rho = 1$ in all separating equilibria. If $w(\theta)$ is increasing (decreasing) in θ, meaning that higher levels of performance receive more (less) weight, then $\rho > 1$ ($\rho < 1$) in all separating equilibria. If $w(\theta)$ places full weight on some lowest type θ_m, then $\rho = \theta_m/\bar{\theta}$ in all equilibria.

A.1.3 The Net Image Payoff

For the sake of parsimony, we focus on the characteristics-signaling model, as the arguments for the action-signaling model are nearly identical.

We establish the following simple result: Assume that S is increasing. If S is concave and w is increasing, then the net image payoff is negative. If S is convex and w is decreasing, then the image payoff is positive. Suppose that S is concave and that w is increasing. Then Jensen's inequality implies that

$$\begin{aligned}
&\int_{\theta'\in\Theta}\int_{\sigma} S\left(\mathbb{E}[\theta|\sigma] - \frac{\int_{x\in\Theta} xw(x)dH^*(x)}{\int_{x\in\Theta} w(x)dH^*(x)}\right) dG(\sigma|\boldsymbol{a}(\theta'))dF(\theta') \\
\leq &S\left(\mathbb{E}[\theta|\sigma]dG(\sigma|\boldsymbol{a}(\theta'))dF(\theta') - \frac{\int_{x\in\Theta} xw(x)dH^*(x)}{\int_{x\in\Theta} w(x)dH^*(x)}\right) \\
= &S\left(\int_{x\in\Theta} xdH^*(x) - \frac{\int_{x\in\Theta} xw(x)dH^*(x)}{\int_{x\in\Theta} w(x)dH^*(x)}\right) && \text{(A.1)} \\
\leq &S(0) = 0. && \text{(A.2)}
\end{aligned}$$

Line (A.2) follows from line (A.1) because S is increasing and $Cov_{H^*}[x, w(x)] > 0$ by assumption.

The case in which S is convex and w is decreasing follows analogously.

A.2 Deadweight Loss Relative to Financial Incentives

A.2.1 Unidimensional Heterogeneity

Suppose first that types are one-dimensional, meaning that the type space Θ is a subset of $\mathbb{R}$. Assume also that all individuals share the same structural PRU S. In any equilibrium, possibly not fully separating, let $R : \mathcal{A} \to \mathbb{R}$ denote the resulting reduced-form PRU. Thus, individuals choose a to maximize $u(a;\theta) + R(a) + y$, where y is numeraire consumption. We let $\boldsymbol{a}(\theta)$ denote individuals' choices.

We can construct a revenue-neutral financial incentive scheme that induces exactly the same decisions $\boldsymbol{a}(\theta)$ as follows. Revenue-neutrality could be obtained in the YMCA setting, for example, by giving individuals a per-attendance incentive, and raising money for that by increasing the membership fees. Let $p(a)$ be the financial reward that individuals receive for choosing action a, and set $p(a) = R(a) - \int_{\theta\in\Theta} R(\boldsymbol{a}(\theta))dF(\theta)$, where F is the distribution over types θ. By construction, $\boldsymbol{a}(\theta)$ maximizes $u(a;\theta)+p(a)+y$, and $\int_{\theta\in\Theta} p(\boldsymbol{a}(\theta))dF(\theta) = 0$.

Plainly, every individual will be better (worse) off under the revenue-neutral financial incentive scheme if $\int_{\theta\in\Theta} R(\boldsymbol{a}(\theta))dF(\theta)$ is negative (positive). In other words, if the net image payoff from public recognition is negative, then every individual will be made better off if the public recognition intervention is instead replaced by the revenue-neutral financial incentive scheme $p(a)$. The difference in each individuals' utility will be $-\int_{\theta\in\Theta} R(\boldsymbol{a}(\theta))dF(\theta)$. We thus refer to $-\int_{\theta\in\Theta} R(\boldsymbol{a}(\theta))dF(\theta)$ as the deadweight loss of public recognition relative to financial incentives. Note that if the image payoff from public recognition are on net positive ($\int_{\theta\in\Theta} R(\boldsymbol{a}(\theta))dF(\theta) > 0$), then welfare with public recognition is higher than with the equivalent revenue-neutral financial incentive scheme.

A.2.2 Costly Public Funds and Constraints on the Sign of the Incentive Scheme

Above, we assumed that it is possible to use a revenue-neutral incentive scheme. In the YMCA context, this revenue-neutral scheme could involve raising monthly or annual membership fees to finance a per-attendance incentive. However, this may not always be possible. In such cases, the relative benefits of public recognition versus financial incentives are more nuanced where there is a shadow cost of public funds.

In particular, let the marginal value of public funds be $1+\lambda$, where $\lambda \geq 0$ is the shadow cost of raising funds due to distortionary effects. When $\lambda > 0$, financial incentives are particularly attractive relative to public recognition if they can be implemented as additional taxes or fines, since doing so raises government revenue. Examples include taxing behaviors that generate environmental externalities (e.g., energy use), or fining behaviors that violate the law (e.g., tax delinquency). However, there are other cases where financial incentives most naturally take the form of positive rewards, such as incentivizing charitable behavior by making it tax-deductible. In these cases there is an additional cost to using financial incentives in lieu of public recognition.

Formally, consider a non-revenue-neutral financial incentive scheme $p(a) = p_0 + R(a)$ that induces the same behavior change as does public recognition. Under public recognition, the net image payoff experienced by individuals is, as before, $\int_{\theta\in\Theta} R(\boldsymbol{a}(\theta))dF(\theta)$. Under the incentive scheme, individuals' earnings change by $\bar{p} = \int_{\theta\in\Theta} p(\boldsymbol{a}(\theta))dF(\theta)$ in total, and the cost to the government is $\lambda\bar{p}$. Thus, the net advantage of financial incentives versus public recognition is given by

$$(1-\lambda)\bar{p} - \int_{\theta\in\Theta} R(\boldsymbol{a}(\theta))dF(\theta).$$

When $\bar{p}$ is negative, meaning that on net the planner collects revenue, financial incentives are particularly attractive. When $\bar{p}$ is positive, meaning that on net the planner gives out financial rewards, financial incentives are less attractive. But when $\lambda = 1$ or when the incentive scheme is revenue-neutral, the relative advantage of financial incentives over public recognition is simply given by $-\int_{\theta\in\Theta} R(\boldsymbol{a}(\theta))dF(\theta)$, the net image payoff.

As an example, suppose that $p(a)$ is required to be non-negative, and return to the welfare estimate in column (1) of Table 1.9a, where the net image payoff was found to be -3.41. Assume also that the predicted 1.75 attendance change could be obtained with a \$1 per attendance financial incentive, as implied by participants' forecasts. For the social costs of a \$1 per attendance subsidy to be higher than the costs of using public recognition, the cost of public funds would need to be approximately $\lambda = 0.7$, which is substantially higher than the typical estimate of 0.3 (Finkelstein, 2019).[1]

A.2.3 Multidimensional Heterogeneity

We now consider the case where types θ are multidimensional because, for example, individuals have varying sensitivities to public recognition. For each individual of type θ, let $\Delta(\theta)$ denote the behavior change induced by public recognition, and let $e(\theta)$ denote the marginal social value of increasing type θ's choice of a. Let $r(\theta)$ denote each individual's realization of public recognition utility, and let $\bar{r} = \int_{\theta\in\Theta} r(\theta)dF(\theta)$ denote the net image payoff. In the one-dimensional case, $r(\theta) = R(\boldsymbol{a}(\theta))$. The total behavior change is given by $\bar{\Delta} = \int_{\theta\in\Theta} \Delta(\theta)dF(\theta)$, and the average marginal benefit of increasing a is $\bar{e} = \int_{\theta\in\Theta} e(\theta)dF(\theta)$. The incremental welfare effect of public recognition is given by

$$\begin{aligned} \Delta W^R &= \int_{\theta\in\Theta} \left(\Delta(\theta)e(\theta) + r(\theta)\right) dF(\theta) \\ &= \bar{\Delta}\bar{e} + \bar{r} + Cov[\Delta(\theta), e(\theta)]. \end{aligned} \tag{A.3}$$

Consider now an incentive scheme $p(a)$ that changes each type θ's behavior by $\Delta_p(\theta)$, such that $\int_{\theta\in\Theta} \Delta_p(\theta)dF(\theta) = \bar{\Delta}$. Let $\bar{p} = \int_{\theta\in\Theta} p(\boldsymbol{a}(\theta))dF(\theta)$ denote the net financial transfer to individuals. The incremental effect of these financial incentives is given by

$$\begin{aligned} \Delta W^p &= \int_{\theta\in\Theta} \left(\Delta_p(\theta)e(\theta) + p(\boldsymbol{a}(\theta))\right) dF(\theta) - \lambda \int_{\theta\in\Theta} p(\boldsymbol{a}(\theta))dF(\theta) \\ &= \bar{\Delta}\bar{e} + Cov[\Delta_p(\theta), e(\theta)] + (1-\lambda)\bar{p}. \end{aligned} \tag{A.4}$$

Equations (A.3) and (A.4) imply that the difference between the welfare effect of public recognition and financial incentives is given by

$$\underbrace{-\bar{r}}_{\text{image payoff}} + \underbrace{Cov[(\Delta_p(\theta) - \Delta(\theta), e(\theta)]}_{\text{relative targeting}} + \underbrace{(1-\lambda)\bar{p}}_{\text{cost of public funds}}. \tag{A.5}$$

Equation (A.5) shows that in addition to the image payoff, two other terms determine the welfare effects of financial incentives versus public recognition. The relative targeting term depends on the extent to which the two policy instruments affect the behavior of individuals whose behavior change generates the highest social benefits. This term can be nonzero if individuals' sensitivity to public recognition is, e.g., more correlated with $e(\theta)$ than their responsiveness to financial incentives. In the case where the benefits of behavior change are due to environmental, health, or fiscal externalities—such as energy

[1] A 1.75 attendance increase would lead to average attendance of $3.14 + 1.75 = 4.89$, and thus to generate a per-person social cost of \$3.41, the cost of public founds would need to be $3.41/4.89 \approx 0.7$.

consumption, vaccinations, or tax delinquency—it is reasonable that $e(\theta)$ is either constant, or at least uncorrelated with $\Delta_p(\theta)$ and $\Delta(\theta)$. In this case, the relative targeting term drops out. In other cases, where the need for behavior change arises from "internalities" such as individuals not attending their health club enough due to self-control problems, $e(\theta)$ is likely to be heterogeneous and could in principle be correlated with incentive effects. However, it is not obvious why $e(\theta)$ would be differentially correlated with responsiveness to financial incentives versus public recognition.

The last term, the impact on the costs of public funds, is discussed above in A.2.2. This term is zero when the incentive-scheme is revenue-neutral, or when $\lambda = 1$. As we discussed, there are also some natural cases where financial incentives in the form of taxes and fines are clearly doubly beneficial because they create additional revenue, but there are also other cases where financial incentives most naturally take the form of subsidies that must be financed by distortionary taxation.

A.3 Supplementary Empirical Results for YMCA Experiment

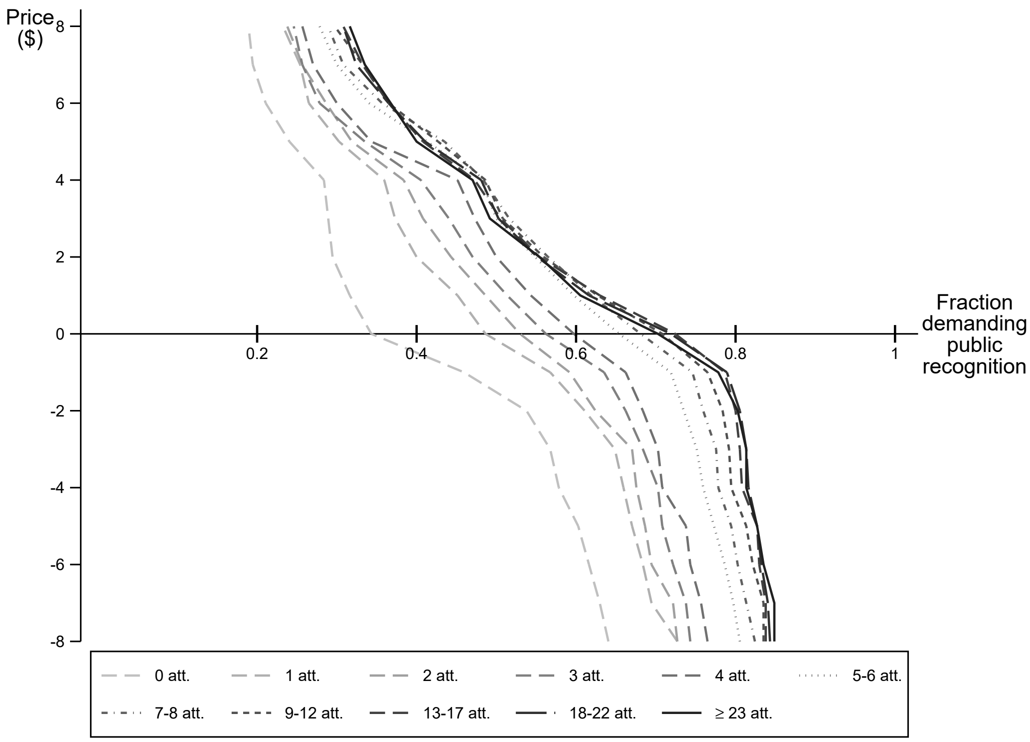
A.3.1 Demand for Public Recognition

A.3.2 Actual Versus Forecasted Attendance

A.3.3 Additional Results about the PRU and Past Attendance

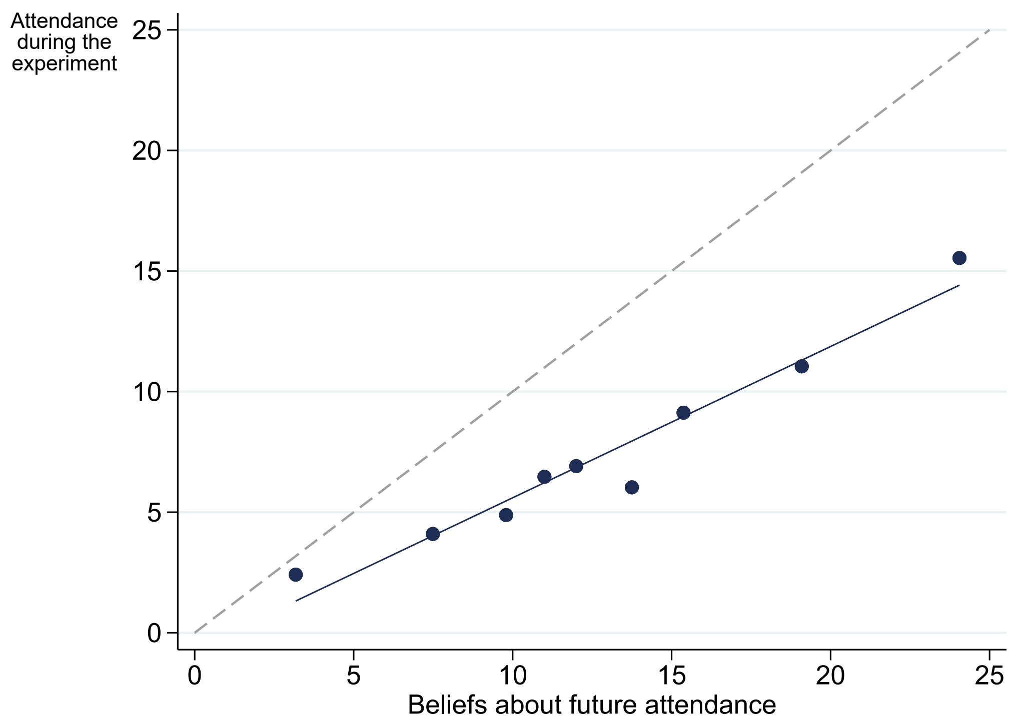
The first table shows that there is no significant interaction between past attendance and the PRU. The second table is analogous to Table 1.4, but considers visits within 4 of past attendance, rather than expectations.

Figure A.3.1: Demand Curves for Public Recognition



Notes: This figure plots the demand curves for public recognition by attendance interval. The analysis excludes 15 participants with "incoherent" preferences for public recognition.

Figure A.3.2: Actual versus forecasted attendance in the YMCA experiment



Notes: This figure plots the relationship between participants' forecasted and actual attendance. For participants in the public recognition group, we compare attendance to their beliefs about attendance if they are randomized into the public recognition group. For participants not in the public recognition group, we compare attendance to their beliefs about attendance if they are randomized to not be in the public recognition group. The analysis excludes 15 participants with "incoherent" preferences for public recognition.

Table A.3.1: WTP for public recognition by YMCA attendance: heterogeneity along average past attendance

	(1)	(2)	(3)	(4)	(5)	(6)
Model	OLS	OLS	OLS	Tobit	Tobit	Tobit
Dependent var.	WTP	WTP	WTP	WTP	WTP	WTP
N. visits	0.29***	0.43***	0.39***	0.52***	0.72***	0.68***
	(0.05)	(0.05)	(0.05)	(0.10)	(0.10)	(0.10)
N. visits sq.	-0.01***	-0.01***	-0.01***	-0.01***	-0.02***	-0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Avg. past att.			0.02			0.02
			(0.06)			(0.11)
N. visits × Past att.			-0.01			-0.01
			(0.01)			(0.01)
N. visits sq. × Past att.			0.00			0.00
			(0.00)			(0.00)
Constant	-0.19	-0.95**	-0.70	-0.86	-1.81**	-1.48*
	(0.46)	(0.44)	(0.45)	(0.94)	(0.83)	(0.87)
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})$	0.068	0.070	–	0.065	0.069	–
95% CI	[0.059, 0.077]	[0.063, 0.077]	–	[0.055, 0.076]	[0.061, 0.076]	–
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})\times$ SD	0.332	0.340	–	0.318	0.335	–
95% CI	[0.288, 0.376]	[0.306, 0.375]	–	[0.269, 0.368]	[0.299, 0.372]	–
Restriction	Coh	Coh	Coh	Coh	Coh	Coh
Observations	Above med.	Below med.	–	Above med.	Below med.	–
N. Subjects	2035	2035	4070	2035	2035	4070
N_clust	185	185	370	185	185	370

Notes: This table reports regression estimates from linear and quadratic models of willingness to pay for public recognition by attendance. This analysis excludes 15 participants with "incoherent" preferences for public recognition. Standard errors are clustered at the participant level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.3.2: WTP for public recognition by YMCA attendance: using number of visits within 4 of past attendance

	(1)	(2)	(3)	(4)
Model	OLS	OLS	Tobit	Tobit
Dependent var.	WTP	WTP	WTP	WTP
N. visits	0.23***	0.52***	0.41***	0.88***
	(0.05)	(0.11)	(0.10)	(0.22)
N. visits sq.		-0.02***		-0.03***
		(0.01)		(0.01)
Constant	-0.40	-1.01**	-0.98	-1.98**
	(0.42)	(0.47)	(0.84)	(0.93)
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})$	–	0.085	–	0.083
95% CI	–	[0.051, 0.118]	–	[0.043, 0.124]
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})\times$ SD	–	0.412	–	0.406
95% CI	–	[0.250, 0.575]	–	[0.209, 0.603]
Observations	1645	1645	1645	1645
N. Subjects	370	370	370	370

Notes: These tables report regression estimates from linear and quadratic models of willingness to pay for public recognition by attendance, restricting to intervals with a midpoint within 4 visits of a participant's average past attendance. The standard deviation of the difference between average past attendance and attendance during the month of the experiment is 4.51for the monotonic sample control group, 4.42 for the coherent sample control group, and 3.19 for the general YOTA population. Measures of the curvature of the estimated reduced-form public recognition function are $-R''_{exp}/R'_{exp}(\bar{a}_{pop})$ and $-R''_{exp}/R'_{exp}(\bar{a}_{pop}) \times SD$, where $\bar{a}_{pop}$ and $SD = 4.86$ are the average attendance and standard deviation of attendance for the general YOTA population, respectively. This analysis excludes 15 participants with "incoherent" preferences for public recognition. Standard errors are clustered at the participant level and reported in parentheses. 95 percent confidence intervals for the curvature statistics are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

A.3.4 Excluding High Visits Intervals

A.3.5 Rescaling the Visits Intervals to Have Equal Width

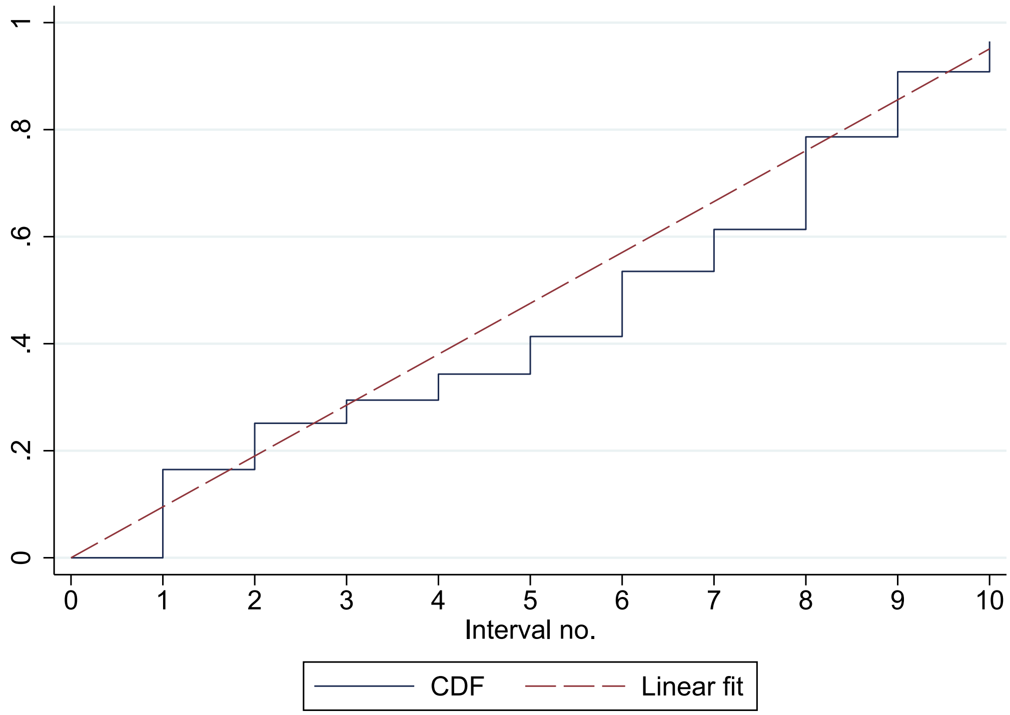
One potential concern with the intervals used in the YMCA intervals chosen is that participants might generate a WTP profile that changes by the same amount with each successive interval, either because of confusion or perceived experimenter demand. This may bias the results to overestimate concavity PRU; for example, if participants had a PRU that is linear

Table A.3.3: WTP for public recognition by YMCA attendance, excluding high number of visits questions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model	OLS	OLS	Tobit	Tobit	OLS	OLS	Tobit	Tobit
Dependent var.	WTP	WTP	WTP	WTP	WTP	WTP	WTP	WTP
N. visits	0.15***	0.47***	0.26***	0.81***	0.21***	0.57***	0.37***	0.99***
	(0.02)	(0.05)	(0.04)	(0.09)	(0.02)	(0.06)	(0.05)	(0.12)
N. visits sq.		-0.02***		-0.03***		-0.03***		-0.04***
		(0.00)		(0.00)		(0.00)		(0.01)
Constant	-0.01	-0.79**	-0.39	-1.74***	-0.24	-0.96***	-0.81	-2.04***
	(0.31)	(0.32)	(0.60)	(0.64)	(0.31)	(0.32)	(0.62)	(0.65)
$-R''/R'(\bar{a}_{pop})$	–	0.092	–	0.090	–	0.120	–	0.118
95% CI	–	[0.084, 0.100]	–	[0.081, 0.099]	–	[0.106, 0.135]	–	[0.101, 0.134]
$-R''/R'(\bar{a}_{pop}) \times$ SD	–	0.446	–	0.439	–	0.586	–	0.573
95% CI	–	[0.406, 0.486]	–	[0.395, 0.482]	–	[0.514, 0.657]	–	[0.493, 0.653]
Excl. int.	Coh	Coh	Coh	Coh	Coh	Coh	Coh	Coh
Observations	Top	Top	Top	Top	Top 2	Top 2	Top 2	Top 2
N. Subjects	3700	3700	3700	3700	3330	3330	3330	3330
N_clust	370	370	370	370	370	370	370	370

Notes: This table reports regression estimates from linear and quadratic models of willingness to pay for public recognition by attendance. Columns (1)-(4) exclude data from the top interval (23 or more attendances) while columns (5)-(8) exclude data from the top two intervals (18 or more attendances). The fraction of the sample who predicted 18 or more attendances is 0.26, and the fraction who predicted 23 or more attendances is 0.10. Measures of the curvature of the estimated reduced-form public recognition function are $-R''_{exp}/R'_{exp}(\bar{a}_{pop})$ and $-R''_{exp}/R'_{exp}(\bar{a}_{pop}) \times SD$, where $\bar{a}_{pop}$ and $SD = 4.86$ are the average attendance and standard deviation of attendance for the general YOTA population, respectively. This analysis excludes 15 participants with "incoherent" preferences for public recognition. Standard errors are clustered at the participant level and reported in parentheses. 95 percent confidence intervals for the curvature statistics are computed using the delta method. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Figure A.3.3: Distribution of Grow & Thrive attendance over elicitation intervals



Notes: This figure plots the cumulative distribution function for the fraction of participants with attendance below the minimum of each interval of attendance used in the WTP elicitation. Interval number takes values from 0 to 10, corresponding to the 11 intervals of future attendance. The analysis excludes 15 participants with "incoherent" preferences for public recognition.

in the index of the interval, it would look concave plotted against the midpoints of intervals that are increasing in length.

In this section, we provide evidence against this potential confound. First, Figure A.3.3 shows that the cumulative distribution function of attendance during Grow & Thrive is approximately linear in the attendance interval number. Thus, the intervals that included a wider range of visits did not actually include a larger share of realized attendance values. Second, Tables A.3.4 and A.3.5 show that the PRU is still estimated to be highly concave when we index intervals not by their midpoint, but instead by their sequential order. Moreover, our estimate of curvature is, if anything, slightly higher with respect to this recoding. This suggests that our results about concavity are not driven by participants trying to generate a WTP profile that is linearly increasing in the interval numbers.

Table A.3.4: WTP for public recognition by index of attendance interval

	(1)	(2)	(3)	(4)
Model	OLS	OLS	Tobit	Tobit
Dependent var.	WTP	WTP	WTP	WTP
Interval no.	0.33***	0.73***	0.58***	1.24***
	(0.04)	(0.08)	(0.07)	(0.16)
Interval no. sq.		-0.04***		-0.07***
		(0.01)		(0.01)
Constant	-0.53	-1.15***	-1.33**	-2.31***
	(0.34)	(0.32)	(0.65)	(0.64)
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})$	–	0.171	–	0.158
95% CI	–	[0.121, 0.220]	–	[0.105, 0.212]
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})\times$ SD	–	0.830	–	0.771
95% CI	–	[0.590, 1.070]	–	[0.508, 1.033]
Observations	4070	4070	4070	4070
N. Subjects	370	370	370	370

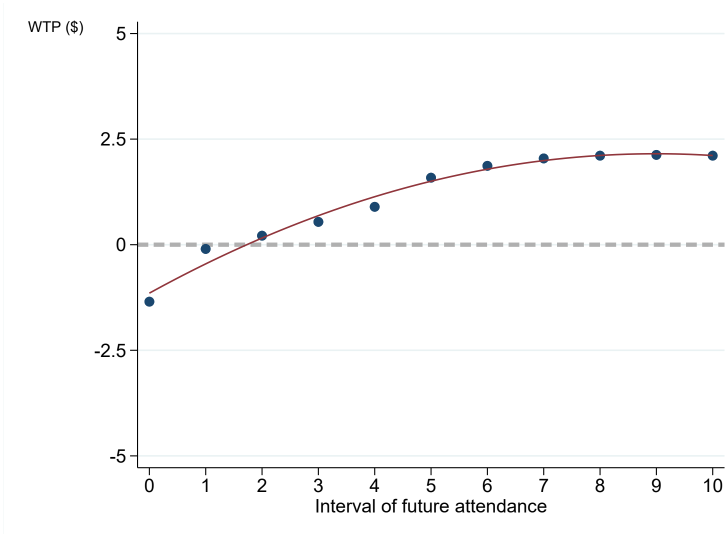
Notes: These tables report regression estimates from linear and quadratic models of willingness to pay for public recognition, by index of the interval. The interval index takes values from 0 to 10, corresponding to the 11 intervals of future attendance. Measures of the curvature of the estimated reduced-form public recognition function are $-R''_{exp}/R'_{exp}(\bar{a}_{pop})$ and $-R''_{exp}/R'_{exp}(\bar{a}_{pop}) \times SD$, where $\bar{a}_{pop}$ and $SD = 4.86$ are the average attendance and standard deviation of attendance for the general YOTA population, respectively. The analysis excludes 15 participants with "incoherent" preferences for public recognition. Standard errors are clustered at the participant level and reported in parentheses. 95 percent confidence intervals for the curvature statistics are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

Table A.3.5: WTP for public recognition by index of attendance interval, restricting to number of visits questions within 4 of predicted PR attendance

	(1)	(2)	(3)	(4)
Model	OLS	OLS	Tobit	Tobit
Dependent var.	WTP	WTP	WTP	WTP
Interval no.	0.70***	0.99***	1.17***	1.50**
	(0.12)	(0.34)	(0.23)	(0.67)
Interval no. sq.		-0.03		-0.03
		(0.03)		(0.06)
Constant	-3.18***	-3.75***	-5.56***	-6.22***
	(0.90)	(1.06)	(1.67)	(2.09)
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})$	–	0.067	–	0.049
95% CI	–	[-0.063, 0.197]	–	[-0.110, 0.207]
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})\times$ SD	–	0.327	–	0.236
95% CI	–	[-0.304, 0.957]	–	[-0.534, 1.006]
Observations	923	923	923	923
N. Subjects	370	370	370	370

Notes: These tables report regression estimates from linear and quadratic models of willingness to pay for public recognition, by index of the interval. The interval index takes values from 0 to 10, corresponding to the 11 intervals of future attendance. Data is restricted to visits intervals with a midpoint within 4 of a participant's predicted attendance if assigned to the public recognition group. Measures of the curvature of the estimated reduced-form public recognition function are $-R''_{exp}/R'_{exp}(\bar{a}_{pop})$ and $-R''_{exp}/R'_{exp}(\bar{a}_{pop}) \times SD$, where $\bar{a}_{pop}$ and $SD = 4.86$ are the average attendance and standard deviation of attendance for the general YOTA population, respectively. The analysis excludes 15 participants with "incoherent" preferences for public recognition. Standard errors are clustered at the participant level and reported in parentheses. 95 percent confidence intervals for the curvature statistics are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

Figure A.3.4: WTP for public recognition by index of interval



Notes: These figures plot the average WTP for public recognition by each of the eleven intervals of possible future attendance. Interval number takes values from 0 to 10, corresponding to the 11 intervals of future attendance. The analysis excludes 15 participants with "incoherent" preferences for public recognition.

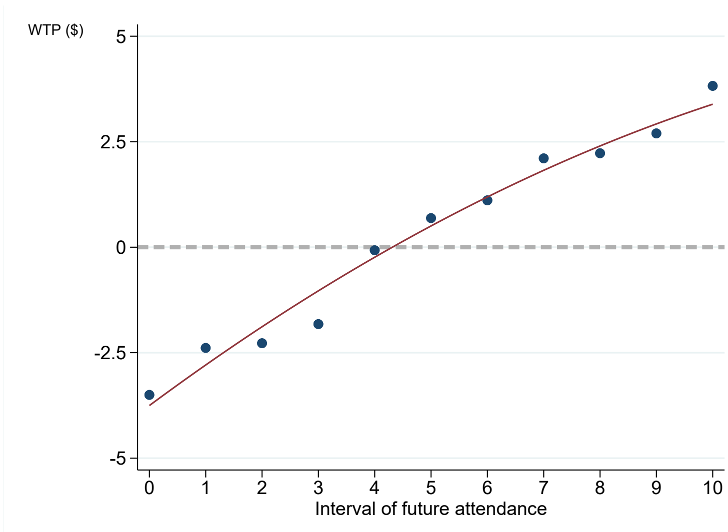
A.3.6 Interaction between Demand for Commitment and WTP for Public Recognition

To develop our measure of the WTP for motivation, we follow Carrera et al. (forthcoming) and Allcott et al. (forthcoming). Letting w_i be individual i's WTP for a \$1 attendance incentive, and letting $\alpha_i(0)$ and $\alpha_i(1)$ be this individual's expected visits in the absence and presence of the attendance incentive, Carrera et al. (forthcoming) and Allcott et al. (forthcoming) show that

$$m_i = w_i - \frac{\alpha_i(0) + \alpha_i(1)}{2}$$

is a measure of individuals' perceived time-inconsistency. This measure equals 0 for individuals who perceive themselves to be time-consistent, is positive for individuals who would like to attend the YMCA more, and is negative for individuals who believe that they attend the YMCA too much. Below, we study whether this measure relates to participants' profile of WTP for public recognition. We present regression results in Table A.3.6 and graphical results in Figure A.3.6.

Figure A.3.5: The reduced-form public recognition function: by interval, restricting to number of visits questions within 4 predicted PR attendance



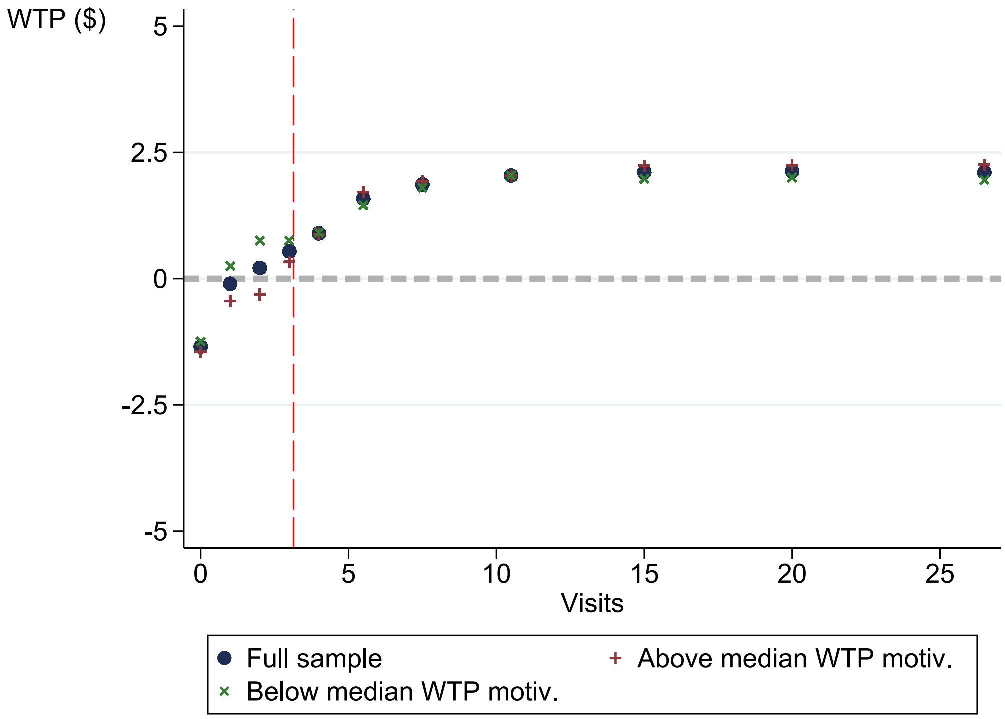
Notes: These figures plot the average WTP for each of the eleven intervals of possible future attendances to the YMCA during the experiment, restricting to intervals whose midpoint is within 4 visits of a participant's predicted attendance if assigned public recognition. For intervals including more than one number of visits (e.g., "between 7 and 8 visits"), the WTP is plotted at the average point of visits. The analysis excludes 15 participants with "incoherent" preferences for public recognition.

Table A.3.6: WTP for public recognition by YMCA attendance: heterogeneity along demand for commitment

	(1)	(2)	(3)	(4)
Model	OLS	OLS	Tobit	Tobit
Dependent var.	WTP	WTP	WTP	WTP
N. visits	0.36***	0.62***	0.59***	0.92***
	(0.04)	(0.08)	(0.14)	(0.26)
N. visits sq.	-0.01***	-0.02***	-0.01***	-0.02**
	(0.00)	(0.00)	(0.00)	(0.01)
WTP motivation	-0.03	-0.04	0.08	0.14
	(0.03)	(0.07)	(0.10)	(0.20)
N. visits $\times$ WTP motiv.	0.00	-0.00	-0.01	-0.01
	(0.00)	(0.01)	(0.01)	(0.03)
N. visits sq. $\times$ WTP motiv.	-0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Constant	-0.64*	-1.45**	-2.81***	-4.77***
	(0.33)	(0.63)	(0.91)	(1.69)
Restriction	All	All	≤ 4	≤ 4
Observations	4070	4070	923	923
N. Subjects	370	370	370	370

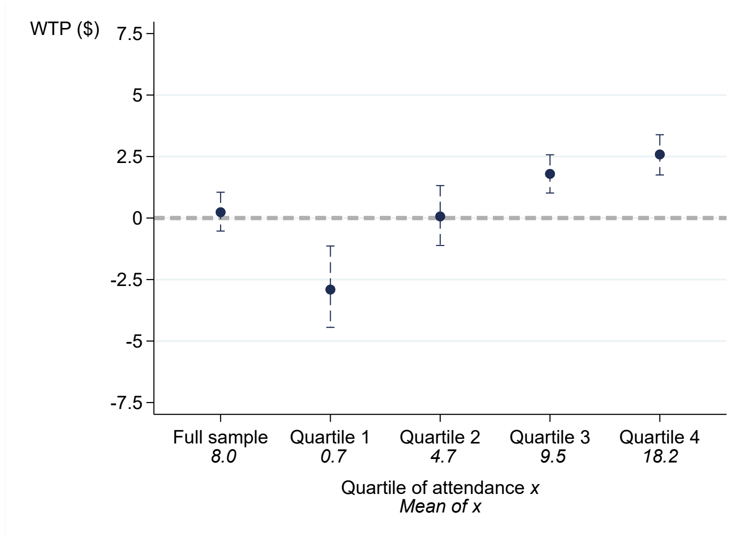
Notes: This table reports regression estimates of quadratic models of willingness to pay for public recognition by YMCA attendance. Columns (1)-(2) use all 11 intervals of future attendance, while columns (3)-(4) restrict to intervals with a midpoint within 4 of a participant's predicted attendance if assigned public recognition. WTP for motivation, m_i, is defined as $m_i := w_i - \frac{\alpha_i(0)+\alpha_i(1)}{2}$, where w_i is individual i's WTP for a \$1 attendance incentive, and $\alpha_i(0)$ and $\alpha_i(1)$ are the individual's expected visits in the absence and presence of the attendance incentive. The analysis excludes 15 participants with "incoherent" preferences for public recognition. Standard errors are clustered at the participant level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure A.3.6: WTP for public recognition by YMCA attendance: heterogeneity along demand for commitment



Notes: This figure plots the average WTP for public recognition by YMCA attendance. For intervals including more than one value of visits (e.g., "5 or 6 visits"), the WTP is plotted at the midpoint the interval. The figure separately reports the average WTP for the whole sample of coherent participants, and for coherent participants whose average attendance prior the experiment was below/above the median WTP for motivation. WTP for motivation, m_i, is defined as $m_i := w_i - \frac{\alpha_i(0)+\alpha_i(1)}{2}$, where w_i is individual i's WTP for a \$1 attendance incentive, and $\alpha_i(0)$ and $\alpha_i(1)$ are the individual's expected visits in the absence and presence of the attendance incentive. The average YOTA attendance is indicated by the dashed red line. The analysis excludes 15 participants with "incoherent" preferences for public recognition.

Figure A.3.7: The image payoff in the YMCA experiment



Notes: These figures plot the average realized payoff from public recognition, of participants assigned public recognition. We present results for both the full sample and each quartile of actual attendance. The average attendance is reported below each subsample label. The analysis excludes 15 participants with "incoherent" preferences for public recognition. Bootstrapped percentile-based confidence intervals, sampled by participant with 1000 iterations, are displayed.

A.3.7 Additional Results on Realized Image Payoffs

To construct the figures below, we instead estimated the reduced-form PRU non-parametrically. We define a participants' realized payoff as follows: If the participant attended the YMCA a times, then we compute $R_{exp}(a)$ to be the average WTP reported by participants for the elicitation interval containing a visits. To counter potential scaling bias, we continue limiting to data where the midpoints of the visits intervals are within 4 of participants' expected number of visits.

A.3.8 Replication of Main Results Restricting to Participants with Monotonic Preferences for Public Recognition

In this Appendix, we replicate our main analyses excluding an additional 31 participants with non-monotonic preferences for public recognition. This *monotonic sample* is of particular interest because it is consistent with the typical monotonicity assumptions of the models in Section 2.2.

Table A.3.7: The impact of public recognition on YMCA attendance

	(1)	(2)	(3)
Public recognition	1.20	1.26***	1.34***
	(0.73)	(0.48)	(0.47)
Avg. past att.		0.89***	0.78***
		(0.04)	(0.05)
Beliefs			0.20***
			(0.05)
Control mean	6.95	6.95	6.95
	(0.49)	(0.49)	(0.49)
N. Subjects	339	339	339

Notes: This table reports regression estimates of the effects of public recognition on attendance during the experiment. “Beliefs” reports the expectations YMCA members had about their attendance assuming that they would be part of the public recognition treatment. The analysis excludes 46 participants with “incoherent” preferences for public recognition. The control mean is the average attendance for participants in the experiment who are not in the public recognition program. Standard errors are clustered at the participant level and reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.3.8: WTP for public recognition by YMCA attendance

	(1)	(2)	(3)	(4)
Model	OLS	OLS	Tobit	Tobit
Dependent var.	WTP	WTP	WTP	WTP
N. visits	0.13***	0.39***	0.25***	0.68***
	(0.01)	(0.04)	(0.03)	(0.08)
N. visits sq.		-0.01***		-0.02***
		(0.00)		(0.00)
Constant	-0.14	-0.91***	-0.69	-2.00***
	(0.32)	(0.34)	(0.63)	(0.68)
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})$	–	0.062	–	0.061
95% CI	–	[0.058, 0.067]	–	[0.056, 0.065]
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})\times$ SD	–	0.303	–	0.294
95% CI	–	[0.282, 0.325]	–	[0.270, 0.318]
Observations	3729	3729	3729	3729
N. Subjects	339	339	339	339

Notes: This table reports regression estimates from linear and quadratic models of willingness to pay for public recognition by attendance. Measures of the curvature of the estimated reduced-form public recognition function are $-R''_{exp}/R'_{exp}(0)$ and $-R''_{exp}/R'_{exp}(0) \times SD$, where $SD = 4.86$ is the standard deviation attendance for the general YOTA population. The analysis excludes 46 participants with "incoherent" preferences for public recognition. Standard errors are clustered at the participant level and reported in parentheses. 95 percent confidence intervals for the curvature statistics are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

A.4 Supplementary Empirical Results for Charitable Contribution Experiments

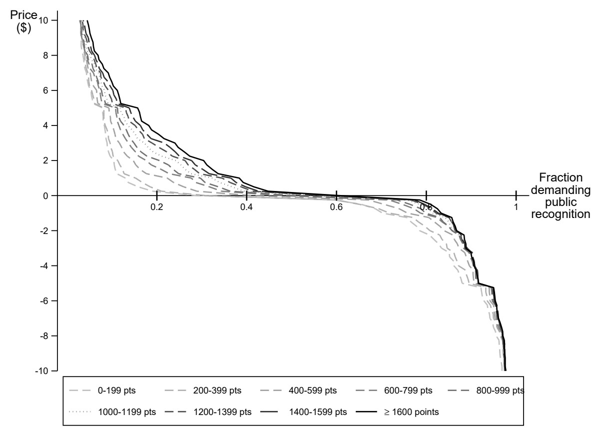
A.4.1 Demand for Public Recognition

Table A.3.9: WTP for public recognition by YMCA attendance, restricting to questions about visits close to participants' expectations

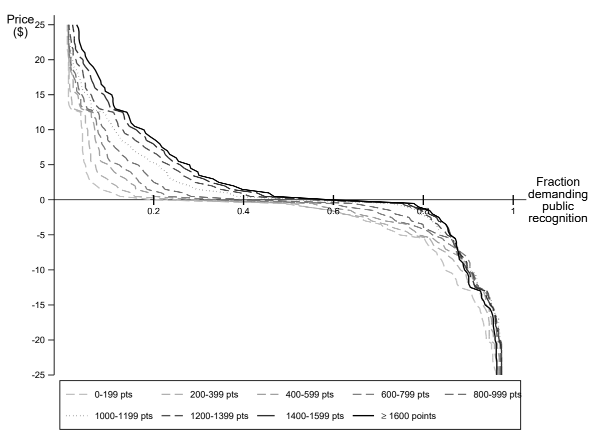
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model	OLS	OLS	Tobit	Tobit	OLS	OLS	Tobit	Tobit
Dependent var.	WTP	WTP	WTP	WTP	WTP	WTP	WTP	WTP
N. visits	0.27***	0.65***	0.47***	1.04***	0.22***	0.67***	0.43***	1.19***
	(0.05)	(0.14)	(0.09)	(0.27)	(0.05)	(0.18)	(0.10)	(0.38)
N. visits sq.		-0.02***		-0.02**		-0.01**		-0.03**
		(0.00)		(0.01)		(0.01)		(0.01)
Constant	-1.83***	-3.41***	-3.54***	-5.87***	-0.94	-3.67***	-2.51*	-7.11***
	(0.69)	(0.94)	(1.28)	(1.81)	(0.74)	(1.29)	(1.41)	(2.59)
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})$	–	0.057	–	0.053	–	0.052	–	0.050
95% CI	–	[0.040, 0.073]	–	[0.031, 0.074]	–	[0.034, 0.069]	–	[0.028, 0.071]
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})\times$ SD	–	0.275	–	0.257	–	0.252	–	0.241
95% CI	–	[0.194, 0.356]	–	[0.152, 0.361]	–	[0.166, 0.338]	–	[0.135, 0.347]
Restriction	≤ 4	≤ 4	≤ 4	≤ 4	Exact	Exact	Exact	Exact
Observations	830	830	830	830	339	339	339	339
N. Subjects	339	339	339	339	339	339	339	339

Notes: These tables report regression estimates from linear and quadratic models of willingness to pay for public recognition by attendance. Columns (1)-(4) restrict to visits intervals with a midpoint within 4 of a participant's predicted attendance if assigned to the public recognition group. Columns (5)-(8) restrict to intervals that contain the participant's predicted attendance if assigned to the public recognition group. Measures of the curvature of the estimated reduced-form public recognition function are $-R''_{exp}/R'_{exp}(\bar{a}_{pop})$ and $-R''_{exp}/R'_{exp}(\bar{a}_{pop}) \times SD$, where $\bar{a}_{pop}$ and $SD = 4.86$ are the average attendance and standard deviation of attendance for the general YOTA population, respectively. The analysis excludes 46 participants with "incoherent" preferences for public recognition. Standard errors are clustered at the participant level and reported in parentheses. 95 percent confidence intervals for the curvature statistics are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

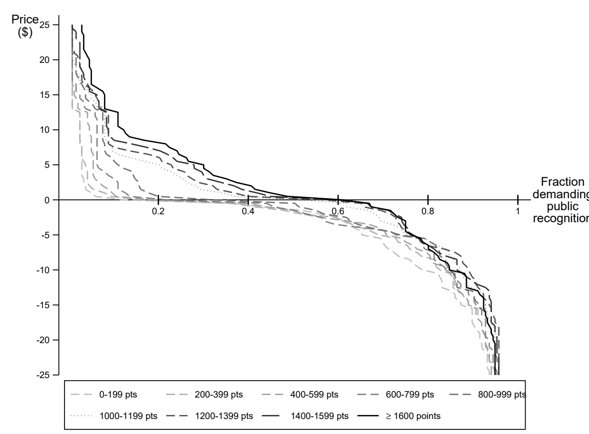
Figure A.4.1: Demand Curves for Public Recognition



(a) Prolific



(b) Berkeley



(c) BU

Notes: This figure plots the demand curves for public recognition by attendance interval. The analysis excludes 40 Prolific participants, 11 Berkeley participants, and 2 BU participants with "incoherent" preferences for public recognition

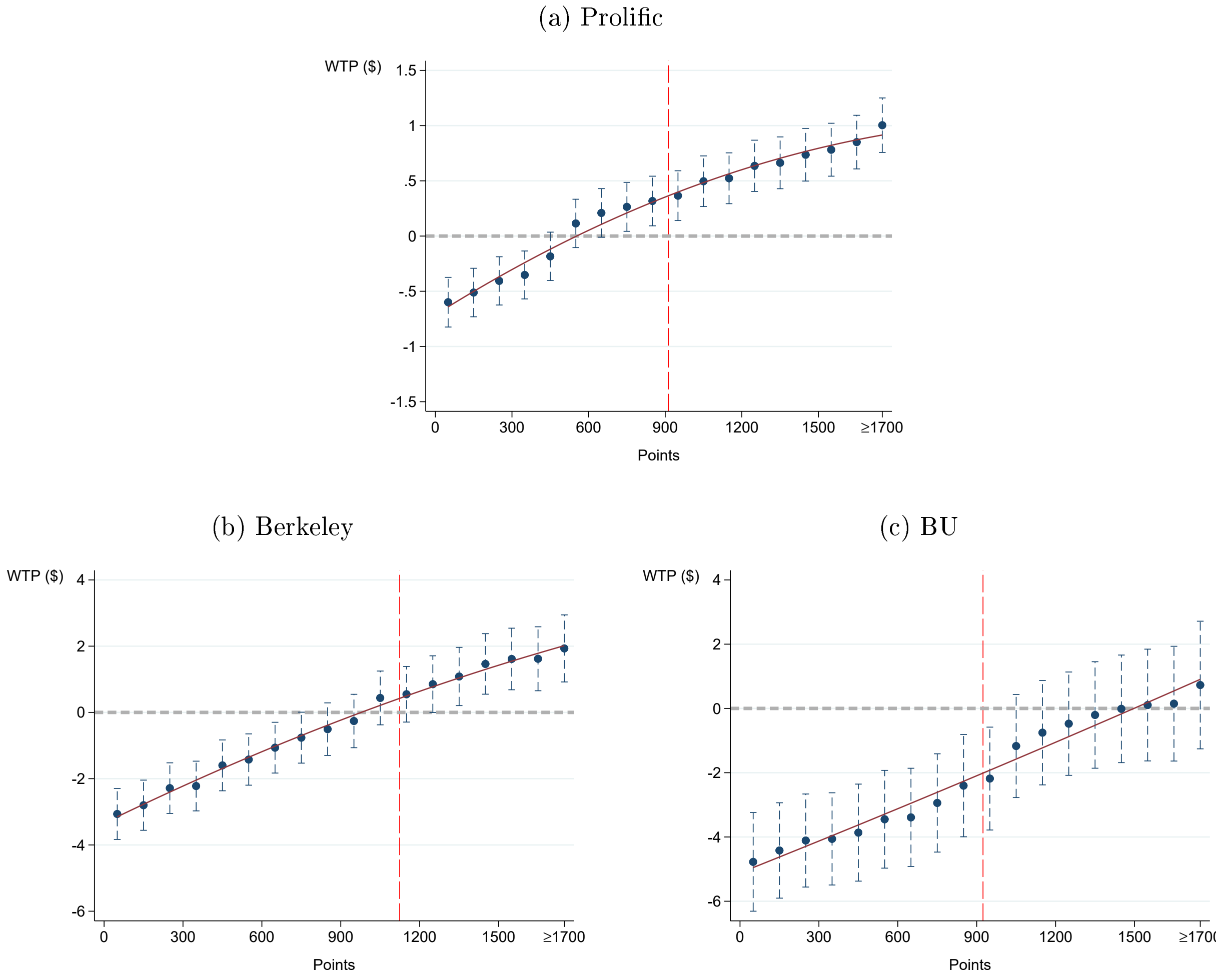
A.4.2 Robustness and Heterogeneity Analysis

Table A.4.1: The effect of public recognition on points scored, first round only

	(1)	(2)	(3)
Model	OLS	OLS	OLS
Dependent var.	Points	Points	Points
Public recognition	104.33***	132.68**	-27.67
	(39.85)	(58.75)	(130.50)
Financial incentives	174.83***	153.18**	-50.94
	(38.31)	(59.45)	(123.83)
Control mean	824.0	1012.4	974.8
	(26.7)	(42.5)	(91.0)
Sample	Prolific	Berkeley	BU
N. Subjects	968	384	118

Notes: This table reports regression estimates of the effects of public recognition and financial incentives on points scored and is limited to observations from the first round randomly assigned to be completed by each participant. The control mean is the mean points scored in the Anonymous Effort Round. The analysis excludes 40 Prolific participants, 11 Berkeley participants, and 2 BU participants with "incoherent" preferences for public recognition. Heteroskedasticity-robust standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure A.4.2: WTP for public recognition by effort in the charitable contribution experiments



Notes: These figures plot the average WTP for public recognition with 95 percent confidence intervals for each of the eighteen intervals of possible points scored in the round selected for public recognition. The WTP is plotted at the midpoint of each of the first seventeen intervals and at ≥1700 points for the 1700 or more points interval. The mean Publicly-Shared Effort Round scores are indicated by dashed red lines. The analysis excludes 40 Prolific participants, 11 Berkeley participants, and 2 BU participants with "incoherent" preferences for public recognition. 95 percent confidence intervals are constructed using standard errors clustered by participant.

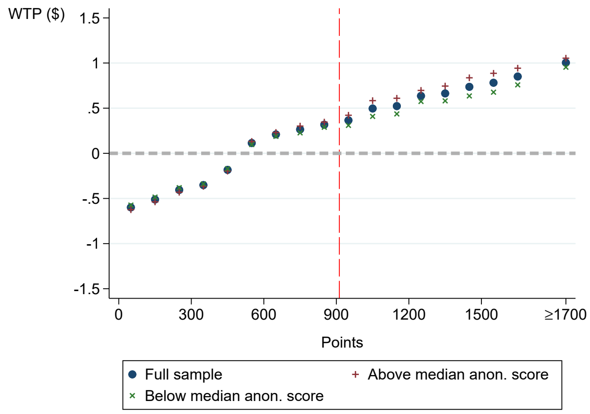
Table A.4.2: WTP for public recognition by effort in the charitable contribution experiments, restricting to questions about scores that are "close" to participants' actual scores

	(1)	(2)	(3)	(4)	(5)	(6)
Model	OLS	OLS	OLS	OLS	OLS	OLS
Dependent var.	WTP	WTP	WTP	WTP	WTP	WTP
Points (00s)	0.106***	0.150***	0.371***	0.376	0.390***	0.341
	(0.020)	(0.055)	(0.072)	(0.230)	(0.135)	(0.288)
Points (00s) sq.		-0.003		-0.000		0.003
		(0.003)		(0.010)		(0.014)
Constant	-0.591***	-0.737**	-3.520***	-3.538***	-5.298***	-5.145***
	(0.214)	(0.286)	(0.804)	(1.252)	(1.178)	(1.483)
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})$	–	0.046	–	0.001	–	-0.015
95% CI	–	[-0.044, 0.136]	–	[-0.107, 0.110]	–	[-0.162, 0.133]
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}}) \times \mathrm{SD}$	–	0.174	–	0.007	–	-0.095
95% CI	–	[-0.073, 0.421]	–	[-0.558, 0.573]	–	[-1.184, 0.994]
Sample	Prolific	Prolific	Berkeley	Berkeley	BU	BU
Observations	8602	8602	3330	3330	982	982
N. Subjects	968	968	383	383	118	118

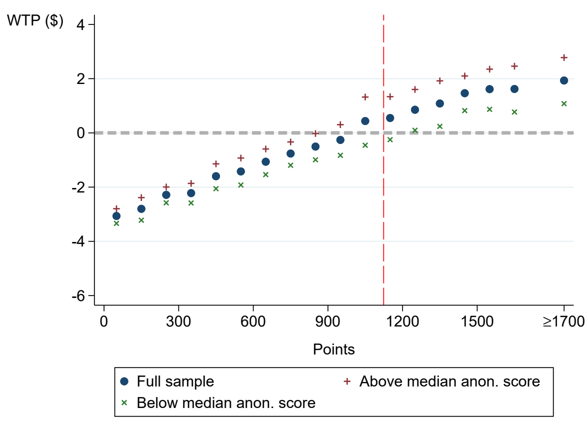
Notes: This table reports regression estimates from linear and quadratic models of willingness to pay for public recognition by the level of publicized effort. The data is restricted to observations in which the midpoint of the points interval for which willingness to pay is reported is within 500 points of the participant's average score across the three experimental rounds. Effort is measured in 100s of points scored. The regressions exclude the ≥1700 points interval. Measures of the curvature of the estimated reduced-form public recognition function are $-R''_{exp}/R'_{exp}(\bar{a}_{pop})$ and $-R''_{exp}/R'_{exp}(\bar{a}_{pop}) \times SD$, where $\bar{a}_{pop}$ and $SD = 4.86$ are the average and standard deviation of points scored in the anonymous round (in units of hundreds of points), respectively. The analysis excludes 40 Prolific participants, 11 Berkeley participants, and 2 BU participants with "incoherent" preferences for public recognition. Standard errors are clustered at the participant level and reported in parentheses. 95 percent confidence intervals for the curvature statistics are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

Figure A.4.3: Willingness to pay for public recognition by effort in the charitable contribution experiments, setting top interval at average performance

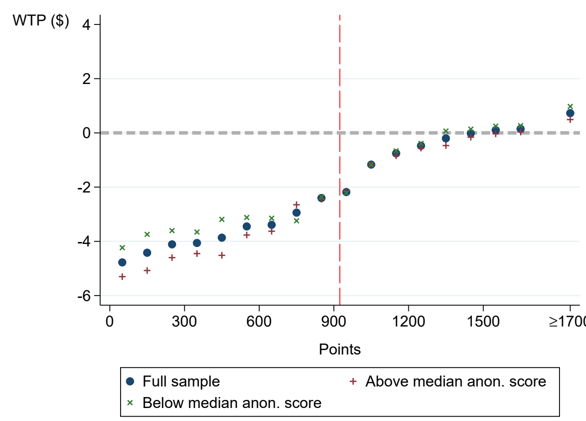
(a) Prolific



(b) Berkeley



(c) BU



Notes: These figures plot the average WTP for public recognition by each of the 18 possible intervals of points scored. The WTP is plotted at the midpoint of each of the first seventeen intervals. For the ≥ 1700 points interval, publicized effort is calculated as the average of the mean points scored among participants whose score lies in that interval for the public recognition round. Panel (a) presents results for the Prolific sample, panel (b) presents results for the Berkeley sample, and panel (c) presents results for the BU sample. The mean Publicly-Shared Effort Round scores are indicated by dashed red lines. The analysis excludes 40 Prolific participants, 11 Berkeley participants, and 2 BU participants with "incoherent" preferences for public recognition.

Table A.4.3: WTP for public recognition by effort in the charitable contribution experiments, including the ≥ 1700 interval

	(1)	(2)	(3)
Model	OLS	OLS	OLS
Dependent var.	WTP	WTP	WTP
Points (00s)	0.146***	0.405***	0.356***
	(0.017)	(0.065)	(0.107)
Points (00s) sq.	-0.003***	-0.006**	-0.001
	(0.001)	(0.003)	(0.005)
Constant	-0.713***	-3.383***	-5.184***
	(0.120)	(0.422)	(0.816)
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})$	0.063	0.040	0.006
95% CI	[0.039, 0.087]	[-0.001, 0.081]	[-0.048, 0.060]
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}}) \times \mathrm{SD}$	0.217	0.153	0.032
95% CI	[0.162, 0.271]	[0.042, 0.264]	[-0.249, 0.313]
Sample	Prolific	Berkeley	BU
Observations	17424	6912	2124
N. Subjects	968	384	118

Notes: This table reports regression estimates from linear and quadratic models of willingness to pay for public recognition by the level of publicized effort. Effort is measured in 100s of points scored. For the ≥ 1700 points interval, publicized effort is calculated as the average of the mean points scored among participants whose score lies in that interval for the public recognition round. Measures of the curvature of the estimated reduced-form public recognition function are $-R''_{exp}/R'_{exp}(\bar{a}_{pop})$ and $-R''_{exp}/R'_{exp}(\bar{a}_{pop}) \times SD$, where $\bar{a}_{pop}$ and $SD = 4.86$ are the average and standard deviation of points scored in the anonymous round (in units of hundreds of points), respectively. The analysis excludes 40 Prolific participants, 11 Berkeley participants, and 2 BU participants with "incoherent" preferences for public recognition. Standard errors are clustered at the participant level and reported in parentheses. 95 percent confidence intervals for the curvature statistics are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

Table A.4.4: WTP for public recognition by effort in the charitable contribution experiments: heterogeneity in sensitivity

	(1)	(2)	(3)	(4)	(5)	(6)
Model	OLS	OLS	OLS	OLS	OLS	OLS
Dependent var.	WTP	WTP	WTP	WTP	WTP	WTP
Points (00s)	0.069***	0.131***	0.252***	0.304***	0.333***	0.418**
	(0.009)	(0.025)	(0.047)	(0.086)	(0.092)	(0.187)
Points (00s) sq.		-0.004***		-0.003		-0.005
		(0.001)		(0.004)		(0.009)
Above med. PR impact	-0.171	-0.168	-0.861	-0.955	-1.141	-0.440
	(0.226)	(0.242)	(0.799)	(0.839)	(1.580)	(1.621)
Points (00s) ×	0.047***	0.046	0.117*	0.150	0.028	-0.219
Above med. PR impact	(0.014)	(0.035)	(0.066)	(0.140)	(0.121)	(0.232)
Points (00s) sq. ×		0.000		-0.002		0.015
Above med. PR impact		(0.002)		(0.007)		(0.011)
Constant	-0.471***	-0.649***	-2.699***	-2.847***	-4.616***	-4.856***
	(0.162)	(0.173)	(0.628)	(0.635)	(0.997)	(1.046)
Sample	Prolific	Prolific	Berkeley	Berkeley	BU	BU
Observations	16456	16456	6528	6528	2006	2006
N. Subjects	968	968	384	384	118	118

Notes: This table reports coefficient estimates from linear and quadratic models of willingness to pay for public recognition at different levels of points scored, in units of hundreds of points. It includes an indicator for the difference between the participant's scores in the anonymous and public recognition rounds being above the median as well as its interactions with points levels. The analysis excludes 40 Prolific participants, 11 Berkeley participants, and 2 BU participants with "incoherent" preferences for public recognition. Standard errors are clustered at the participant level and reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.4.5: WTP for public recognition by effort in the charitable contribution experiments: heterogeneity by intrinsic motivation

	(1)	(2)	(3)	(4)	(5)	(6)
Model	OLS	OLS	OLS	OLS	OLS	OLS
Dependent var.	WTP	WTP	WTP	WTP	WTP	WTP
Points (00s)	0.083***	0.142***	0.275***	0.333***	0.315***	0.177
	(0.010)	(0.025)	(0.049)	(0.110)	(0.083)	(0.166)
Points (00s) sq.		-0.003***		-0.003		0.008
		(0.001)		(0.006)		(0.009)
Above med. anon. score	-0.077	-0.094	0.548	0.488	-0.998	-1.550
	(0.227)	(0.242)	(0.800)	(0.841)	(1.572)	(1.605)
Points (00s) ×	0.018	0.024	0.070	0.091	0.064	0.258
Above med. anon. score	(0.015)	(0.035)	(0.066)	(0.140)	(0.120)	(0.232)
Points (00s) sq. ×		-0.000		-0.001		-0.011
Above med. anon. score		(0.002)		(0.007)		(0.011)
Constant	-0.518***	-0.686***	-3.405***	-3.570***	-4.679***	-4.287***
	(0.168)	(0.178)	(0.573)	(0.615)	(0.920)	(0.919)
Sample	Prolific	Prolific	Berkeley	Berkeley	BU	BU
Observations	16456	16456	6528	6528	2006	2006
N. Subjects	968	968	384	384	118	118

Notes: This table reports coefficient estimates from linear and quadratic models of willingness to pay for public recognition at different levels of points scored, in units of hundreds of points. It includes an indicator for the participant having scored above the median number of points in the anonymous round as well as its interactions with points levels. The analysis excludes 40 Prolific participants, 11 Berkeley participants, and 2 BU participants with "incoherent" preferences for public recognition. Standard errors are clustered at the participant level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.4.6: WTP for public recognition by effort in the charitable contribution experiments: heterogeneity along public recognition group size

	(1)	(2)
Model	OLS	OLS
Dependent var.	WTP	WTP
Points (00s)	0.098***	0.159***
	(0.011)	(0.027)
Points (00s) sq.		-0.004***
		(0.001)
Group of 300	0.121	0.141
	(0.256)	(0.268)
Group of 300 × Points (00s)	-0.016	-0.024
	(0.017)	(0.039)
Group of 300 × Points (00s) sq.		0.001
		(0.002)
Group of 15	0.332	0.305
	(0.293)	(0.307)
Group of 15 × Points (00s)	-0.001	0.009
	(0.018)	(0.044)
Group of 15 × Points (00s) sq.		-0.001
		(0.002)
Constant	-0.676***	-0.852***
	(0.163)	(0.175)
Observations	16456	16456
N. Subjects	968	968

Notes: This table reports regression estimates from linear and quadratic models of willingness to pay for public recognition by the level of publicized effort in the Prolific sample. Effort is measured in 100s of points scored. The regressions exclude the $\geq$1700 points interval. The regressions include interactions with group size variables in the Prolific sample, which indicate the approximate number of individuals in the participant's randomly assigned public recognition group. The omitted group size category is 75 participants. The analysis excludes 40 Prolific participants, 11 Berkeley participants, and 2 BU participants with "incoherent" preferences for public recognition. Standard errors are clustered at the subject level and reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

A.4.3 Results for QM221 Sample

Before the experiment started, we preregistered our analysis plan on the AEA RCT Registry (AEARCTR-0005737). We had originally planned to also create a fourth sample from the QM221 statistics class for first-year students (who know each other less well than the QM222 students), but the response rate was too low to make use of this data. For transparency, we report reduced-form results for the QM 221 class below.

Table A.4.7: **Reduced-form results for the QM 221 sample**

(a) The effect of public recognition on points scored

	(1)
Model	OLS
Dependent var.	Points
Public recognition	122.20*
	(72.11)
Financial incentives	156.36**
	(71.77)
Control mean	910.7
	(80.1)
Observations	156
N. Subjects	52

(b) WTP for public recognition by effort in the charitable contribution experiments

	(1)	(2)
Model	OLS	OLS
Dependent var.	WTP	WTP
Points (00s)	0.297***	0.397*
	(0.109)	(0.217)
Points (00s) sq.		-0.006
		(0.009)
Constant	-1.797	-2.080
	(1.109)	(1.272)
95% CI	–	0.040
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})$	–	[-0.082, 0.163]
95% CI	–	0.171
$-\mathrm{R}''/\mathrm{R}'(\bar{a}_{\mathrm{pop}})\times$ SD	–	[-0.208, 0.549]
95% CI	884	884
Observations	52	52

Notes: Panel (a) reports regression estimates of the effects of public recognition and financial incentives on points scored..The control mean is the mean points scored in the Anonymous Effort Round. Dummy variables for the order in which the round appeared (first, second, or third) are included. Panel (b) reports regression estimates from linear and quadratic models of willingness to pay for public recognition by the level of publicized effort in the Prolific sample. Effort is measured in 100s of points scored. The regressions exclude the ≥1700 points interval. Measures of the curvature of the estimated reduced-form public recognition function are $-R''_{exp}/R'_{exp}(\bar{a}_{pop})$ and $-R''_{exp}/R'_{exp}(\bar{a}_{pop})\times SD$, where $\bar{a}_{pop}$ and $SD = 4.86$ are the average and standard deviation of points scored in the anonymous round (in units of hundreds of points), respectively. The analysis excluded participants with "incoherent" preferences for public recognition. Standard errors are clustered at the subject level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

A.4.4 Model Selection

We formally test for the appropriate functional form using the Bayesian Information Criteria (BIC), which penalizes model complexity and rewards goodness-of-fit:[2]

$$BIC = \underbrace{-2 \cdot \ln(l)}_{\text{reward for fit}} + \underbrace{k \cdot \ln(n)}_{\text{penalty for complexity}} .$$

Here l denotes the maximized value of the likelihood function for the model, k the number of parameters in the model, and n the sample size. The magnitude of the BIC is not on its own informative, but is instead to be compared with the BIC of other models. When considering similar models, the one with the lowest BIC best fits the data. Kass and Raftery (1995) and Raftery (1995) discuss guidelines for interpreting the differences in magnitudes of the BIC, which we list in Table A.4.8.

Table A.4.8: Guidelines for comparing BIC magnitude

$BIC(modelA) - BIC(modelB)$	Interpretation
$\in (-\infty, -10]$	Decisive evidence for model A
$\in (-10, -6]$	Strong evidence for model A
$\in (-6, -2]$	Positive evidence for model A
$\in (-2, 2)$	Weak evidence for either model
$\in [2, 6)$	Positive evidence for model B
$\in [6, 10)$	Strong evidence for model B
$\in [10, \infty)$	Decisive evidence for model B

Sources: Kass and Raftery (1995) and Raftery (1995). While Kass and Raftery (1995) label a difference of $(-2, 0)$ as weak evidence for model A and $(0, 2)$ as weak evidence for model B, we follow Raftery (1995) in labeling these as weak evidence for either model.

For the Prolific sample, the BIC is minimized for the quadratic specification, where it is roughly 9 points lower than the BIC of the cubic model, and 18 points lower than the quartic model. Per the guidelines in Table A.4.8, this provides strong evidence to reject the cubic model in favor of the quadratic model, and decisive evidence to reject the quartic model.

While the BIC provides strong to decisive evidence to reject the cubic and quartic models, visual examination of the PRUs suggests moderate jumps in the WTP around 500 and 1000 points, which makes the PRU look S-shaped. We hypothesize that these jumps are attributable to a round-number heuristic. Under this hypothesis, participants might heuristically feel most compelled to significantly adjust their WTP when they pass a multiple of 500. This hypothesis is consistent with the fact that we see the jumps appear most

[2] The BIC was first developed in Schwarz (1978), which now has over 46,000 Google Scholar citations. The approach is widely used in model selection for social science research, including economics (see e.g., Kim (1998) and Steel (2020)).

prominently for the Prolific sample, where participants move through the experiment more quickly than in the university samples, and thus may be more likely to rely on heuristics.

To test this round-number heuristic, we re-estimate the linear to quartic models including dummy variables for having the points scored exceed 500, 1000, or 1500 points. For all three samples, third- and fourth-order terms are no longer significant. However, the quadratic term in the Prolific sample is still significant. The results suggest the statistical significance of higher-order polynomial terms is likely not due to multiple inflection points in the aggregate PRU, but rather due to some modest round-number bias.

While we present estimates of the model with the dummy variables to highlight the round-number bias, including these is not costless. Including these fixed effects reduces the precision of our estimates, particularly in the BU sample where the sample size is the smallest. Additionally, when comparing the BIC of all columns for the Prolific sample, where the round-number bias appears most prominent, we again see strong to decisive evidence to prefer the quadratic specification. These results suggest that the linear and quadratic specifications should be the primary specification.

Similarly, for the Berkeley and BU samples, we see strong evidence to prefer the linear specification and to reject the higher-order specifications.

A.5 Individual Differences Analysis

In this section we allow for heterogeneity in individuals' reduced-form PRU. Each individual i's reduced-form PRU is given by $r_{0i} + r_{1i}a + r_{2i}a^2$, where the parameters r_{0i}, r_{1i}, and r_{2i} are jointly distributed as follows:

$$\begin{bmatrix} r_{0i} \\ r_{1i} \\ r_{2i} \end{bmatrix} \sim N\left(\begin{bmatrix} E[r_0] \\ E[r_1] \\ E[r_2] \end{bmatrix}, \begin{bmatrix} Var[r_0] & 0 & 0 \\ 0 & Var[r_1] & Cov[r_1, r_2] \\ 0 & Cov[r_1, r_2] & Var[r_2] \end{bmatrix} \right)$$

To estimate the moments in this joint distribution, we use a mixed effects model. Specifically, we define w_{ij} to denote the WTP for public recognition of individual i if their performance lies in interval j, and estimate the following model:

$$w_{ij} = \beta_0 + \beta_1 a_{ij} + \beta_2 a_{ij}^2 + u_{0j} + u_{1i} a_{ij} + u_{2i} a_{ij}^2 + \varepsilon_{ij} \tag{A.6}$$

Here β_0, β_1, and β_2 identify the population average reduced-form PRU, with $\hat{\beta}_0 = \mathbb{E}[r_0]$, $\hat{\beta}_1 = \mathbb{E}[r_1]$, and $\hat{\beta}_2 = \mathbb{E}[r_2]$. u_{0i}, u_{1i}, and u_{2i} are mean-zero random effects on the scalar, linear and quadratic terms of the reduced-form PRU, respectively, and capture individual deviations from the population average. We estimate the variance-covariance matrix of the random effects using maximum likelihood, imposing zero covariance between the random effect for the constant and those for the linear and quadratic terms. By construction, the estimated variance of the random effects u_{0i}, u_{1i}, and u_{2i} identify $Var[r_0]$, $Var[r_1]$, and $Var[r_2]$, respectively, and the estimated covariance between u_{1i} and u_{2i} identifies $Cov[r_1, r_2]$.

Tables A.5.1-A.5.2 present the results for the YMCA sample and the charity samples, respectively. Across all samples, we estimate small variances for r_1 and r_2, and a large negative covariance between r_1 and r_2. Collectively, these results suggest that the ratio

Table A.5.1: Individual differences: YMCA sample

Parameter	(1) Point estimate	(2) Std. Error	Parameter	(3) Point estimate	(4) Std. Error
$\mathbb{E}[r_0]$	-1.125	0.701	$Var[r_0]$	19.669	1.772
$\mathbb{E}[r_1]$	0.410	0.095	$Var[r_1]$	0.262	0.070
$\mathbb{E}[r_2]$	-0.011	0.003	$Var[r_2]$	0.00024	0.00009
			$Cov[r_1, r_2]$	-0.008	0.003

Notes: This table reports regression estimates of willingness to pay for public recognition by YMCA attendance using equation A.6. .The analysis restricts to intervals with a midpoint within 4 of a participant's predicted attendance if assigned public recognition, and analysis excludes 31 participants with "incoherent" preferences for public recognition. Standard errors are clustered at the subject level.

of r_2/r_1 is nearly constant across individuals, and thus that there is little heterogeneity in curvature. We also estimate larger variances for r_0, suggesting larger heterogeneity in the reference point parameter ρ.

A.6 Structural Estimation Details

A.6.1 Action-signaling Model

Public recognition utility has the form $\nu S^a(a - \rho^a \bar{a}_{pop}) = \gamma_1^a(a - \rho^a \bar{a}_{pop}) + \gamma_2^a(a - \rho^a \bar{a}_{pop})^2$, where participants compare their action to a multiple of the average action $\bar{a}_{pop}$ of the general population. Total utility $U(a;\theta)$ is thus:

$$U(a;\theta) = \theta a - \frac{c}{2}a^2 + y + pa + \gamma_1^a(a - \rho^a \bar{a}_{pop}) + \gamma_2^a(a - \rho^a \bar{a}_{pop})^2 \tag{A.7}$$

Unless otherwise noted, we make the simplifying assumption that $p = 0$.

Before introducing public recognition, the population is initially in equilibrium given by $a^*(\theta) = \theta/c$ and $\bar{a}^0_{pop} := \mathbb{E}[\theta/c]$. We verify this by taking the F.O.C. of equation (A.7) with respect to a when $\gamma_1^a = \gamma_2^a = 0$ and solving for a.

Equilibrium Behavior

We now consider the impact of introducing public recognition to the population. We first solve for the equilibrium action function $a^*(\theta)$:

$$a^*(\theta) = \frac{\theta/c + \gamma_1^a/c - 2\gamma_2^a \rho^a \bar{a}^{eq}_{pop}/c}{1 - 2\gamma_2^a/c} \tag{A.8}$$

Table A.5.2: Individual differences: charity samples

(a) Prolific sample

Parameter	(1) Point estimate	(2) Std. Error	Parameter	(3) Point estimate	(4) Std. Error
$\mathbb{E}[r_0]$	-0.733	0.121	$Var[r_0]$	13.624	0.943
$\mathbb{E}[r_1]$	0.155	0.018	$Var[r_1]$	0.277	0.038
$\mathbb{E}[r_2]$	-0.004	0.001	$Var[r_2]$	0.00055	0.00008
			$Cov[r_1, r_2]$	-0.011	0.002

(b) Berkeley sample

Parameter	(1) Point estimate	(2) Std. Error	Parameter	(3) Point estimate	(4) Std. Error
$\mathbb{E}[r_0]$	-3.325	0.420	$Var[r_0]$	62.907	7.724
$\mathbb{E}[r_1]$	0.379	0.070	$Var[r_1]$	1.647	0.298
$\mathbb{E}[r_2]$	-0.004	0.004	$Var[r_2]$	0.00420	0.00074
			$Cov[r_1, r_2]$	-0.072	0.014

(c) BU sample

Parameter	(1) Point estimate	(2) Std. Error	Parameter	(3) Point estimate	(4) Std. Error
$\mathbb{E}[r_0]$	-5.076	0.810	$Var[r_0]$	70.867	13.540
$\mathbb{E}[r_1]$	0.309	0.116	$Var[r_1]$	1.299	0.483
$\mathbb{E}[r_2]$	0.002	0.006	$Var[r_2]$	0.00309	0.00093
			$Cov[r_1, r_2]$	-0.053	0.021

Notes: This table reports regression estimates of willingness to pay for public recognition by the level of publicized effort in the Prolific sample using equation A.6. The Effort is measured in 100s of points scored. The regressions exclude the $\geq$1700 points interval. The analysis excludes 40 Prolific participants, 11 Berkeley participants, and 2 BU participants with “incoherent” preferences for public recognition. Standard errors are clustered at the subject level.

Here $\bar{a}^{eq}_{pop}$ denotes the equilibrium average attendance, and its form depends on whether we are in a partial equilibrium or a full equilibrium. In a partial equilibrium, the reference population is not receiving public recognition and $\bar{a}^{eq}_{pop}$ thus remains constant at its initial value $\bar{a}^{0}_{pop}$:

$$\bar{a}^{eq}_{pop} = \bar{a}^{0}_{pop} \tag{A.9}$$

In a full equilibrium, public recognition is scaled up to the entire population, and so average attendance will increase until it reaches an equilibrium value $\bar{a}^{eq}_{pop}$. This equilibrium value is given by:

$$\bar{a}^{eq}_{pop} = \frac{\bar{a}^{0}_{pop} + \gamma^a_1/c}{(1 - 2(1-\rho^a)\gamma^a_2/c)} \tag{A.10}$$

In sum, the partial equilibrium is defined by equations (A.8) and (A.9), and the full equilibrium is defined by equations (A.8) and (A.10).

To see why these equations define the equilibrium, we take the F.O.C. of equation (A.7) with respect to a and solve for $a^*(\theta)$. From this we immediately recover equation (A.8). To verify equation (A.10), we take the expectation of both sides of equation (A.8), recalling that we are in the case where everyone receives public recognition:

$$\mathbb{E}[a^*(\theta)|PR = 1] = \frac{\mathbb{E}[\theta/c] + \gamma^a_1/c - 2\gamma^a_2\rho^a\bar{a}^{eq}_{pop}/c}{1 - 2\gamma^a_2/c}$$

To simplify the above expression, we first note that since everyone is receiving public recognition $\mathbb{E}[a^*(\theta)|PR = 1] = \bar{a}^{eq}_{pop}$. Second, we recall that $E[\theta/c] = \bar{a}^{0}_{pop}$. Substituting both of these into the equation above yields the following expression:

$$\bar{a}^{eq}_{pop} = \frac{\bar{a}^{0}_{pop} + \gamma^a_1/c - 2\gamma^a_2\rho^a\bar{a}^{eq}_{pop}/c}{1 - 2\gamma^a_2/c}$$

By isolating $\bar{a}^{eq}_{pop}$ in the equation above, we recover equation (A.10).

The Predicted Impact of Financial Incentives

With a financial incentive p per a and no public recognition, the utility function is given by $U(a;\theta) = \theta a - \frac{c}{2}a^2 + y + pa$. We use the first order condition to solve for a as a function of p:

$$a^*(\theta; p) = \theta/c + p/c \tag{A.11}$$

The impact of financial incentives on attendance, $a^*(\theta; p) - a^*(\theta; 0)$, is thus equal to p/c.

Mapping the Model Parameters to a Reduced-Form PRU

If the structural PRU is quadratic, it is immediate that the reduced-form PRU is also quadratic. We denote the reduced-form PRU by $R(a) = r_0 + r_1 a + r_2 a^2$. Unlike νS^a, the reduced-form PRU $R(a)$ is estimable from our data. In this section we derive mapping equations to express the structural parameters γ_1^a, γ_2^a, and ρ^a from a partial equilibrium to the reduced-form parameters r_0, r_1, r_2:

$$\gamma_1^a = \sqrt{r_1^2 - 4r_0r_2} \tag{A.12}$$

$$\gamma_2^a = r_2 \tag{A.13}$$

$$\rho^a = \frac{\sqrt{r_1^2 - 4\gamma_2^a r_0} - r_1}{2\bar{a}_{pop}^0 r_2} \tag{A.14}$$

To see why equations (A.12)-(A.14) hold, we begin by regrouping the terms in $\nu S(a - \bar{a}_{pop}^0)$:

$$\nu S(a - \rho^a \bar{a}_{pop}^0) = \left[\gamma_2^a(\rho^a\bar{a}_{pop}^0)^2 - \gamma_1^a(\rho^a\bar{a}_{pop}^0)\right] + \left[\gamma_1^a - 2\gamma_2^a(\rho^a\bar{a}_{pop}^0)\right] a + \gamma_2^a \cdot a^2$$

We next map this equation to $R(a) = r_0 + r_1 a + r_2 a^2$, which results in the following system of equations:

$$\gamma_2^a(\rho^a\bar{a}_{pop}^0)^2 - \gamma_1^a(\rho^a\bar{a}_{pop}^0) = r_0 \tag{A.15}$$

$$\gamma_1^a - 2\gamma_2^a(\rho^a\bar{a}_{pop}^0) = r_1 \tag{A.16}$$

$$\gamma_2^a = r_2 \tag{A.17}$$

Below we outline the algebra to isolate the structural parameters from mapping equations (A.15)-(A.17). First, equation (A.17) immediately verifies equation (A.13). Using $\gamma_2^a = r_2$ and the quadratic formula, we solve for ρ^a in terms of γ_1^a:

$$\rho^a = \frac{\gamma_1^a - \sqrt{(\gamma_1^a)^2 + 4r_0r_2}}{2\bar{a}_{pop}^0 r_2} \tag{A.18}$$

By substituting the above expression and $\gamma_2^a = r_2$ into equation (A.16), we recover equation (A.12). By substituting equation (A.12) into equation (A.18), we recover equation (A.14).

Identifying the Model

The reduced-form public recognition function: For the YMCA sample, $R(a)$ corresponds to the quadratic Tobit regression of WTP on visits in column (2) of Table 1.4b, which restricts to intervals of attendance within four of the participant's predicted attendance with public recognition. For the samples in the charitable contribution experiment, $R(a)$ corresponds to the quadratic OLS regression of WTP on hundreds of points in columns (2), (4), and (6) of Table 1.6.

The effects of public recognition on performance: We define $\bar{\tau} := \mathbb{E}[a|PR = 1] - \mathbb{E}[a|PR = 0]$ as the difference in average intensity between the experimental population that received public recognition ($PR = 1$) and the experimental population that did not ($PR = 0$). For the YMCA sample, we estimate $\bar{\tau}$ by controlling for past attendance. For the charitable contribution experiments, we estimate $\bar{\tau}$ by controlling for order effects, and allow it to vary by sample. For all samples, $\bar{a}_0 := \mathbb{E}[a|PR = 0]$ is directly observable as the average YMCA attendance from the no PR treatment, or as the average performance in the Anonymous Effort Round.

The cost parameter c: We estimate c using the following equation:

$$c = \frac{r_1 + 2r_2 \left(\bar{a}_0 + \bar{\tau}\right)}{\bar{\tau}} \tag{A.19}$$

To see why this equation recovers c, we recall the partial-equilibrium action function from equation (A.8):

$$a^*(\theta; 0) = \frac{\theta/c + \gamma_1^a/c - 2\gamma_2^a \rho^a \bar{a}^0_{pop}/c}{1 - 2\gamma_2^a/c}$$

We next take the expectation of both sides, recalling that we are in the case where $PR = 1$:

$$\mathbb{E}[a|PR = 1] = \frac{\mathbb{E}[\theta/c] + \gamma_1^a/c - 2\gamma_2^a \rho^a \bar{a}_{pop}/c}{1 - 2\gamma_2^a/c}$$

We substitute $\mathbb{E}[\theta/c] = \bar{a}^0_{pop}$ and $\mathbb{E}[a|PR = 1] = \bar{a}_0 + \bar{\tau}$ into the expression above, and solve for c:

$$\begin{aligned}
\bar{a}_0 + \bar{\tau} &= \frac{\bar{a}^0_{pop} + \gamma_1^a/c - 2\gamma_2^a \rho^a \bar{a}_{pop}/c}{1 - 2\gamma_2^a/c} \\
c &= \frac{\gamma_1^a - 2\gamma_2^a \rho^a \bar{a}^0_{pop} + 2\gamma_2^a \left(\bar{a}_0 + \bar{\tau}\right)}{\bar{\tau}} \qquad \text{(A.20)}
\end{aligned}$$

Finally, we substitute the expressions for γ_1^a, γ_2^a, and ρ^a into the equation above, which recovers equation (A.19).

The net image payoff from scaling up public recognition: Using the attendance rules from equations equations (A.8) and (A.10), we estimate each YOTA member's counterfactual attendance when public recognition is scaled up at specified values of γ_1^a, γ_2^a, and ρ^a. We then use the attendances and equation (A.7) to estimate the net-image payoff.

Estimating confidence intervals: Because the parameters are highly nonlinear functions of these empirical moments, we compute confidence intervals without relying on asymptotic normality approximations. Instead, we compute 95 percent confidence intervals for the estimates reported in Tables 1.7 and 1.9 using a percentile-based bootstrap blocked at the individual level.

A.6.2 Characteristic-Signaling Model

Public recognition utility has the form $\nu S^{\theta}(\mathbb{E}[\theta|a]-\rho^{\theta}\bar{\theta}) = \gamma_1^{\theta}(\mathbb{E}[\theta|a]-\rho^{\theta}\bar{\theta})+\gamma_2^{\theta}(\mathbb{E}[\theta|a]-\rho^{\theta}\bar{\theta})^2$, where participants compare the signal of their type, $\mathbb{E}[\theta|a]$, given their action to a multiple of the average type, $\rho^{\theta}\bar{\theta}$. Total utility $U(a;\theta)$ is thus:

$$U(a;\theta) = \theta a - \frac{c}{2}a^2 + y + pa + \gamma_1^{\theta}(\mathbb{E}[\theta|a] - \rho^{\theta}\bar{\theta}) + \gamma_2^{\theta}(\mathbb{E}[\theta|a] - \rho^{\theta}\bar{\theta})^2 \tag{A.21}$$

As with the action-signaling model, we make the simplifying assumption that $p = 0$ unless otherwise noted. We also note that, absent public recognition, the optimal action function $a^*(\theta)$ is given by $a^*(\theta) = \theta/c$.

Equilibrium Behavior

We first characterize the unique separating equilibrium under the D1 criterion. We prove that there exist scalars r_0, r_1, and r_2 given by equations (A.22)-(A.24) below and an equilibrium action function $a^*(\theta) = \frac{\theta}{c-2r_2} + \frac{r_1}{c-2r_2}$ such that $\nu S^{\theta}(\mathbb{E}[\theta|a] - \rho^{\theta}\bar{\theta})$ is given by $r_0 + r_1 a + r_2 a^2$, with $r_2 - \frac{c}{2} < 0$ and $R\left(a^*(\rho^{\theta}\bar{\theta})\right) = 0$.[3] In terms of the structural parameters, we will show that the scalars r_j are given by:

$$r_2 = \frac{1 + 4c\gamma_2^{\theta} - \sqrt{1 + 8c\gamma_2^{\theta}}}{8\gamma_2^{\theta}} \tag{A.22}$$

$$r_1 = \frac{(c - 2r_2)^2\gamma_1^{\theta} - 2r_2\rho^{\theta}\bar{\theta}}{c} \tag{A.23}$$

$$r_0 = -r_1\frac{\rho^{\theta}\bar{\theta}/c + r_1/c}{1 - 2r_2/c} - r_2\left(\frac{\rho^{\theta}\bar{\theta}/c + r_1/c}{1 - 2r_2/c}\right)^2 \tag{A.24}$$

We first show that there exists a linear equilibrium where $a^*(\theta)$ is linear in θ. Note that if all other players had linear equilibrium action functions, then since $\nu S^{\theta}(\mathbb{E}[\theta|a] - \rho^{\theta}\bar{\theta})$ is quadratic in θ, this implies that the reduced-form public recognition function is quadratic. Let this reduced-form PRU be given by $R(a) = r_0 + r_1 a + r_2 a^2$. Given this reduced-form public recognition function, total utility can then be expressed in terms of $R(a)$ as follows:

$$U(a;\theta) = \theta a(\theta) - \frac{c}{2}a(\theta)^2 + y + r_0 + r_1 a(\theta) + r_2 a(\theta)^2 \tag{A.25}$$

We now verify that each type's best response is $a^*(\theta) = \frac{\theta}{c-2r_2} + \frac{r_1}{c-2r_2}$. We do so by taking the F.O.C of equation (A.25) with respect to a:

$$\begin{aligned} 0 &= \theta - ca^*(\theta) + r_1 + 2r_2 a^*(\theta) \\ \Leftrightarrow a^*(\theta) &= \frac{\theta/c + r_1/c}{1 - 2r_2/c} \end{aligned} \tag{A.26}$$

[3] The condition $r_2 - \frac{c}{2} < 0$ ensures that S is quadratic, and that our solutions are well-defined.

We next verify that equations (A.22)-(A.24) map $\nu S^\theta(\mathbb{E}[\theta|a] - \rho^\theta\bar{\theta})$ to $R(a)$. To see this, we begin with $R(a^*(\theta))$ and substitute in equation (A.26):

$$\begin{aligned} R(a^*(\theta)) &= r_0 + r_1 a^*(\theta) + r_2 a^*(\theta)^2 \\ &= r_0 + r_1 \cdot \frac{\theta/c + r_1/c}{1 - 2r_2/c} + r_2 \left(\frac{\theta/c + r_1/c}{1 - 2r_2/c}\right)^2 \end{aligned}$$

The above expression is algebraically equivalent to the following:

$$\begin{aligned} R\left(a^*(\theta)\right) = r_0 + r_1 \frac{\rho^\theta\bar{\theta}/c + r_1/c}{1 - 2r_2/c} + r_2 \left(\frac{\rho^\theta\bar{\theta}/c + r_1/c}{1 - 2r_2/c}\right)^2 \\ + \frac{cr_1 + r_1 r_2 + 2r_2\rho^\theta\bar{\theta}}{(c - 2r_2)^2}(\theta - \rho^\theta\bar{\theta}) + \frac{r_2}{(c - 2r_2)^2}(\theta - \rho^\theta\bar{\theta})^2 \end{aligned} \tag{A.27}$$

The first three terms in the equation above sum to $R\left(a^*(\rho^\theta\bar{\theta})\right)$, which equals 0 if and only if $r_0 = -r_1 \frac{\rho^\theta\bar{\theta}/c + r_1/c}{1-2r_2/c} - r_2\left(\frac{\rho^\theta\bar{\theta}/c + r_1/c}{1-2r_2/c}\right)^2$. This verifies equation (A.24).

From equation (A.27), we see that $R\left(a(\theta)\right)$ maps to $\nu S^\theta(\mathbb{E}[\theta|a] - \rho^\theta\bar{\theta})$ if and only if the following two equations hold:

$$\gamma_1^\theta = \frac{cr_1 + r_1 r_2 + 2r_2\rho^\theta\bar{\theta}}{(c - 2r_2)^2} \tag{A.28}$$

$$\gamma_2^\theta = \frac{r_2}{(c - 2r_2)^2} \tag{A.29}$$

Solving these equations for r_1 and r_2 recovers equations (A.22) and (A.23), respectively. This completes the proof that $a^*(\theta)$ is an equilibrium action function. Since $a^*(\theta)$ is linear in θ, it also defines a separating equilibrium.

Finally, we argue that $a^*(\theta)$ is the unique equilibrium action function. Because the material utility function $\theta a^*(\theta) - \frac{c}{2}a^*(\theta)^2$ satisfies the single-crossing property, i.e., the derivative with respect to $a^*(\theta)$, $\theta - ca^*(\theta)$, is increasing in θ, the results of Mailath (1987) imply that this separating equilibrium must be a unique separating equilibrium.

The Impact of Scaling up Public Recognition

We consider the counterfactual where public recognition is applied to the full population, and restrict attention to the YMCA case. Because we have an approximately continuous strategy space, the equilibrium in the characteristic-signaling model is a separating equilibrium, in which each type's optimal choice of a depends on the structural public recognition function S and on $\bar{\theta}$, but not on any other moments of the distribution of θ. This implies that even though the types that are in the experiment are not representative of those in the population, the equilibrium choice of action of any given type will be the same. The property that a type's choice of action is independent of the distribution of types, beyond $\bar{\theta}$, generally holds for any signaling model with a continuous action space and a utility function that satisfies the single-crossing property (Mailath, 1987).

We thus take the expectation of the optimal attendance rule in equation (A.26) to predict equilibrium attendance $\bar{a}_{eq}$:

$$\bar{a}^{eq}_{pop} = \frac{\bar{a}^0_{pop} + r_1/c}{1 - 2r_2/c}$$

The optimal attendance, $a^*(\theta)$, is the same as in equation (A.26). Below we write it in terms of the structural parameters γ^θ_1, γ^θ_2, and ρ^θ which we use in our simulations that exogenously vary these parameters:

$$a^*(\theta) = \frac{4c\gamma^\theta_2}{1 - \sqrt{1 + 8c\gamma^\theta_2}} \cdot a_0 - \frac{1 + 4c\gamma^\theta_2 - \sqrt{1 + 8c\gamma^\theta_2}}{1 - \sqrt{1 + 8c\gamma^\theta_2}} \cdot \rho^\theta \bar{a}^0_{pop} + \frac{1 - \sqrt{1 + 8c\gamma^\theta_2}}{4c\gamma^\theta_2} \cdot \gamma^\theta_1 \quad \text{(A.30)}$$

Here $\bar{a}^0_{pop} := \bar{\theta}/c$ corresponds to the action taken by the average type, and $a_0 := \theta/c$ corresponds to the action the individual would take absent public recognition.

Mapping the Model Parameters to a Reduced-Form PRU

In Section A.6.2, we derived a mapping for r_0, r_1, r_2 in terms of the structural parameters γ^θ_1, γ^θ_2, and ρ^θ. In this section we derive equations representing γ^θ_1, γ^θ_2, and ρ^θ in terms of r_0, r_1, and r_2:

$$\gamma^\theta_1 = \frac{\sqrt{r_1^2 - 4r_0r_2}}{c - 2r_2}. \quad \text{(A.31)}$$

$$\gamma^\theta_2 = \frac{r_2}{(c - 2r_2)^2} \quad \text{(A.32)}$$

$$\rho^\theta = \frac{\sqrt{r_1^2 - 4r_0r_2} - r_1}{2\bar{a}^0_{pop}r_2} - \frac{\sqrt{r_1^2 - 4r_0r_2}}{c\bar{a}^0_{pop}} \quad \text{(A.33)}$$

To verify these equations, we first note that we recovered equation (A.32) as equation (A.29) in Section A.6.2. We also recovered equation (A.28), which defines γ^θ_1 in terms of the reduced-form parameters and ρ^θ. We thus next verify equation (A.33). To do so, we first note that the optimal action function at $\rho^\theta\bar{\theta}$ equals:

$$a^*(\rho^\theta\bar{\theta}) = \frac{\rho^\theta\bar{\theta}/c + r_1/c}{1 - 2r_2/c}$$

Using $\bar{\theta}/c = \bar{a}^0_{pop}$, we rewrite this as:

$$a^*(\rho^\theta\bar{\theta}) = \frac{\rho^\theta\bar{a}^0_{pop} + r_1/c}{1 - 2r_2/c}$$

We next substitute the above expression into $R\left(a(\rho^\theta\bar{\theta})\right) = 0$:

$$0 = r_0 + r_1 \frac{\rho^\theta \bar{a}^0_{pop} + r_1/c}{1 - 2r_2/c} + r_2 \left(\frac{\rho^\theta \bar{a}^0_{pop} + r_1/c}{1 - 2r_2/c} \right)^2$$

We solve this equation for ρ^θ via the quadratic formula, which yields equation (A.33). Finally by substituting (A.33) into equation (A.28), we recover equation (A.31).

Estimating the Model

The reduced-form public recognition function: We use the same procedure as in the action-signaling model.

The effects of public recognition on performance: We again use the same procedure as in the action-signaling model.

The cost parameter c: We recover the same estimate for c as in the action-signaling model:

$$c = \frac{r_1 + 2r_2\left(\bar{a}^0_{pop} + \bar{\tau}\right)}{\bar{\tau}} \tag{A.34}$$

To see why this equation recovers c, we recall the equilibrium action function from equation (A.26):

$$a^*(\theta) = \frac{\theta/c + r_1/c}{1 - 2r_2/c}$$

We next take the expectation of both sides, recalling that we are in the case where $PR = 1$:

$$\mathbb{E}[a^*(\theta)|PR = 1] = \frac{\mathbb{E}[\theta/c] + r_1/c}{1 - 2r_2/c}$$

We substitute $\mathbb{E}[\theta/c] = \bar{a}^0_{pop}$ and $\mathbb{E}[a|PR = 1] = \bar{a}_0 + \bar{\tau}$ into the expression above, and solve for c:

$$\begin{aligned} \bar{a}_0 + \bar{\tau} &= \frac{\bar{a}^0_{pop} + r_1/c}{1 - 2r_2/c} \\ c &= \frac{r_1 + 2r_2\left(\bar{a}^0_{pop} + \bar{\tau}\right)}{\bar{\tau}} \end{aligned}$$

The net image payoff from scaling up public recognition: Using the optimal attendance rule from equation (A.30), we estimate each YOTA member's counterfactual attendance when public recognition is scaled up at specified values of γ_1^θ, γ_2^θ, and ρ^θ. We then use these values and equation (A.21) to estimate the net-image payoff.

Estimating confidence intervals: As with the action-signaling model, we compute 95 percent confidence intervals for the estimates reported in Tables 1.7 and 1.9 using a percentile-based bootstrap blocked at the individual level.

A.6.3 Incorporating Heterogeneity

Consider heterogeneity in marginal costs, so that the cost of effort is given by $C(a, \xi) = ca^2/2 + \xi a$. For simplicity, assume that $\mathbb{E}[\xi|\theta] = 0$ and that $Pr(\xi + \theta < 0) = 0$. Then the optimal action given a reduced-form PRU $R(a) = r_0 + r_1 a + r_2 a^2$ is

$$a^*(\theta, \xi) = \frac{(\theta - \xi)/c}{1 - 2r_2/c} + \frac{r_1/c}{1 - 2r_2/c} \tag{A.35}$$

and thus

$$\mathbb{E}[a^*(\theta, \xi)|\theta] = \frac{\theta/c}{1 - 2r_2/c} + \frac{r_1/c}{1 - 2r_2/c} \tag{A.36}$$

In other words, the expected action of a person with intrinsic motivation θ remains unchanged. This immediately implies that all of the conclusions derived above for the action-signaling model remain unchanged, since the reduced-form PRU will be quadratic if and only if the structural PRU is quadratic.

Consider now the characteristics-signaling model, where individuals derive utility about the audience's impression of their intrinsic motivation θ, but not the marginal cost ξ. We show that we can microfound a quadratic reduced-form PRU with an approximately quadratic structural PRU. From equation (A.35), note that if $Var[\xi|\theta]$ is sufficiently small, then $\mathbb{E}[\theta|a] = (c - 2r_2)a - r_1 + O(Var[\xi|\theta])$, where terms $O(Var[\xi|\theta])$ are negligible. In Bénabou and Tirole (2006), this linear approximation holds when θ and ξ are distributed normally, and the domain of a is all of $\mathbb{R}$. As long as this linear approximation is valid, the structural PRU in the characteristics-signaling model can again be written as $\nu S(\theta - \rho^\theta \bar{\theta}) = r_0 + r_1 \cdot a^*(\theta) + r_2 \cdot a^*(\theta)^2$, where $a^*(\theta) = \frac{\theta}{c - 2r_2} + \frac{r_1}{c - 2r_2}$.

Appendix B

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

B.1 Additional theoretical results

B.1.1 Shannon model with heterogeneous priors

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

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

1. Consumers choose a joint distribution 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 if and only if $v - p_s - \sigma(\rho t + (1-\rho)\hat{t}) > 0$. We will use $b(\rho) \in \{0, 1\}$ to note whether it is optimal for a consumer to buy given 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. (2019).

Proposition 3. *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 3 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

2.5 and 2.6, 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 3 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 3 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 underestimation, 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.

B.1.2 The Gabaix (2014) model

A second model that allows for a continuous range of cognitive effort is the Gabaix (2014) model. We utilize the binary action extension of the model.[1] 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 2.2.2, 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 $B - (1-m)^2 B$.

The special case $\alpha = 0$ corresponds to our binary attention example in Section 2.2.2. 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 4. *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.

[1] 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.

2. The valuation weight $\theta_{\Xi,\sigma}$ is increasing in $\hat{t}$, with $\theta_{\Xi,\sigma} = 1$ when $\hat{t} = t$. Moreover, $|1 - \theta_{\Xi,\sigma}|$ is decreasing in r.

Like Proposition 3, Proposition 4 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 moprimitives when closed-form solutions exist for the choice m^*, as in the simple examples for $\alpha = 1, 2$ that we summarized above.

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

B.2 General results and proofs about costly attention models

B.2.1 Lemma for revealed valuation weight representation

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

Lemma 1. *Suppose that the probability that consumer i chooses to buy the product given a valuation v, salient price p_s, and opaque price p_o is given by $G(v - p_s, p_o)$, with G increasing in the first argument. 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 1 implies that any attention model that predicts that consumers are more likely to buy when the transparent surplus $v - p_s$ is higher can be represented using the reduced-form attention weight model. The additional statements in the Lemma help provide further structure on the attention weights. For example, when the buying decision is not stochastic, as in the Gabaix (2014) model, the reduced-form valuation weight will not be stochastic either.

B.2.2 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 2. *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 2 implies that the ex-ante expected likelihood of buying is increasing in $u = v - p_s$. However, it does not by itself imply that the ex-ante expected likelihood of buying is increasing in every state, and $\omega = q_o$ in particular. If as u increases, the relative likelihood of buying in $\omega = q_o$ decreases sufficiently quickly, then the likelihood of buying in that state would not decrease. Although we have not confirmed this exhaustively, this seems like an unlikely result. Below, we confirm that the likelihood of buying in state $\omega = q_o$ is increasing in u when attention costs are proportional to (Shannon) entropy reduction.

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

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

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

Since the right-hand side of (B.1) is increasing in both Q and u, Lemma 2 implies that $Pr(buy|\omega = q_o)$ is increasing in u. □

The last result immediately leads to the following:

Proposition 5. *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 3 implies that the probability that the consumer buys the product can be written as $G(v - p_s, p_o)$, where G is increasing in the first argument. Lemma 1 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 6. *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 1 then implies the result. □

Generalization to other cost functions: In general, Proposition 6 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 7. *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. (2019), Q must satisfy

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

with equality if $Q > 0$. Now

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

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 approaches ∞, $Pr(buy|lv, lp_s, lp_o) \to 0$ 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. □

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 7 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.

B.2.3 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,

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

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

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

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

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

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

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

Proposition 8. *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 9. *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

$$u^\dagger = m(u^\dagger)\sigma q_o + (1 - m(u^\dagger))\sigma\bar{q}_o \tag{B.2}$$

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

Then

$$\theta = \frac{\sigma q_o}{u^\dagger} = \frac{q_o}{m(u^\dagger)q_o + (1 - m(u^\dagger))\bar{q}_o} \tag{B.3}$$

Note that θ is a function of m and $u^\dagger$ only, and does not directly depend on stakes. The combination of (B.2) and (B.3) 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 (B.2) 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 $u > \sigma\bar{q}_o$

$$B(\sigma', u') = \int_{\omega \geq u'/\sigma'} (\sigma'\omega - u')\mu(\omega) = \frac{\sigma'}{\sigma}\int_{\omega}(\sigma\omega - u)\mu(\omega) = \frac{\sigma'}{\sigma}B(\sigma, u) \tag{B.4}$$

Since the perceived benefit of increasing m is linear in B, equation (B.4) 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 10. *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 9, 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 1.

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

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

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

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

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

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

B.2.4 Proof of Proposition 3

Proposition 5 establishes that the model has a revealed valuation weight representation. Proposition 7 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. (2019) 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

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

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

$$\rho_0 = \frac{e^{\frac{u-\hat{t}}{\lambda}} - 1}{e^{\frac{u-\hat{t}}{\lambda}} - e^{\frac{u-t}{\lambda}}} \tag{B.5}$$

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

$$\rho_0 = \frac{1-e^{\frac{\hat{t}-u}{\lambda}}}{1-e^{\frac{\hat{t}-t}{\lambda}}} \tag{B.6}$$

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 (B.5), since $e^{\frac{u-t}{\lambda}} < 1$ and $e^{\frac{u-\hat{t}}{\lambda}} > 1$ but is decreasing in $\hat{t}$.[2] 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 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 $\hat{t}$.

[2] 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 $e^{\frac{u-\hat{t}}{\lambda}}$.

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.2.5 Proof of Proposition 4

Proof. Proposition 8 establishes that the model has a revealed valuation weight representation. Proposition 9 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 1, and the fact that the revealed valuation weight representation has $\theta = 1$ when $\hat{t} = t$, this will imply the remaining statement of the proposition.

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

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

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

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

B.3 Proofs of Propositions in the body of the paper

B.3.1 Proof of Proposition 1

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

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

Again by the law of iterated expectations,

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

The first statement of the proposition is therefore equivalent to

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

which holds by the Cauchy-Schwarz inequality. More generally, if 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. □

B.3.2 Proof of Proposition 2

We start with the more general statement.

Proposition 11. *Let Y be a random variable supported on $[\underline{Y}, \bar{Y}]$. Then*

$$Pr(Y > y) \geq \frac{E[(Y - \underline{Y})^2] - (y - \underline{Y})E[Y - \underline{Y}]}{(\bar{Y} - y)(\bar{Y} - \underline{Y})} \tag{B.7}$$

$$Pr(Y < y) \geq \frac{E[(\bar{Y} - Y)^2] - (\bar{Y} - y)E[\bar{Y} - Y]}{(y - \underline{Y})(\bar{Y} - \underline{Y})} \tag{B.8}$$

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. □

We obtain Proposition 2 as a corollary. When $\underline{Y}=0$ and $y=1$, equation (B.7) 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 (B.8) translates to

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

B.4 Predictions 3 and 4 with heterogeneity in attention costs

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

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

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

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

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

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

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

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

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

$$\eta_i = a_0 + a_1\tilde{\theta}_i + \epsilon \tag{B.9}$$

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

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

$$\begin{aligned} Cov[\theta_i, \eta_i(1-\tilde{\theta}_i)] &= Cov[\theta_i, (a_0 + a_1\tilde{\theta}_i)(1-\tilde{\theta}_i)] \\ &= -(a_0 - a_1)Var[\tilde{\theta}_i] - a_1 Cov[\tilde{\theta}_i, \tilde{\theta}_i^2] \end{aligned} \tag{B.10}$$

Now suppose that, as in our data, that $E[\eta_i(1-\tilde{\theta}_i)] = \sigma\frac{dE[\tilde{\theta}_i]}{d\sigma} > 0$. Multiplying (B.9) by $1 - \tilde{\theta}_i$ and taking expectations yields

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

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

$$a_0 > \frac{a_1(E[\tilde{\theta}_i^2] - E[\tilde{\theta}_i])}{(1 - E[\tilde{\theta}_i])} > a_1 \tag{B.11}$$

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

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

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

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

$$\begin{aligned} -(a_0 - a_1)Var[\tilde{\theta}_i] - a_1 Cov[\tilde{\theta}_i, \tilde{\theta}_i^2] &\leq -(a_0 - a_1)Var[\tilde{\theta}_i] + \frac{a_0}{\bar{\theta}}(\bar{\theta} - E[\tilde{\theta}_i])Var[\tilde{\theta}_i] \\ &< -(a_0 - a_1)Var[\tilde{\theta}_i] + a_0 Var[\tilde{\theta}_i] \\ &= a_1 Var[\tilde{\theta}_i] \end{aligned} \tag{B.12}$$

Putting (B.11) and (B.12) together to sign (B.10), we see that: (i) if a_1 is positive, then both terms of (B.9) must be negative by (B.11). On the other hand, (ii) if a_1 is negative, then the expression in (B.10) must be negative by (B.12).

B.4.1 Simple example with binary attention costs

For concreteness, consider the simple example in Section 2.2.2. In this example, if $\theta_i < 1$ then either $\frac{d\theta_i}{d\sigma} = 0$ or $\frac{d\theta_i}{d\sigma} = \frac{\lambda_i}{\sigma^2 tr} = (1-\theta_i)/\sigma$. Similarly, if $\theta_i \geq 1$ then either $\frac{d\theta_i}{d\sigma} = 0$ or $\frac{d\theta_i}{d\sigma} = (1-\theta_i)/\sigma$. Thus the elasticity of misreaction is either $\eta_i = 0$ or $\eta_i = 1$. Simple algebra reveals that

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

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

B.5 Theories of bounded rationality inconsistent with our predictions

If consumers hold incorrect beliefs, or lack the financial literacy to integrate the opaque price into their final estimate, then their mistakes should be unresponsive to stakes. Eliciting beliefs or financial literacy is a complementary test of this possibility, and one that we conduct. However, these are one-sided tests, as evidence of incorrect beliefs or financial illiteracy among some consumers does not rule out costly attention as an important mechanism. Moreover, because "behavioral" consumers may not necessarily act on the answers they give to abstract beliefs questions (e.g., Bernheim and Taubinsky, 2018b), we consider our predictions about behavior a more satisfactory test.

The commonly used attention models in which attention is exogenous to stakes but responds to non-pecuniary stimuli, such as those summarized in DellaVigna (2009), are ambiguous about Prediction 2, and are not consistent with the other four predictions. Beyond a qualitative assessment of these simpler attention models, our tests also serve as a critical examination of the extent to which exogenous attention might be a reasonable approximation. For example, even if consumers do exert more mental effort to consider sales taxes for big purchases like cars, the endogeneity of mental effort to taxes may be negligible for most goods that consumers purchase. Thus our measures of the *attention elasticities* are a key quantitative input for enriching the modeling of attention.[3]

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

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

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

Finally, the predictions differentiate costly attention models from two starkly different theories about the interaction between incentives, mental effort, and decision quality. First, Kahneman (2003) and others argue that even if mental effort increases with incentives, this does not have to translate to better decisions.[4] Second, Ariely et al. (2009b) argue for and test the hypothesis—building on a line of reasoning dating to Kahneman (1973)—that mental effort is not fully controlled because it is influenced by affective states of arousal, and thus higher incentives may lead mental effort to become misdirected, which would decrease decision quality. For example, a consumer considering a big-ticket purchase may become distracted by the pressure and gravity of the decision, and consequently omit considering sales taxes.

B.6 Relation to sales tax literature

Table B.6.1 summarizes the prior literature on misreaction to sales taxes. We mark a cell with "Yes" if the prediction is confirmed, with "Unclear" if the evidence is inconclusive, and we leave the cell blank if the the prediction is not tested. We include the following papers: Chetty, Looney, and Kroft (2009, **CK**); Goldin and Homonoff (2013, **GH**); Feldman and Ruffle (2015, **FR**); Taubinsky and Rees-Jones (2018, **TRJ**); Feldman, Goldin, and Homonoff (2018, **FGH**); and this paper (**MT**)

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

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

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

Rows 4 and 5 summarize evidence on how misreaction changes as the stakes increase. As discussed in Section 2.2.4, we verify in our paper that average misreaction decreases both as tax rates increase and as the posted prices increase. TRJ find that average underreaction is lower in their high-tax treatment than in their standard-tax treatment, but they generally lack statistical power to find differences across posted prices, and can only detect a difference between small prices and all other prices in the high-tax-rate condition. Feldman and Ruffle (2015) provide an indirect test of the effect of changing prices in their Table 5, but caveat that there are potential confounds and thus "do not push these particular results too far."

[4]See, e.g., Enke and Zimmermann (2019) for evidence in line with this conjecture.

This test is not a main focus of their paper. Like TRJ, Feldman et al. (2018) vary tax rates in a between-subject design. However, Feldman et al. (2018) report results that are statistically imprecise and consistent both with the null of exogenous attention as well as the results we report in this paper. The Feldman et al. (2018) is not set-up to identify $E[\theta]$ or $Var[\theta]$, and otherwise features the same weaknesses as TRJ, which we discuss below.

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

Table B.6.1: Summary of prior literature

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

B.6.1 Detailed comparisons to TRJ

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

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

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

The differences in statistical power are likely a consequences of two design differences. First, TRJ have more "noise" in any given decision created by the apparently more confusing BDM mechanism. Second, TRJ do not fully randomize the order of all the tax conditions within-subject. Consequently, they must use a "control" arm to estimate the order effects, which decreases statistical power.

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

Appropriate designs for applying our econometric methods. TRJ create data that violates the key assumption underlying our approach to individual differences, stated in Section 2.5.1: 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 TRJ would lead to correlated measurement error and violate our assumption. Consequently, our new methods for bounding $Var[\theta]$ are not applicable to experimental datasets such as those in TRJ or Feldman et al. (2018) that do not vary the order of tax environments.

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

B.7 Counterfactual demand curve construction

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 $\tilde{D}_{jB}(p_n) := \sum_i \left[\frac{0.15-\tau_i}{0.15} D_{ijA}(p_n) + \frac{\tau_i}{0.15} D_{ijA}(p_{n+1})\right]$, 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.[5] For store C, if $\tau_{iC} < 0.15$ we use the same interpolation as in the store B counterfactual demand; if $\tau_{iC} > 0.15$, we calculate $\tilde{D}_{jC}(p_n) := \frac{0.30-3\tau_i}{0.15} D_{jA}(p_{n+1}) + \frac{3\tau_i-0.15}{0.15} D_{jA}(p_{n+2})$. To construct $\tilde{D}_{jC}(p_9)$, we use the self-reported maximum willingness to pay to see if individuals willing to purchase at price p_{10} would be willing to purchase at price $1.15p_{10}$.

B.8 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, $f'(x) = -xf(x)$:

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

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

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 instrument $X \in \{0, 1\}$. 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,[6] 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 2.1b and 2.2 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 2.1a we get $0.25 \times 1.20 + 0.75 \times 0.25 = 0.78$. When we average the two values in table 2.1b we get $0.25 \times .86 + 0.75 \times 0.76 = 0.785$. These results suggest that there is not a significant bias.

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

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.

B.9 Point estimates and confidence intervals for Figures 2.2a and 2.2b

Table B.9.1: Average revealed valuation weights in Figure 2.2a

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

Table B.9.1 presents the estimates for $E[\theta]$ and average tax owed displayed in figure 2.2a. 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 (2.4) for prices below the cutoff. Standard errors are clustered at the subject level.

Table B.9.2: Average revealed valuation weights in Figure 2.2b

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

Table B.9.2 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 (2.4), described in equation (2.5). Standard errors are clustered at the subject level.

B.10 Relationship between average revealed valuation weights and marginal utility of money

One potential concern with our estimation procedure in Section 2.4.3 is that the set of consumers on the margin at each price mechanically have different product valuations. If the valuation for the product is correlated with attention, this would confound our results about how average valuation weights covary with price.

In this appendix, we present additional evidence that our results are robust to this concern. Specifically, we utilize the split-sample techniques described in Section 2.5.1 to analyze whether, holding price constant, participants with lower marginal utility of money, and hence higher propensity to pay for the products in our experiment, are more attentive to taxes.

First, we use one product to divide consumers into two groups. The *low* marginal utility of money (MU) group consists of those with above-median values of willingness to pay in the no-tax environment and the *high* MU group consists of those with below-median values of willingness to pay. The intuition is that participants who have a higher valuation of the product were willing to forego more money to obtain the product, or equivalently have a lower marginal utility of money. We then estimate equation (2.5) *on the other two products* to estimate average valuation weights at each of the five price sets for the high and low MU groups. We then repeat this process for the other two products, and average the resulting estimates to obtain average valuation weights at each price pair P_n, $E[\theta_{ijk}|k = K, p \in P_{n,}]$, for both the low and high groups. To hold prices constant, we compute separate average valuation weights for each price-pair.

More concretely, we index each of the three products seen by each person with $j \in \{1,2,3\}$. First, we start with $j = 1$ and we split the sample into two groups: those with p^*_{ijA} in the top 50% of the population and those with p^*_{ijA} in the bottom 50% of the population, where p^*_{ijA} denotes the WTP for product j in store A.[7] We then use decisions in the *other two* products to estimate the average valuation weights $E[\theta_{ijk}|k = K, LowMU_{i1}, p \in P_n, j \neq 1]$ for each price pair P_n using equation (2.5), where $K \in \{B, C\}$ and $LowMU_{i1}$ is defined as an indicator for the low MU group. We repeat the procedure twice using products 2 and 3 to generate $LowMU_{i2}$ and $LowMU_{i3}$, and estimate $E[\theta_{ijk}|k = K, LowMU_{i2}, p \in P_n, j \neq 2]$ and $E[\theta_{ijk}|k = K, LowMU_{i3}, p \in P_n, j \neq 3]$. Finally, we average the estimates from each of these three iterations to get an overall average estimate of θ_{ijk} for those in the high and low groups:

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

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

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

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

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

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

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

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

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

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

[7] Section 2.5.1 details the methodology used to construct p^*_{ijA}.

B.11 Covariates of attention 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 (2.4) separately for the above-median and below-median tax groups to create figures analogous to Figure 2.2a. We include state fixed effects to capture some of the geographic variation.

Figure B.11.1 presents the results. Panel (a) uses the main sample and is identical to Figure 2.2a. Panel (b) restricts to participants with a local sales tax rate above 7.00%, the median of the sample. Panel (c) restricts to participants with a local sales tax rate at or below 7.00%. The results provide some evidence that participants in high sales tax locations have lower revealed valuation weights than those from low sales tax locations.

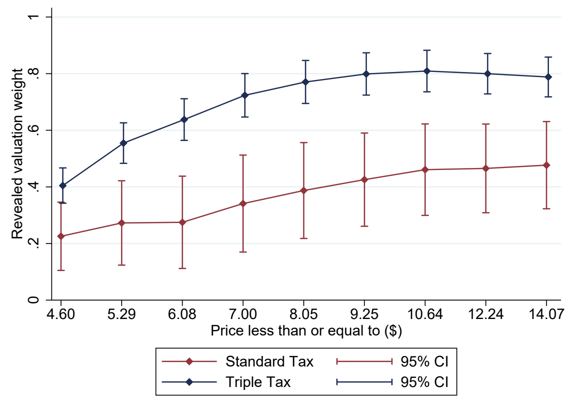
Table B.11.1 presents store-specific estimates of average θ by tax group using all prices. These estimates match the rightmost points of the series in Figure B.11.1. The third column presents the difference, which is statistically significant for store C.

We next move from splitting the sample at the median tax rate to dividing the sample into quartiles of local tax rates. The median tax in our sample is 7.00% and the interquartile range is 6.00%-8.15%. Table B.11.2 presents the results. The differences in tax rates are statistically significant for store C ($p = 0.001$) and just outside the 10% significance level for store B ($p = 0.105$). For both stores B and C, the lowest tax quartile sample has the highest average valuation weights, 0.62 and 0.86, respectively. We similarly see that the highest tax quartile group has the lowest average revealed valuation weights of 0.34 and 0.57, respectively.

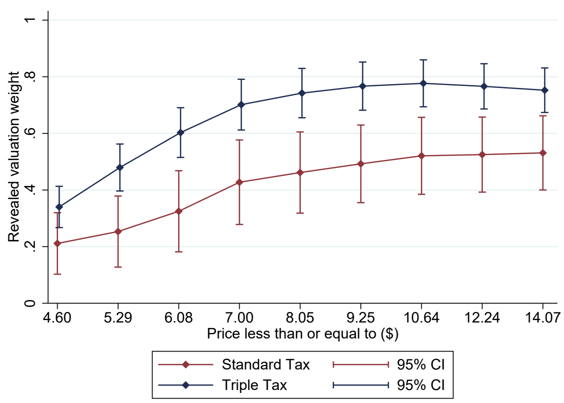
Although this association between valuation weights and local tax rates appears to be opposite to the relationship seen in Figure 2.2a, we note that it is likely a confounded test of costly attention models because local tax variation could be related to a number of differences in geography, including consumers' views and preferences about tax rates, or consumers' attention to tax rates. For example, higher-tax rate jurisdictions tend to be more urban and in more liberal states, and the observed differences in average valuation weights may reflect sorting into urban versus rural jurisdictions. Moreover, to the extent that states and counties set their tax rates that follow some version of a standard inverse-elasticity rule, this will tend to lead to higher tax rates being set in places where consumers tend to be least attentive to taxes (a "reverse-causality" mechanism).

Figure B.11.1: Average revealed valuation weight for posted prices at or below a cutoff

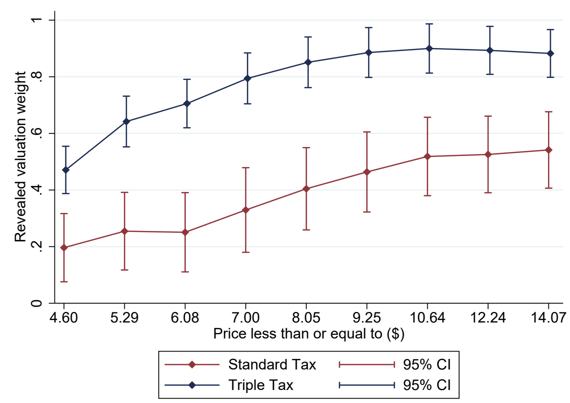
(a) Main sample



(b) Local sales tax rate above 7.00%



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



Panels (b) and (c) of Figure B.11.1 recreate Figure 2.2a, restricting to participants above and below the median local sales tax rate respectively.

Table B.11.1: Average revealed valuation weights by tax group

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

Table B.11.1 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 (2.4), interacting the covariate with price and tax. Standard errors are clustered at the subject level.

Table B.11.2: Average revealed valuation weights by tax rate quartile

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

Table B.11.2 presents store-specific estimates of $E[\theta]$ by tax rate 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 7.00% and the interquartile range is 6.00%-8.15%. The results are estimated using equation (2.4), interacting the covariate with price and tax. Standard errors are clustered at the subject level.

B.11.1 Demographics

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

Political party: Table B.11.3 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 (39.4% of our sample).[8] 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, 0.29]).

[8] 30.6% of participants self-identify as independent and 8.9% of participants self-identify as other.

Education: Table B.11.4 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]). In Table B.11.5, we similarly split the sample into those with advanced degrees versus those without advanced degrees. We again do not detect statistically significant differences in average valuation weights between the groups.

Income: Table B.11.6 presents average θ estimates for each income quartile. Individuals in the top income quartile have self-reported annual income above $80000, in the second quartile from $49000-$80000, in the third quartile from $25000-$49000, and in the bottom quartile below $25000.

All quartiles have average θ_C point estimates in the 0.77-0.79 range. Individuals in the top three quartiles have average θ_B estimates in the 0.39-0.52 range, while individuals in the bottom quartile have an average θ_B of 0.52 (95% CI: [0.22, 0.81]). 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.

Beliefs: Table B.11.7 presents average θ estimates separately for (i) participants who exactly know their local tax rate (51.0% of our sample), (ii) participants who know their local tax rate with one percentage point but do not know it exactly (31.0% of our sample), and (iii) participants who do not know their tax rate within one percentage point (18.0% of our sample). The means of participants' estimates of their sales tax rates in these three groups are 7.08%, 7.46%, and 8.41%, respectively. Compared to participants who do not know their local tax rate within one percentage point, we see that participants with exact knowledge have higher average θ_B (0.54 vs. 0.54, 95% CI for difference: [−0.21, 0.77]) and θ_C (0.85 vs. 0.61, 95% CI for difference: [0.03, 0.45]), though these differences are only statistically significant in the triple-tax environment. The results provide some evidence that participants with less accurate knowledge about their local tax rates underreact to taxes more on average than do participants with more accurate knowledge about them.

Table B.11.3: 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, 0.87]
(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 B.11.3 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 (2.4), interacting the covariate with price and tax. Standard errors are clustered at the subject level.

Table B.11.4: 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 B.11.4 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 (2.4), interacting the covariate with price and tax. Standard errors are clustered at the subject level.

Table B.11.5: Average revealed valuation weights by education: graduate school versus no graduate school

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

Table B.11.5 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. Graduate includes participants with master's degrees or more advanced degrees. The results are estimated using equation (2.4), interacting the covariate with price and tax. Standard errors are clustered at the subject level.

Table B.11.6: Average revealed valuation weights by income quartile

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

Table B.11.6 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 (2.4), interacting the covariate with price and tax. Standard errors are clustered at the subject level.

Table B.11.7: Average revealed valuation weights by knowledge of local sales tax rate

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

Table B.11.7 presents store-specific estimates of $E[\theta]$ by participants' knowledge of their local tax rate. 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 first row includes participants who know their local sales tax rate exactly, the second row includes participants who have an error in their beliefs of less than one percentage point, and the third row includes participants who do know their local sales within one percentage point. The results are estimated using equation (2.4), interacting the covariate with price and tax. Standard errors are clustered at the subject level.

B.12 Alternative construction of proxies for valuation weights

Figure B.12.1: Average revealed valuation weight by posted price

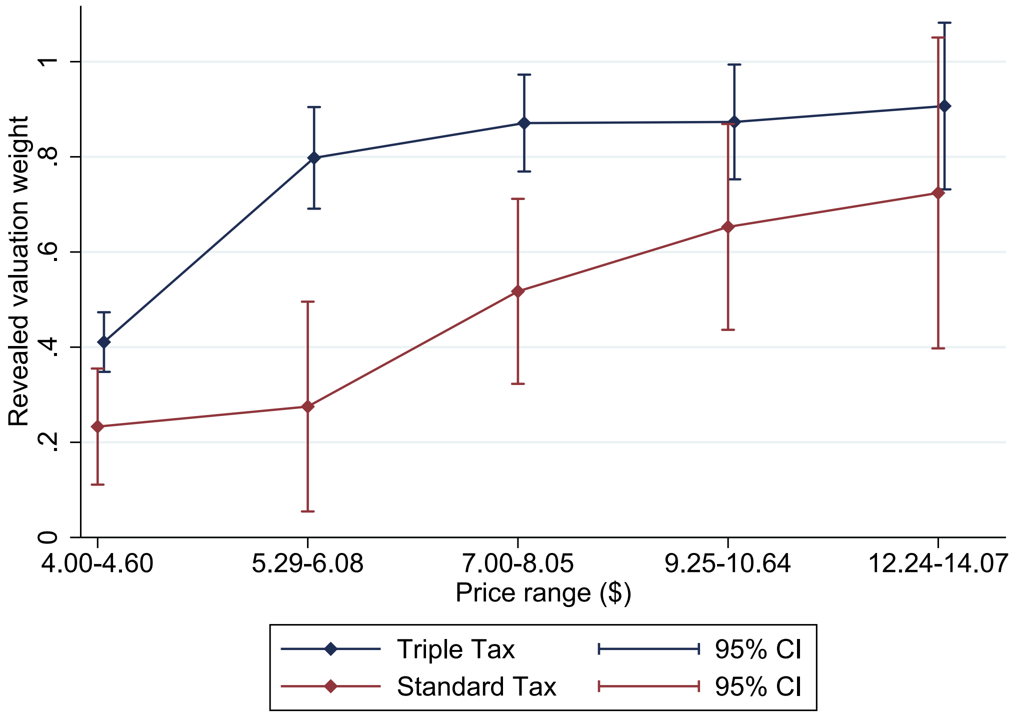


Figure B.12.1 presents store-specific estimates $E[\theta]$ by the posted price. θ is defined as the revealed valuation weight that consumers place on the sales tax, with $\theta = 0$ corresponding to complete neglect of the tax and $\theta = 1$ corresponding to the equal weight of the tax and salient price. Each point is estimated using equation (2.4) for the specified posted prices. Standard errors are clustered at the subject level.

Table B.12.1: Average revealed valuation weights by group: using a 50th percentile cutoff

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

Table B.12.1 repeats Table 2.1a with an alternative split of consumers into high and low valuation weight groups In this table high valuation weight individuals are those with $F(\hat{\theta}_{ijB}) > 0.50$ and low valuation weight individuals are those with $F(\hat{\theta}_{ijB}) \leq 0.50$.

Table B.12.2: Average revealed valuation weights by valuation weight group: using an 80th percentile cutoff

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

Table B.12.2 repeats Table 2.1a with an alternative split of consumers into high and low valuation weight groups. In this table high valuation weight individuals are those with $F(\hat{\theta}_{ijB}) > 0.80$ and low valuation weight individuals are those with $F(\hat{\theta}_{ijB}) \leq 0.80$.

Table B.12.3: Average revealed valuation weights by valuation weight group: using an 85th percentile cutoff

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

Table B.12.3 repeats Table 2.1a with an alternative split of consumers into high and low valuation weight groups In this table high valuation weight individuals are those with $F(\hat{\theta}_{ijB}) > 0.85$ and low valuation weight individuals are those with $F(\hat{\theta}_{ijB}) \leq 0.85$.

Table B.12.4: Average revealed valuation weights by adjustment group: using a 50th percentile cutoff

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

Table B.12.4 repeats Table 2.1b with an alternative split of consumers into high and low adjustment groups. For this table high adjustment individuals are those with $F(\hat{\theta}_{i1C} - \hat{\theta}_{i1B}) > 0.50$ and low adjustment individuals are those with $F(\hat{\theta}_{i1C} - \hat{\theta}_{i1B}) \leq 0.50$.

Table B.12.5: Average revealed valuation weights by adjustment group: using a 20th percentile cutoff

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

Table B.12.5 repeats Table 2.1b with an alternative split of consumers into high and low adjustment groups. For this table high adjustment individuals are those with $F(\hat{\theta}_{i1C} - \hat{\theta}_{i1B}) > 0.20$ and low adjustment individuals are those with $F(\hat{\theta}_{i1C} - \hat{\theta}_{i1B}) \leq 0.20$.

Table B.12.6: Average revealed valuation weights by adjustment group: using a 15th percentile cutoff

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

Table B.12.6 repeats Table 2.1b with an alternative split of consumers into high and low adjustment groups. For this table high adjustment individuals are those with $F(\hat{\theta}_{i1C} - \hat{\theta}_{i1B}) > 0.15$ and low adjustment individuals are those with $F(\hat{\theta}_{i1C} - \hat{\theta}_{i1B}) \leq 0.15$.

B.13 Replication of results restricting to participants with nearly-accurate beliefs and high computational ability

Do participants know their true sales tax rate, and if not, are incorrect beliefs a mechanism driving the results? Consistent with Chetty et al. (2009) and Taubinsky and Rees-Jones (2018), we find that participants generally have correct beliefs: 51.0% of our sample know their tax rate exactly, 70.3% within 0.5 percentage points, and 82.0% within one percentage point.[9] We also do not find any evidence of systematic underestimation of tax rates. The mean of participants' estimates of their sales tax rates is 7.44%, which is negligibly higher

[9]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.

than the actual mean of 7.24%. We refer to the 70.3% who know their tax within 0.5 percentage points as the "nearly-accurate beliefs" sample.

Another potential mechanism driving the results is the inability of participants to compute the sales tax they would need to pay for an item. We test for this mechanism by asking participants to report how much sales tax they would owe for an \$8.00 item purchased in their city of residence. 44.1% of participants are able to calculate their tax burden within \$0.01, and 62.9% are able to compute their tax burden within \$0.05.[10]

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 B.13.1 and B.13.2 present the results.

We next repeat our individual differences analysis, restricting to the "nearly-accurate beliefs and computation" sample. Tables B.13.1-B.13.3 recreate tables 2.1a-2.2 restricting to this sample. Consistent our main results, the low valuation weight group exhibits a larger increase in the revealed valuation weights than the high valuation weight group when tax rates are tripled. The adjustments and their difference are similar in magnitude to our main sample results.

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

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

Figure B.13.1: Average revealed valuation weight for posted prices at or below a cutoff: nearly-accurate beliefs subsample

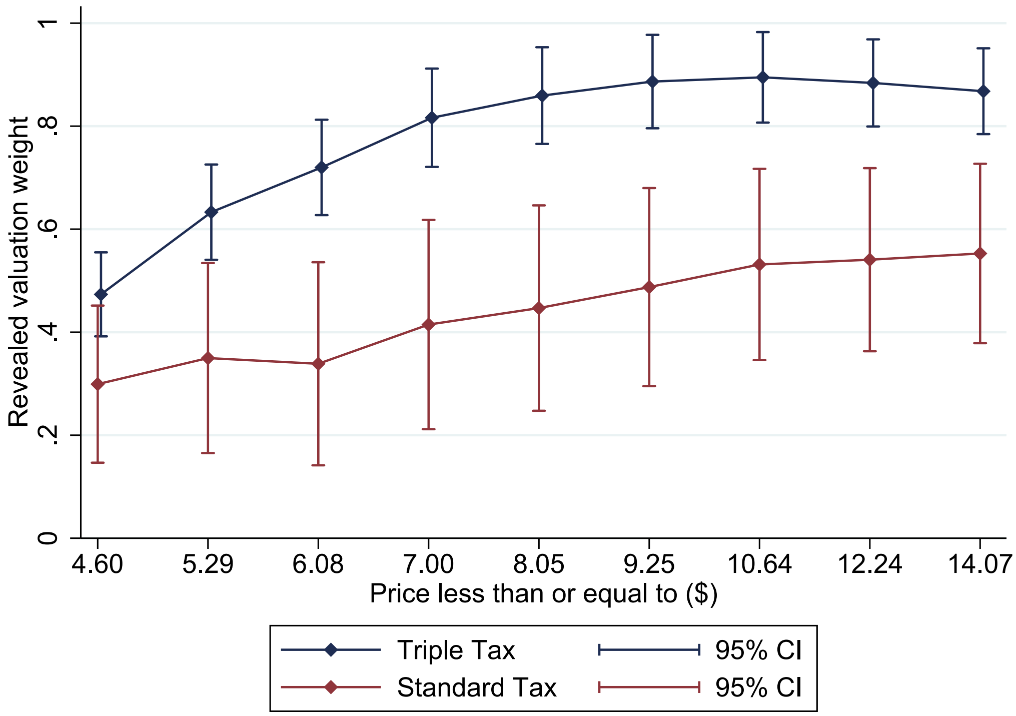


Figure B.13.1 recreates Figure 2.2a, restricting to the 70.3% of the main sample who could identify their local sales tax rate within 0.5 percentage points.

Figure B.13.2: Average revealed valuation weight for posted prices at or below a cutoff: restricting to participants with strong computational ability

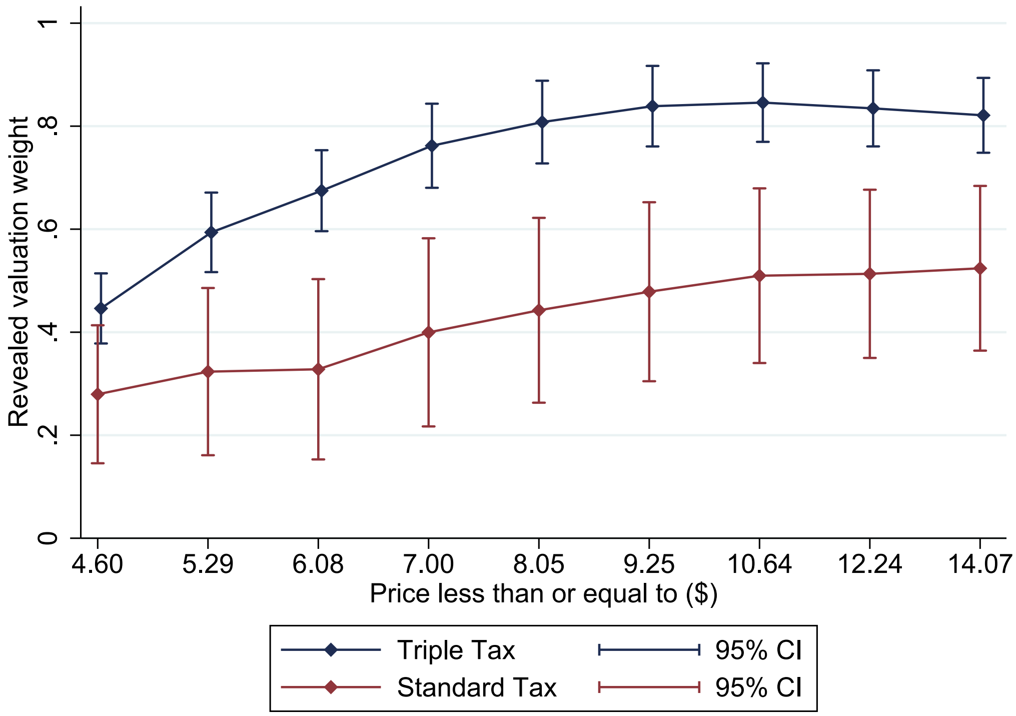


Figure B.13.2 recreates Figure 2.2a, restricting to the 62.9% of the main sample who could identify their tax burden within $0.05 on an $8.00 item purchased in their city of residence.

Table B.13.1: 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	0.40
	[0.85, 1.37]	[1.18, 1.45]	[0.24, 0.55]
(2): Low valuation wgt.	0.34	0.73	0.20
	[0.13, 0.55]	[0.64, 0.83]	[−0.01, 0.42]
(3): (1) − (2)	0.77	0.58	−0.19
	[0.47, 1.08]	[0.44, 0.73]	[−0.44, 0.06]
(4): Full sample	0.56	0.88	0.33
	[0.37, 0.74]	[0.79, 0.97]	[0.19, 0.47]

Table B.13.1 repeats Table 2.1a, restricting to the 59.9% of the main sample who could identify their local sales tax rate within 0.5 percentage points and compute the sales tax they would owe for an $8.00 item purchased in their city of residence within $0.05.

Table B.13.2: 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	0.48	0.43
	[0.23, 0.63]	[0.20, 0.76]	[0.28, 0.58]
(3): (1) − (2)	0.86	0.08	−0.39
	[0.76, 0.96]	[−0.04, 0.21]	[−0.62, −0.17]
(4): Full sample	0.56	0.88	0.33
	[0.37, 0.74]	[0.79, 0.97]	[0.19, 0.47]

Table B.13.2 repeats Table 2.1b, restricting to the 59.9% of the main sample who could identify their local sales tax rate within 0.5 percentage points and compute the sales tax they would owe for an $8.00 item purchased in their city of residence within $0.05.

Table B.13.3: 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 B.13.3 repeats Table 2.2, restricting to the 59.9% of the main sample who could identify their local sales tax rate within 0.5 percentage points and compute the sales tax they would owe for an $8.00 item purchased in their city of residence within $0.05.

B.14 Replication of main results without excluding study participants failing comprehension questions or violating monotonicity

In our primary analyses we exclude 255 respondents who incorrectly answered one or more of the comprehension questions and an additional 47 respondents who had monotonicity violations within a price list. Figure B.14.1 repeats Figure 2.2a including these 302 participants. We again find strong evidence for Prediction 1, indicating that poor computational ability was not the sole mechanism driving consistency with the prediction. The estimates are of smaller magnitude than the full sample results, but are consistent with the theory which predicts average valuation weights are increasing in the absolute size of the tax. Using all prices we estimate an average revealed valuation weight of 0.36 (95% CI [0.22, 0.51]) for the standard tax environment in the restricted sample compared to 0.48 (95% CI [0.32, 0.63]) in the main sample. Similarly, we estimate an average revealed valuation weight of 0.67 (95%

CI [0.60, 0.74]) for the triple tax environment in the restricted sample, which is only slightly lower than the estimate in the main sample 0.79 (95% CI [0.72, 0.86]).

Tables B.14.1-B.14.3 replicate tables 2.1a-2.2 including the respondents who failed the comprehension checks. We still exclude participants with monotonicity violations, as our estimation procedure in Section 2.5.1 assumes monotonic preferences in estimating a willingness-to-pay.

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

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

Including participants who failed comprehension checks leads to a lower variance bound on adjustment (0.32, 5% confidence bound of 0.18) than the bound of (0.86, 5% confidence bound of 0.31) in our main sample. Additionally, we estimate an upper bound on $\underline{\Delta}$ to be 0.15 (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 B.14.1: Average revealed valuation weight for posted prices at or below a cutoff: including participants who fail comprehension checks

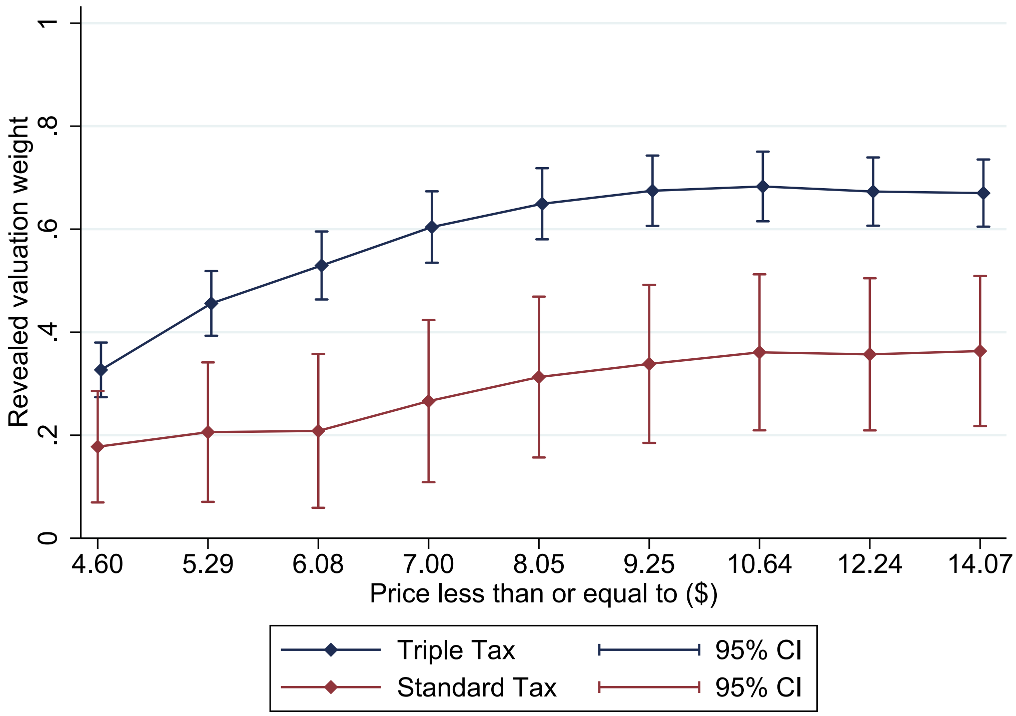


Figure B.14.1 recreates figure 2.2a, including the 302 participants who failed comprehension checks or had monotonicity violations in purchase decisions.

Table B.14.1: 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, 0.52]	[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 B.14.1 repeats Table 2.1a including 255 participants who were excluded from our main sample solely for failing our comprehension check.

Table B.14.2: 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 B.14.2 repeats Table 2.1b including 255 participants who were excluded from our main sample solely for failing our comprehension check.

Table B.14.3: Bounds on the dispersion of revealed valuation weights: including participants who fail comprehension checks

	Standard	Triple	Standard-Triple
Variance (Lower Bound)	0.72	0.72	0.32
	[0.51]	[0.59]	[0.18]
Supremum (Lower Bound)	2.27	1.73	0.15
	[1.70]	[1.55]	[−0.06]

Table B.14.3 repeats Table 2.2 including 255 participants who were excluded from our main sample solely for failing our comprehension check.

B.15 Replication of main results excluding participants who always or never buy a product in at least one store

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

Table B.15.1 presents the results for average valuation weights. The estimates are of smaller magnitude than the full sample results, but are consistent with our prediction that average valuation weights are increasing in the absolute size of the tax. Using all prices we estimate an average revealed valuation weight of 0.36 (95% CI [0.22, 0.50]) 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 0.60 (95% CI [0.53, 0.67]) for the triple tax environment in the restricted sample, which is only slightly lower than the estimate from our main sample, 0.79 (95% CI [0.72, 0.86]).

Tables B.15.2-B.15.3 replicate Tables 2.1a-2.1b excluding the respondents who either always buy or never buy at least one product in at least one store. As with our main results, the low valuation weight group exhibits a larger increase in the revealed valuation weights than the high valuation weight group when tax rates are tripled (0.20 vs. 0.28; 95% CI for difference [−0.27, 0.13]). The adjustments and their difference are similar in magnitude to our main sample results (0.16 vs. 0.39; 95% CI for difference [−0.43, −0.04]).

When dividing consumers by adjustment group, the estimates are also broadly similar: consumers in the low adjustment group increase their valuation weights by an average of 0.15 (95% CI [−0.02, 0.31]) compared to 0.01 (95% CI [−0.15, 0.17]) in our main sample. Similarly, those in the high adjustment group increase their revealed valuation weights by an average of 0.28 (95% CI [0.16, 0.40]) compared to 0.43 (95% CI [0.30, 0.55]) in our main sample. Consistent with our main results, we find that consumers in the low adjustment group have higher valuation weights in both the standard tax regime (0.47 vs. 0.28; 95% CI for difference [−0.02, 0.43]) and in the triple tax regime (0.62 vs. 0.55; 95% CI for difference [−0.03, 0.17]). In summary, we find that the estimates for this subsample conform to Predictions 1-4.

Table B.15.1: Average revealed valuation weights excluding those who always or never buy

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

This table presents store-specific estimates of the average valuation weight. Row (1) presents estimates excluding 726 individuals who either always choose to purchase the product or never choose to purchase the product in a given store, while row (2) presents results for the full sample of 1534 individuals. 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 results are estimated using equation (2.4). Standard errors are clustered at the subject level.

Table B.15.2: Average revealed valuation weights by group: excluding those who always or never buy

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

This table repeats Table 2.1a excluding 726 individuals who either always choose to purchase the product or never choose to purchase the product in a given store.

Table B.15.3: Average revealed valuation weights by adjustment group: excluding those who always or never buy

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

This table repeats Table 2.1b excluding 726 individuals who either always choose to purchase the product or never choose to purchase the product in a given store.

B.16 Order Effects

A potential concern with our within-subject experimental design is that purchase decisions could be influenced by the order in which the nine purchase decisions are presented to consumers. For example, individuals might be more likely to buy in store A when store A preceded by store B rather than comes after store B, since in the former scenario store A seems like a particularly good deal. In Table B.16.1, we test four potential order effects. First, we examine whether the tax environment first shown to consumers impacts their buy probability. We test for this effect via the following model:

$$1 - Pr(buy_{ijk}|p) = \Phi\left(\frac{\alpha_j + \ln(p) + \bar{\theta}_B \ln(1+\tau_{ik}) \cdot I(k=B) + \bar{\theta}_C \ln(1+\tau_{ik}) \cdot I(k=C)}{\sigma_j} + \frac{\gamma^B First_i^B + \gamma^C First_i^C}{\sigma_j}\right) \tag{B.13}$$

This model modifies equation (2.4) 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 (B.13), using $First_{ij}^k$ instead of $First_i^k$:

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

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

For our third test, we examine whether the last store shown to consumers for a product affects their purchase decision. We construct indicator variables $Last_{ij}^k$ which equal one if the consumer's last purchase decision for product j occurred in store k and equals zero otherwise. We then repeat equation (B.13), using $Last_{ij}^k$ instead of $First_i^k$:

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

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

For our fourth test, we construct indicator variables for each possible combination stores A, B, and C were presented to consumer i for product j. We then estimate the following model for $1 - Pr(buy_{ijk}|p)$ for $\kappa_1, \kappa_2, \kappa_3 = \{A, B, C\}$:[11]

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

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

As a test of whether attention is altered by the within-subject nature of our design, in Table B.16.2 we report estimates of average θ using the first $N = 1, 2, \ldots 9$ decisions that consumers make. Although the results are noisy for the standard tax condition for low values of N, the triple tax condition provides us with greater statistical power, and shows that there is little variation in the estimates of average θ when we use only initial decisions or all decisions.

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

Table B.16.1: 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 B.16.1 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 (2.4). 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).

Table B.16.2: Average revealed valuation weights using the first N purchase decisions

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

Table B.16.2 presents store-specific estimates of $E[\theta]$ for the first N purchase decisions made by participants. 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 final row includes all nine purchases decisions, and matches the estimates from row (4) in Table 2.1a. The results are estimated using equation (2.4), interacting the covariate with price and tax. Standard errors are clustered at the subject level.

B.17 Comparison of demand curves to Amazon.com prices

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

While our experiment is not designed to check consumers' awareness of prices at other online stores, we do observe how frequently consumers are willing to purchase above the

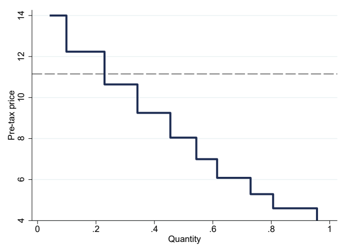
Amazon.com prices. In Figure B.17.1, we plot product–specific demand curve in the no-tax environment compared to the prices listed on Amazon.com near the time of our experiment. These prices range from \$7.73 to \$12.99, which exceed five and nine, respectively, of the prices on our MPL. Depending on the product, approximately 10-30% of consumers choose to buy the product at a price above the Amazon.com price.

There are several caveats in comparing the consumer demand curves to the Amazon.com prices. First, purchasing a product on Amazon.com often requires consumers to pay additional costs such as shipping fees or, in select states, sales taxes.[12] Second, the Amazon.com prices we report are from February 2015, as documented in Taubinsky and Rees-Jones (2018). They may vary over time or by geographic region. Third, we only report the price available on Amazon.com, even though consumers could also buy the products from other online or physical stores.

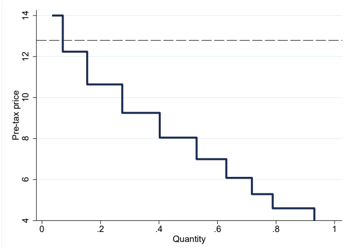
[12] At the time of our experiment (September 2016), Amazon did not collect sales taxes from most states, which is why we choose to compare Amazon.com prices to the demand curves in the no-tax store.

Figure B.17.1: Product-specific demand curves from the no-tax environment compared to Amazon.com prices

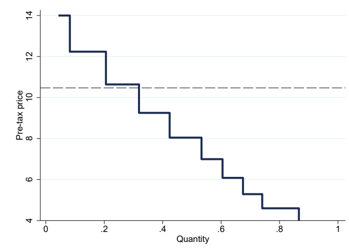
(a) Energizer AA Batteries



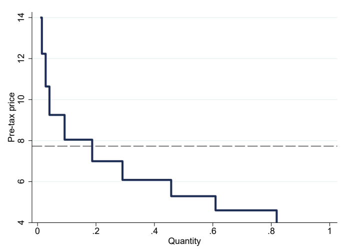
(b) Glad Trash Bags



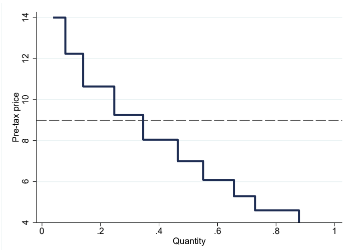
(c) Rubbermaid Lunch Blox



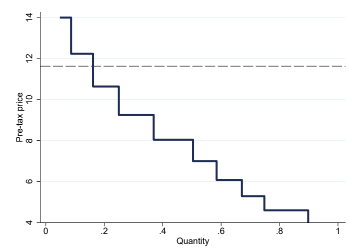
(d) Scotch-Brite Sponge



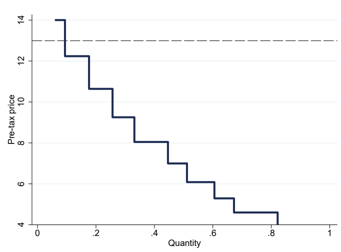
(e) Microban Cutting Board



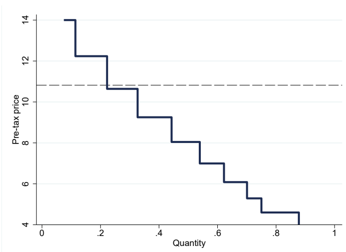
(f) Nordic-Ware Baker's Half Sheet



(g) Libbey White Wine Glasses



(h) Envision Home Microfiber Bath Mat



(i) Carnation Home Fashions Shower Curtain Liner

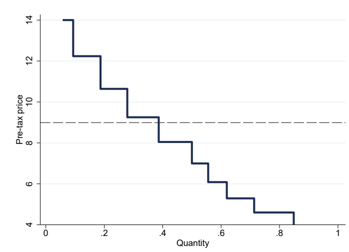


Figure B.17.1 presents product-specific demand curves from the no-tax store. The Amazon.com price is indicated by the dashed line. Prices are from February 2015, as documented in Taubinsky and Rees-Jones (2018). They may vary over time or by geographic region.

B.18 Welfare implications of overreaction

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

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

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

where $\varepsilon_{D,p}$ is the price elasticity of demand, $E[\theta|p,t]$ and $Var[\theta|p,t]$ correspond to the mean and variance of consumers marginal at price p and tax t,[13] and ρ is the pass-through rate of producer taxes (e.g., excise taxes) to prices. The variance of misreaction is a sufficient statistic for the efficiency costs of heterogeneous tax misreaction. The formula generalizes the homogeneous underreaction case analyzed by Chetty et al. (2009), in which making taxes opaque always reduces deadweight loss under the assumption of quasilinear utility.[14] Note that we assume that the tax is sufficiently small that it does not affect the pass-through rate. If pass-through rates are decreasing in θ this can attenuate (amplify) DWL due to overreaction (underreaction).

With heterogeneity, making taxes opaque increases deadweight loss if and only if $\rho E[\theta|p,t]^2 + Var[\theta|p,t] > \rho$. In the leading case of constant marginal costs of production (and thus full pass-through), this reduces to $E[\theta|p,t]^2 + Var[\theta|p,t] > 1$. But as shown in equation (2.8) of Proposition 1, $E[\theta|p,t]^2 + Var[\theta|p,t] \leq E[\theta|p,t]$, and thus deadweight loss is guaranteed to be smaller with shrouded taxes if $\theta < 1$ for all consumers. For example, combining our experimental estimate of $E[\theta] = 0.48$ with the presumption that all consumers underreact would imply that deadweight loss is *at least* 50% smaller when sales taxes are shrouded. Instead, we find significant overreaction, corresponding to $Var[\theta] \geq 0.83$. This implies that $E[\theta|p,t]^2 + Var[\theta|p,t] \geq 1.06$, and thus that shrouding taxes increases deadweight loss. Moreover, because this calculation uses the lower-bound variance estimate, considers full pass-through, and ignores the mental effort costs used to process the opaque taxes, the actual deadweight loss may be significantly higher if (i) the variance is significantly larger, (ii) mental costs of effort are taken into account and (iii) ρ is significantly lower than 1.[15]

B.19 Additional details of the experiment

The experiment proceeded in the following order:

[13] Concretely, this is mathematically equivalent to the formula in Proposition 2 of TRJ. See also Farhi and Gabaix (2020) for similar insights with applications to Ramsey taxation.

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

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

- Consent form: participants were first shown a consent form, which is shown in Figure B.19.1.
- Questions on residence: participants answered the questions in Section B.19.2 about their residence. The city and state selected in Question 1 were entered in future questions.
- Instructions: participants were shown three screens containing the experiment instructions. Figures B.19.2-B.19.4 contain the screenshots.
- Pre-purchase comprehension questions: participants answered the questions in Section B.19.2. Participants must answer all questions correctly to proceed. They were informed if they answered a question incorrectly, and were given unlimited attempts to get the correct answers.
- Purchase decisions: participants made nine purchase decisions. Each participant was randomly assigned three products from the list in Section B.19.3, and shopped for each product in all three stores. For each purchase, participants were first shown a screen detailing which store they entered, as seen in Figure B.19.5a. They then filled out an MPL, as seen in Figure B.19.5b.
- Post-purchase comprehension questions: participants answered the questions in Section B.19.2.
- Additional closing questions: participants answered the questions in Section B.19.2.

Section B.19.1 contains screenshots of the instructions shown to participants. Section B.19.2 contains text of the questions asked of participants, and, where applicable, the correct answer displayed in parenthesis and italics. Section B.19.3 contains a list of the products used, along with the Amazon.com prices and product descriptions.

B.19.1 Instructions

Figure B.19.1: 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 B.19.2: Instructions (screen 1 of 3)

Instructions

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

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

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

RainStoppers 68-Inch Oversize Golf Umbrella

This RainStoppers 68" oversize golf umbrella is large enough to cover three or more people. Umbrella frame constructed with fiberglass shaft and ribs for maximum stability. Canopy is made of 190T Nylon fabric. Complete with a foam non slip handle. Matching sleeve included. Length when closed is 43.

	Yes	No
Would you buy the RainStoppers umbrella for $5?		
Would you buy the RainStoppers umbrella for $4?		
Would you buy the RainStoppers umbrella for $7?		
Would you buy the RainStoppers umbrella for $13?		
Would you buy the RainStoppers umbrella for $10?		
Would you buy the RainStoppers umbrella for $6?		

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

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

Figure B.19.3: Instructions (screen 2 of 3)

Your shopping decisions

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

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

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

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

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

Figure B.19.4: Instructions (screen 3 of 3)

How your decisions in this study impact **you**

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

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

The computer will randomly select one of the stores.

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

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

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

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

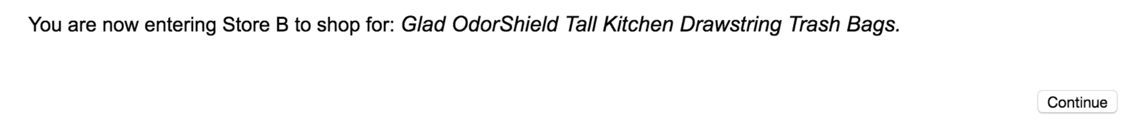
Every store, item, and price has an equal probability of being selected.

The bottom line:

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

Figure B.19.5: Screenshots of a purchase decision

(a) Introduction



(b) Purchase decision

Figure B.19.5 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.

B.19.2 Text of questions

Residence

Question 1: Please select the state, county, and city in which you currently live:
Question 2: How long have you live in [city], [state]

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

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

Pre-purchase comprehension questions

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

- $10
- $16
- $30

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

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

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

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

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

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

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

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

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

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

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

Closing comprehension questions

The questions below are about a hypothetical survey respondent Alex, who lives in [city], [state] just like you. **You must answer these three questions correctly to be eligible for the \$16 shopping budget**.
Question 1: If Alex purchases an item for \$10 in <u>Store A</u>, then... *(Correct answer: option 1)*

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

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

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

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

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

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

Additional closing questions

Question 1: What percent is the sales tax rate in your city of residence, [city], [state]? If your city exempts some goods from the full sales tax, please indicate the rate for a standard non-exempt good. If you're not sure, please make your best guess. (Note: Please enter your answer as a percent. For example, if you think that the tax rate is 1 percent, please enter 1, rather than 0.01. Do not include a percent sign in your answer.)
Question 2: Suppose that you bought a (standard, not tax-exempt) product for $8. What would be the [city], [state] sales tax that you would have to pay for that product? (Note: Please enter your answer in dollar units. For example, if you think that the sales tax is 10 cents, please enter 0.10 and not 10. Do not include a dollar sign in your answer.)
Question 3: Suppose you had $100 in a savings account and the interest rate was 2 percent per year. After 5 years, how much do you think you would have in the account if you left the money to grow? *(Correct answer: option 1)*

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

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

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

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

- True
- False
- Do not know

Question 6: How many people are in your household (including yourself)?
Question 7: What is your marital status?

- Single
- Married or domestic partnership
- Widowed

Question 8: What was your total household income for the year 2015?
Question 9: What is the highest level of education that you have attained?

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

Question 10: Are you currently a student?

- Yes
- No

Question 11: What is your age?
Question 12: Which best describes your political party affiliation?

- Independent
- Republican
- Democrat
- Other

B.19.3 Items used in the study

Product	Amazon.com price	Amazon.com product description
Energizer AA Batteries max Alkaline 20-Pack	\$11.15	Energizer AA max alkaline batteries 20 pack super fresh, Expiration Date: 2024 or better. Packed in original Energizer small box 4 batteries per box x 5 boxes total 20 batteries.
Glad OdorShield Tall Kitchen Drawstring Trash Bags, Fresh Clean, 13 Gallon, 80 Count	\$12.79	Glad OdorShield Tall Kitchen Drawstring Trash Bags backed by the power of Febreze are tough, reliable trash bags that neutralize strong and offensive odors for lasting freshness. These durable bags are great for use in the kitchen, home office, garage, and laundry room.
Rubbermaid Lunch Blox medium durable bag - Black Etch	\$10.47	The Rubbermaid 1813501 Lunch Blox medium durable bag - Black Etch is an insulated lunch bag designed to work with the Rubbermaid Lunch Blox food storage container system. The bag is insulated to achieve the maximum benefit of Blue Ice blocks and keep your food cold. The bag features a bottle holder, side pocket, comfort-grip handle and removable shoulder strap. The lunch Blox bag is durable and looks good for both the professional bringing their lunch to work or the kid taking their lunch to school.
Scotch-Brite Heavy Duty Scrub Sponge 426, 6-Count	\$7.73	O-Cel-O™ sponges and Scotch Brite scrubbers are truly a fashion-meets-function success story. The highly absorbent and durable sponges come in different sizes and scrub levels for the various surfaces around the home. Their assorted colors and patterns follow the current fashion trends to create the perfect accent in any room.
Microban Antimicrobial Cutting Board Lime Green - 11.5x8 inch	\$8.99	The Microban cutting board from Uniware is the perfect cutting board for the health conscious. The cutting board has a soft grip with handle and is dishwasher safe. The cutting board can be reversible, used on both sides, and is non-porous, non-absorbent. The rubber grips prevents slipping on countertop. Doesn't dull knives, juice-collecting groove. Microban is the most trusted antimicrobial product protection in the world. Built-In defense that inhibits the growth of stain and odor causing bacteria, mold, and mildew. Always works to keep the cutting board cleaner between cleanings. Lasts throughout the lifetime of the cutting board. Size: 11.5"x8" Color: Lime Green.

Product	Amazon.com price	Amazon.com product description
Nordic Ware Natural Aluminum Commercial Baker's Half Sheet	\$11.63	Nordic Ware's line of Natural Commercial Bakeware is designed for commercial use, and exceeds expectations in the home. The durable, natural aluminum construction bakes evenly and browns uniformly, while the light color prevents over-browning. The oversized edge also makes getting these pans in and out of the oven a cinch. Proudly made in the USA by Nordic Ware.
Libbey 14-Ounce Classic White Wine Glass, Clear, 4-Piece	\$12.99	Great for any party, this set includes four 14-ounce clear classic white wine glasses which match perfectly with the classic collection by libbey. The glasses are dishwasher safe and made in the USA.
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.

Product	Amazon.com price	Amazon.com product description
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.

Appendix C

The Effects of Price Discounts on Consumer Behavior and Beliefs: Evidence from a Field Experiment in the Apparel Industry

C.1 Survey Appendix

Due to the company's desire to remain anonymous, we remove their name and the names of their products and competitors from the instructions below. The survey proceeded in the following order. First, all participants saw the consent and introduction sections. Based on their choices, they would then answer a series of identical questions on two men's products or two women's products. They would complete all questions for one product before seeing questions for the other product. Finally, they answered a series of closing questions on the company and brand.

For the WTP questions, we utlizied a dynamic MPL, where customers could click on one spot in the MPL and it would automatically fill out other rows based on that response. For example, if participants selected that they preferred to receive the product over $50, the rows above would automatically fill out to say they preferred to receive the product at cheaper prices, and vice versa for rows below it. Participants could then adjust accordingly.

For this appendix, text in italics was not displayed to participants, but rather serves to clarify the questions asked.

C.1.1 Consent

Thank you for your interest! This is a study by researchers at The University of California at Berkeley, and Princeton University. We are interested in what factors people consider when making shopping decisions.
Procedure: If you agree to be in this study, which is completely voluntary, you will be asked to do the following: Answer a few basic questions about your demographics. Answer a few questions about your shopping preferences.
Study time: The study will take 5-10 minutes.
Compensation: In return for your time and effort, you will be entered into a lottery to win one of six prizes, which include one $250 Amazon gift card, four $100 Amazon gift cards, and a [Company] product. The [Company] product chosen will be based on your choices in the survey. You will not be paid any money if you do not complete the study. Confidentiality: [Company] will not have access to your individual survey responses. All responses will be seen only by the research team at The University of California at Berkeley and Princeton University. To minimize the risks to confidentiality, all data will be transmitted via a secure, encrypted connection and stored on an account to which only the research team has access. All identifiable data will be destroyed after the completion of this study. Identifiers will be removed from the identifiable private information. [Company] will not have access to the identifiers. After such removal, the information could be used for future research studies or distributed to other investigators for future research studies without additional informed consent from the subject.
Contact: If you have any questions, please feel free to contact Daniel Morrison at dm31@princeton.edu, or Princeton Institutional Review Board at irb@princeton.edu. This study received Princeton IRB approval (IRB #13993).
If you agree to take part in the research, please click on the "Yes" button below.

- Yes, I wish to participate in this study.

- No, I decline the opportunity to participate in this study.

C.1.2 Introduction

Thank you for participating in our survey! You might be randomly selected to receive a [Company] product based on your responses to the following question. If selected, would you prefer to receive men's apparel or women's apparel?

- Men's apparel
- Women's apparel

C.1.3 Questions on products

Introduction

Thank you for your responses so far! We will first ask you some questions about the [Company product]. *Company product description, taken from their website, is then shown, along with photos and available color options. A competior product, taken from their website, is then shown, along with photos and available color options.*
You may win a prize! In the following questions you will be shown some tables, and asked whether you prefer the option on the left or the option on the right for each row.
At the end of the survey, we will randomly select one participant to receive a product, based on the choices in the table. We have set up the procedure so that **it is in your best interest to answer honestly.**
If you are randomly selected to receive a product, the computer will first randomly select one of the questions about which you made choices. The computer will then randomly choose one row from that table. You will then receive the option you selected. We will then email you to ask for your preferred size, color, and shipping address.

MPL questions

First MPL: In each row of the table below, please tell us whether you would prefer to receive a [Company product], or the specified amount of money.
There's a chance we will honor one of the decisions in this table, so it's in your best interest to answer honestly.
You can fill out the table by selecting your decision in every row, or by clicking on the row where you want to switch from choosing to receive the Stio product to choosing to receive money. *Note: the monetary values ranged from $10 to $210, by increments of $20.*

Second MPL: In each row of the table below, please tell us whether you would prefer to receive a [Competitor product], or the specified amount of money.
There's a chance we will honor one of the decisions in this table, so it's in your best interest to answer honestly.
You can fill out the table by selecting your decision in every row, or by clicking on the row where you want to switch from choosing to receive the Stio product to choosing to receive money. *Note: the monetary values ranged from $10 to $210, by increments of $20.*

Third MPL: Now imagine you have the following options:

- Receive a [Company product] that will be shipped to you on February 15.
- Receive a [Company product] that will be shipped to you on June 15 along with some money (the money would also be sent on June 15).

In each row of the table below, please tell us which option you prefer. There's a chance we will honor one of the decisions in this table, so it's in your best interest to answer honestly. You can learn more about how your decisions affect your prize here. You can fill out the table by selecting your decision in every row, or by clicking on the row where you want to switch from choosing to have the Stio product ship on February 15th to June 15th.
You can fill out the table by selecting your decision in every row, or by clicking on the row where you want to switch from choosing to receive the Stio product to choosing to receive money. *Note: the monetary values ranged from $5 to $55, by increments of $5.*

Beliefs questions

[Company] sometimes offer sales and send promotional discount codes that customers can enter at checkout to take a certain percentage off an item. For how many of the 100 days following February 15th (that is February 15 to May 26) do you think you would be able to get the following price reduction on the [Company product], shown below. This product has a listed price of [list price]. Note: the entries must add to 100 days. The total is displayed at the bottom.

- 0% off
- 1-10% off
- 11-20% off
- 21-30% off
- 31-40% off
- 41-50% off
- 51% off or more

The [Company product] is shown below. This product has a listed price of [list price]. Imagine you are planning to buy this product (in your preferred size and color) sometime between February 15 to May 26 (that is, sometime in the 100 days following February 15). What do you think the lowest price you will be able to purchase this product during that period is? *Note: the prices below were shown both in dollar terms and in %off terms, E.g., $100-$120 (21-30% off).*

- No discount
- 1-10% off

- 11-20% off
- 21-30% off
- 31-40% off
- 41-50% off
- 51% off or more

How confident are you in this answer?

- Not at all confident
- Slightly confident
- Moderately confident
- Mostly confident
- Very confident

What is the lowest price at which you are **very confident** that you will be able to purchase this product between February 15 to May 26 (that is, sometime in the 100 days following February 15)? Note: If you are not very confident that there will be a discount during this period, select the "no discount" option.

- No discount
- 1-10% off
- 11-20% off
- 21-30% off
- 31-40% off
- 41-50% off
- 51% off or more

C.1.4 Closing questions

Participants answered the following questions about the company and three of its competitors.

- How likely are you to recommend each of the following brands to friends or family? *Answered on a slider from 1-10*
- To what extent do you think the following companies care about their customers? *Answered on a slider from 1-10*

- Think about your active-wear clothing and apparel purchases over the past year. Roughly how much of the total expenses went toward each of the following companies? *Options included None at all, A little, A moderate amount, A lot, All*

- For each of the following companies, how often is there a sale on the products you want to buy? *Options included Never, Rarely, Sometimes, Often, Always, and Not Sure*

For the following questions, participants could select any or all of the options, which included the company and three of its competitors. A "none of the above" option was also included.

- From which of the following companies have you received discount codes in the last 30 days?

- From which of the following companies do you regularly receive emails?

What is your age?
Which race or ethnicity best describes you?

- American Indian or Alaska Native
- Asian or Pacific Islander
- Black or African American
- Hispanic or Latino
- White
- Other
- Prefer not to respond

What was your total household income for the past 12 months? Please include wages, salary, bonuses, tips, investment income (for example: interest, dividends, and rental income), government payments (for example: Social Security, welfare payments, and unemployment compensation), and pensions.

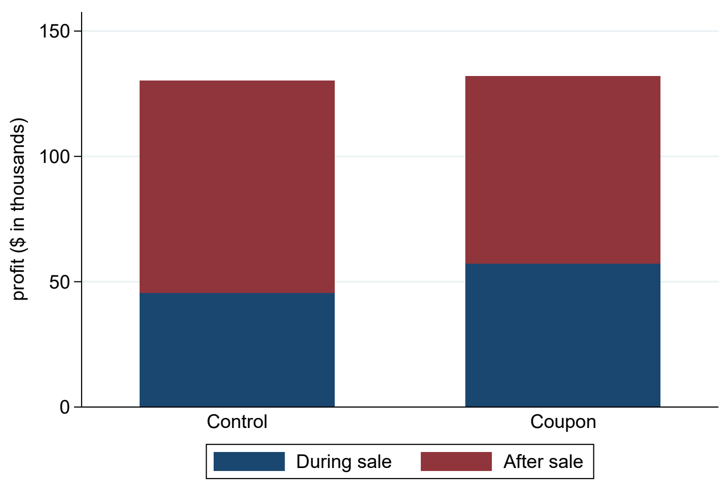
- $0 to $20,000
- $20,001 to $40,000
- $40,001 to $60,000
- $60,001 to $80,000
- $80,001 to $100,000
- $100,001 to $120,000
- $120,001 to $140,000
- $140,001 to $160,000

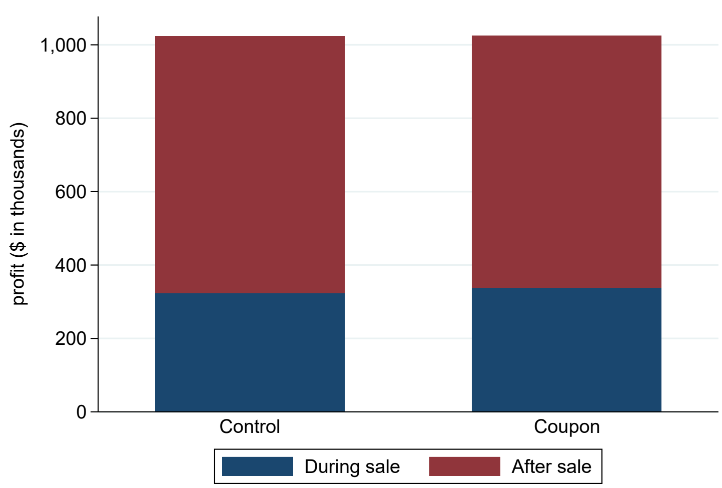
- $160,001 to $180,000
- $180,001 to $200,000
- $200,001 or more

C.2 Additional Figures

Figure C.2.1: Profit from the treatment and control groups during and after the sale

(a) Eligible Products



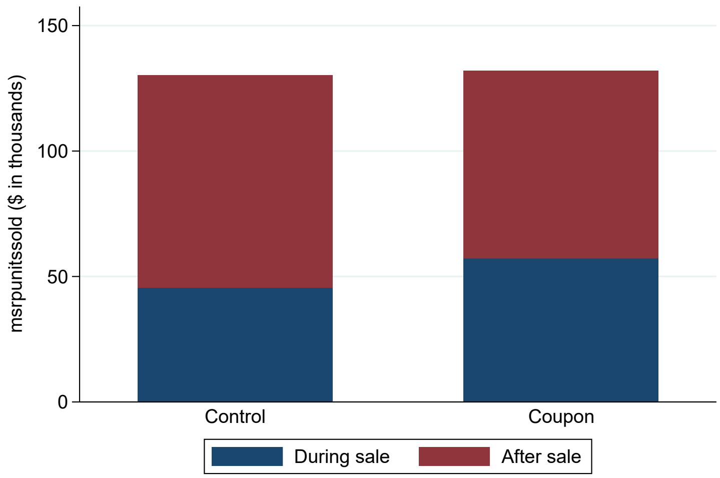


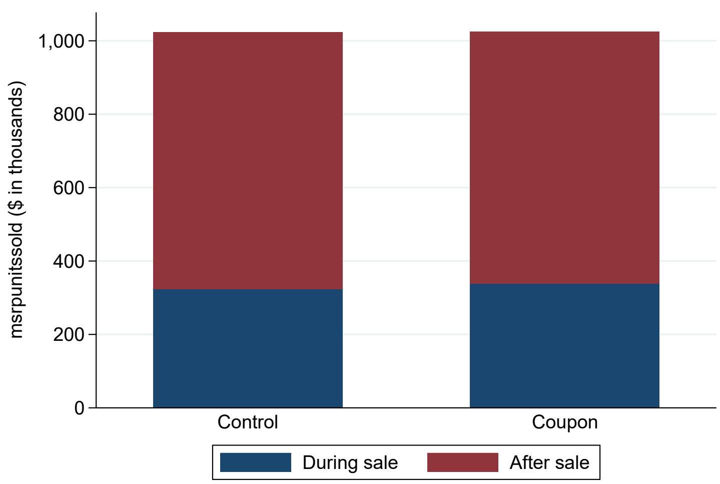
(b) Eligible Products

This figure shows the number of units sold to customers in the experiment before and after the sale. We double the quantities in the control group as there were twice as many participants assigned to the coupon group than the control group.

Figure C.2.2: Units sold to the treatment and control groups during and after the sale

(a) Eligible Products





(b) Eligible Products

This figure shows the number of units sold to customers in the experiment before and after the sale. We double the quantities in the control group as there were twice as many participants assigned to the coupon group than the control group.