NBER WORKING PAPER SERIES

MEASURING THE WELFARE EFFECTS OF SHAME AND PRIDE

Luigi Butera
Robert Metcalfe
William Morrison
Dmitry Taubinsky

Working Paper 25637
http://www.nber.org/papers/w25637

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2019, Revised August 2020

We are indebted to Ryan Johnson and Erik Daubert from the YMCA of the USA for their invaluable guidance and expertise in setting up our research partnership, as well as to Maria-Alicia Serrano for her continuing support. We wish to thank the YMCA of the Triangle Area, particularly Tony Campione, Mark Julian, Brian Spanner, and Janet Sprague, for their outstanding support in implementing our field experiments. We are grateful to Leonardo Bursztyn, Gary Charness, Alain Cohn, Jonathan Davis, Stefano DellaVigna, Ray Fisman, Jana Gallus, David Gill, Alex Imas, John List, Keith Ericson, Fatemeh Momeni, Ricardo Perez-Truglia, Daniel Tannenbaum, as well as numerous conference and seminar participants for helpful comments. The views expressed herein do not necessarily reflect the views of the YMCA of the USA or any other YMCA member association. The experiment was approved by University of Chicago IRB, #IRB15-1647. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Measuring the Welfare Effects of Shame and Pride
Luigi Butera, Robert Metcalfe, William Morrison, and Dmitry Taubinsky
NBER Working Paper No. 25637
March 2019, Revised August 2020
JEL No. D8,D9,H0,I0

ABSTRACT

Public recognition is a frequent tool for motivating desirable behavior, yet its welfare effects are rarely measured. We develop a portable money-metric approach for measuring the direct welfare effects of shame and pride, which we deploy in a series of experiments on exercise and charitable behavior. In all experiments, public recognition motivates desirable behavior but creates highly unequal emotional consequences. High-performing individuals enjoy significant utility gains from pride, while low-performing individuals incur significant utility losses from shame. We estimate structural models of social signaling, and we use the models to explore the social efficiency of public recognition policies.

Luigi Butera
Department of Economics
Copenhagen Business School
Denmark
lbutera2@gmail.com

Robert Metcalfe
Questrom School of Business
Boston University
595 Commonwealth Avenue
Boston, MA 02215
and NBER
rdmet@bu.edu

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

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

Randomized controlled trials registry entries are available at
https://www.socialscienceregistry.org/trials/4004
and
https://www.socialscienceregistry.org/trials/5737

"What do you regard as most humane? To spare someone shame."
– Friedrich Nietzsche, the Joyful Wisdom

"A soldier will fight long and hard for a bit of colored ribbon."
– Napoleon Bonaparte[1]

The human desire to avoid shame and seek out pride 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]

The *financial* costs of utilizing public recognition to motivate behavior are typically low, but the *emotional* costs may not be. Although behavioral scientists sometimes refer to social-influence-based interventions as light-touch, innocuous "nudges" (Halpern, 2015; Benartzi et al., 2017), such a label would not be appropriate for a policy that leads to a significant number of individuals experiencing shame (Bernheim and Taubinsky, 2018). 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).[3] On the other hand, public recognition policies that mostly generate warm feelings of pride are arguably a "win-win."

Unfortunately, psychological theories do not provide clear guidance about when shame or pride will be the more prevalent consequence of public recognition (Leary, 2007; Tangney et al., 2007). Developing quantitative methods for measuring the welfare effects of public recognition is therefore crucial for both positive and normative progress.

[1] We thank an anonymous referee for this quote.

[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. (2009), Jones and Linardi (2014), Bernheim and Exley (2015), Exley (2018), and Birke (2020).

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

In this paper, we develop a portable approach for directly quantifying the emotional effects of public recognition, using a money-metric method that can be immediately incorporated into welfare analysis. 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 emotional effect on people's utility? 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 through pride, and are individuals choosing low levels of the desirable behavior made worse of through shame? Third, are the emotional effects of shame and pride on net 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 shameful behavior transitions to admirable behavior.

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. (2009) 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 to address some open issues from the field experiment. In this real-effort task, participants raise money for the American Red Cross by repeatedly pressing two keys on a computer keyboard. Participants in our study took part in three rounds. In the Anonymous 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 this online experiment simultaneously to three different subject pools that differ in individuals' familiarity with each other: (i) an online panel called 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

[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.

[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.

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 misforecasting 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.[6]

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. As in the YMCA experiment, we elicited WTP for different possible levels of charitable contribution, and participants' elicited preferences were implemented with 10 percent chance. With 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 pay to avoid shame or obtain pride 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' payoffs from public recognition vary with their level of performance. We find that participants' payoffs are strictly increasing in performance in all experiments. Moreover, in all experiments, participants in the bottom quartile of performance receive negative payoffs, on average, while participants in the top quartile of performance receive positive payoffs, on average. The robust presence of negative payoffs from public

[6]This information was common knowledge among participants.

recognition is consistent with leading economics models of social signaling (e.g., Bénabou and Tirole, 2006; Andreoni and Bernheim, 2009), but it is not an implication of psychological theories of shame. 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. For example, raising *any* amount of money for the Red Cross could have been perceived as commendable prosocial behavior.

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 we compare the models' predictions to a direct estimate of the effect of the financial incentive. In the YMCA experiment, we were not able to randomize a financial incentive, but 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 this overestimation is not statistically significant at conventional levels. This suggests that our monetization of social image incentives is accurately capturing the (presumably nuanced) psychological effects of public recognition.

Fifth, we study the shape of the PRU. In our models, whether the emotional effects of pride and shame are on net negative or positive depends on the degree of concavity and the standard for what constitutes pride-worthy versus shameful behavior. Intuitively, more concavity leads individuals to be more sensitive to shame than to pride, while a higher standard increases the fraction of individuals who experience shame. For example, if people derive pride 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 at which shameful behavior transitions to pride-worthy behavior 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 BU samples, is 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 likely to be a less socially efficient means of generating behavior change than are financial incentives. We precisely quantify this deadweight loss, and we numerically examine how it varies with the shape of the PRU. If the shape of the PRU more closely resembled our estimate in the Prolific sample, then public recognition would be a more efficient means of changing behavior than financial incentives.

Collectively, our results illustrate the importance of directly measuring the welfare effects of shame and pride, and the potential benefits of our methodology. Our findings about the prevalence of shame 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 may 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. While there are a number of reasons why caution is warranted in extrapolating from our specific results, one lesson seems clear: it is inappropriate to judge the success of a public recognition policy solely by its effect on behavior. Our methodology could help enrich the applied work on public recognition and social signaling by helping researchers study both behavior and *welfare.* We discuss possible extensions and additional applications in the concluding section.

The remainder of the paper is organized as follows. Section 1 further reviews the related literature. Section 2 introduces our theoretical framework. Section 3 describes the YMCA experiment and Section 4 reports the reduced-form results. Section 5 describes the charitable contribution experiments and Section 6 reports the reduced-form results. Section 7 presents our estimates of structural models of social signaling and welfare implications. Section 8 concludes by discussing limitations, robustness, and questions for future research.

1 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 the effects of experiencing shame and pride on people's wellbeing. We build on this literature by developing a portable approach for measuring the welfare effects of shame and pride, 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). 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. We also add several technical innovations to this important literature. First, our experiments utilize a new design technique, grounded in "strategy method"

approaches typically only used in laboratory experiments, that eliminates the need to rely on the assumption that individuals can correctly forecast their future behavior.[7] We establish the need to relax this assumption in our setting, and we discuss its relevance for other studies. Second, we develop simple principles for comparing the economic efficiency of non-financial policy instruments to that of financial incentives.

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).[8] 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.[9] 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.

2 Theoretical framework for analysis

2.1 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

[7] See also Bernheim and Taubinsky (2018) for a more detailed discussion of the weaknesses assuming that people can forecast their behavior when evaluating non financial policy instruments, as well as the "non-comparability problem" that less theory-grounded approaches such as those of Allcott and Kessler (2019) are subject to.

[8] Karing (2019), Bursztyn et al. (2019), Ariely et al. (2009), and Exley (2018) 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.

[9] We delineate between social pressure and public recognition. Social pressure commonly refers to situations in which individuals take actions to avoid the emotional agitation of another person's pressure in typically face-to-face interaction; an example is DellaVigna et al. (2012), where social pressure is a force layered on top of the information already revealed by choosing to use a do-not-visit tag. Public recognition instead refers to situations in which individuals take actions to influence others' beliefs about them. In some settings both are in play; e.g., when telling a surveyor whether or not one has voted, as in DellaVigna et al. (2017). The implicit assumption of the DellaVigna et al. (2017) model that not answering the door to a pre-announced visit (an action most likely to be taken by those who did not vote) generates no disutility beyond hassle costs is more consistent with a social pressure interpretation and less consistent with leading social signaling models such as those of Bénabou and Tirole (2006) and Andreoni and Bernheim (2009).

[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.

$a^*(\theta)$, 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 shame and pride 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 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):

$$u(a;\theta) + y + \nu S\left(a - \rho\bar{a}\right) \tag{1}$$

where $\bar{a}$ is the average action in the population, and $\rho\bar{a}$ is the standard for what constitutes shameful versus pride-worthy behavior. 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):

$$u(a;\theta) + y + \nu S(\mathbb{E}[\theta|a] - \rho\bar{\theta}) \tag{2}$$

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 shameful versus pride-worthy characteristics.[11]

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

The generalizations in Appendix A imply that in the general case where behavior and/or types are not fully revealed, the standard will depend on the information structure. 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 neither shame nor pride. Roughly speaking the parameter ρ tends to 1 as the information structure coarsens.

[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).

2.2 The net effects of shame and pride

Although theoretical work often makes the simplifying assumption that the net effect of shame and pride is zero by assuming that S is linear and that $\rho = 1$ (e.g., Bénabou and Tirole, 2006, 2011), it is well understood that both assumptions are not without loss of generality. Psychologically, both assumptions can be legitimately challenged. Because shame and pride are separate emotions of different valences (Tangney et al., 2007), people's wellbeing 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.

Plainly, both the curvature of S and the value of ρ determine the net utility effect of shame and pride. In particular, let $a(\theta)$ denote individuals' equilibrium action choices. Then the net effects of shame and pride in the two models are, respectively, given by:

$$\mathbb{E}[S(a(\theta) - \rho\bar{a})] \tag{3}$$

$$\mathbb{E}[S(\mathbb{E}[\theta|a(\theta)] - \rho\bar{\theta})] \tag{4}$$

If S is concave and $\rho \geq 1$, then Jensen's Inequality implies that the net effects of shame and pride 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, the net emotional effect is negative when the function is concave and the standard for behavior/characteristics is at least as demanding as the average. Conversely, the net emotional effect is positive when $\rho \leq 1$ and S is convex. In general, the net emotional effect 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 emotional effect 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 emotion 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 emotional effect $\mathbb{E}[S]$ connects to a key economic question: whether public recognition is an efficient tool for behavior change relative to standard financial

incentives. We show that under some homogeneity assumptions, $\mathbb{E}[S]$ is the welfare effect of public recognition relative to financial incentives. Intuitively, consider revenue-neutral financial incentive schemes that penalize individuals for low levels of a and reward individuals for high levels of a. In the appendix, we show that under some assumptions there exists a financial incentive scheme that produces exactly the same distribution of behavior change as does public recognition. However, by virtue of being revenue-neutral, this financial incentive scheme achieves this behavior change at a net-zero cost or benefit to individuals. In fact, when $\mathbb{E}[S]$ is positive (negative), the revenue-neutral scheme pareto dominates (is pareto dominated by) public recognition.

2.3 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 the equilibrium response of an individual in an experiment may differ from the equilibrium response when public recognition is scaled up to the broader population.

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 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.

Moreover, the net emotional effect could be positive in the experiment even if $\rho = 1$ and S is concave, illustrating the sense in which partial-equilibrium reduced-form results need to be supplemented with structural modeling. Our reading of existing literature studying social comparisons and social pressure is that it often stops at R_{exp}.[12]

[12]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

3 YMCA experiment design

3.1 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 were told that they could sign up for the program by completing the survey and agreeing to participate.[13]

3.2 Experimental protocol

The survey began by explaining the nature of the incentives during the program.[14] 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 provides a screenshot of what this public recognition email entailed.

We then elicited people's willingness to pay for receiving (or avoiding) public recognition using

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.

[13]The "pilot" language was important for our field partner, but we recognize that in principle it could have affect people's perceptions about the longer-run consequences of their choices. However, recent work by DellaVigna and Pope (2019) and de Quidt et al. (2018) suggests that framing effects of this sort seem have muted effects on behavior. DellaVigna and Pope (2019) also put forward the provocative finding that academics seem to overestimate the extent to which such framing matters (at least with respect to the specific but related issue of whether or not subjects are told that they are part of an experiment).

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

a combination of the *strategy method* and the Becker-DeGroot-Marschak (BDM) mechanism. The incentive-compatible method contained eleven two-part questions about possible numbers of visits: 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 public 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?"* The details were then explained in simple and plain language, and participants 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 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.[15]

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

[15]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.

$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.[16]

Because others' behavior plays a role in the models summarized in Section 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 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.[17]

3.3 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.[18] We utilize attendance data for

[16] 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 $b = -v$.

[17] 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.

[18] 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. To account for the risk of participants strategically swiping in and out without accessing YMCA programs and initiatives during their stay, 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. Participants knew that multiple accesses during the same day would only count as one attendance.

non-experimental participants in the out-of-sample predictions in Section 7.

3.4 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.

Preference 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 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 emotions Although our approach does not require people to correctly forecast their future attendance, it does rely on the assumption that people can anticipate the emotional effects 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).

4 Reduced-form results from the YMCA experiment

4.1 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.[19] 43 participants were randomly assigned to receive

[19]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 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. 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 addition to the coherent sample, we also analyze the slightly smaller group of participants whose preferences for public recognition are monotonically increasing in YMCA visits. This *monotonic sample* is of particular interest because it is consistent with the typical monotonicity assumptions of the models in Section 2.

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

4.2 The effect of public recognition on behavior

Figure 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 2. Columns (1)-(3) present results from the monotonic sample, while columns (4)-(6) present results from the slightly larger coherent sample. The table shows that in both samples 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.

4.3 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 4 plots the average WTP by the attendance level that would be publicized to other

the YMCA of the Triangle via email about their treatment assignment the morning of the first day of Grow & Thrive.

participants. These WTP profiles constitute model-free measures of the reduced-form PRU R_{exp} introduced in Section 2.3. We identify each set of possible visits from our elicitation with its midpoint, meaning that the first five sets $\{0\}, \{1\}, \ldots, \{4\}$ 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 from the monotonic sample, 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 all three panels, 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 shameful versus pride-worthy behavior. As discussed in Section 2, the net effect of shame and pride 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 4 also shows that participants' PRUs do not vary with their past attendance. We verify this formally in regression analysis in Table A1 in Appendix C.1. 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 shame and pride. 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 4 would diverge.

Table 3 quantifies the descriptive results in Figure 4 by presenting regressions of WTP for public recognition on the midpoint of the visits intervals. Columns (1)-(4) present results from the monotonic sample, while columns (5)-(8) present results from the coherent sample. 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 3 generate two robust results, which are visually apparent in Figure 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 $-R''_{exp}/R'_{exp}(0)$, which is analogous to the coefficient of absolute risk aversion (ARA). Another measure of curvature is $-R''_{exp}/R'_{exp}(0)$ multiplied by the standard deviation of attendance of YOTA participants.[20] 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 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 5 and Table 4.

Figure 5 presents the WTP data analogously to Figure 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 4, Figure 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 5 is significantly steeper. While the profile in Figure 4 spans payoffs between approximately -$2 and $2, the profile in Figure 5 spans payoffs between approximately -$4 and $4. This difference is consistent with the possibility that the data reported in Figure 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 4 quantifies the results suggested by Figure 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 3, the estimated coefficients in Table 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.

[20] Note that we don't specify an argument for R''_{exp} because our quadratic regression models assume a constant second derivative.

However, our estimates of curvature are very similar to the estimates in Table 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.1 shows that the results in Table 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 7 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.

4.4 Further robustness checks

Potential bias from high visits questions Because only 10 percent of participants expected to attend the YMCA as many as 23 times, the 23-30 visits interval presents an unrealistic hypothetical to many participants, and thus might have undue influence on our estimate of concavity in Table 3. However, as Tables A3 and A4 in Appendix C.2 show, excluding these high visits intervals slightly increases our estimate of curvature. This is consistent with the visual evidence in Figure 5, which shows that the quadratic fit is equally consistent with high visits intervals and low visits intervals.

Potential bias from visits intervals increasing in size One key design decision was to make the intervals of possible visits very fine at low values (e.g., 0 visits, 1 visit, 2 visits), but more coarse at higher levels (e.g., 18 to 22 visits). Our motivation was to roughly equalize the number of participants whose attendance falls within each bin, as well as to avoid overburdening participants with too many decisions. Indeed, as shown in Figure A1 in Appendix C.3, our visits intervals are roughly equal in size according to this metric. Nonetheless, 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 if this might have led us to overestimate concavity, in Appendix C.3 we rescale the attendance intervals such that they are coded as equal in size. Specifically, 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 even with respect to this recoding of the intervals. Moreover, our estimate of curvature, $-R''/R'(0)$ is, if anything, slightly higher with respect to this recoding.[21] 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.

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

Demand for public recognition as commitment To the extent that individuals attend the YMCA to exercise rather than to participate in some other more immediately pleasurable activity, and to the extent that they are (partially) sophisticated about possible self-control problems, they may wish to motivate their future selves to attend the YMCA more. We argue that this is unlikely for several 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. However, an individual can decrease an expected payoff for a low attendance level either by inflating or deflating their WTP for the public recognition treatment at that attendance level. Thus, the bias, if it exists, is unsigned. 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, Augenblick and Rabin (2019), Fedyk (2018), and Yaouanq and Schwardmann (2019) all provide strong evidence against this. These three papers show that even when individuals are (partially) aware of their self-control problems, they are not sophisticated enough to use complex mechanisms to create commitment opportunities.

Second, as shown by Laibson (2015), Carrera et al. (2019), 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.4, we 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 use three additional survey elicitations for this analysis: (i) people's beliefs about their next month's attendance, (ii) their beliefs about the increase in attendance from a hypothetical \$1 incentive, and (iii) their valuation of the hypothetical \$1 incentive. As reviewed in Appendix C.4, Carrera et al. (2019) and Allcott et al. (2020) formally show that if people perceive themselves to be time-inconsistent, then their WTP for the \$1 per attendance incentive should equal the average of their expected attendance with and without the incentive. WTP values above this statistic imply that a person thinks that they don't attend the YMCA enough, while WTP values below this statistic imply that a person thinks that they attend the YMCA too much. We find no relationship between this measure and people's WTP for public recognition.

4.5 Realized payoffs from shame and pride

We end our reduced-form analysis by reporting our experimental participants' realized payoffs from the shame and pride 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 4.3, we estimate payoffs for each level of attendance using the specification in column (4) of the two

panels in Table 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 payoff from pride or shame, 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 C.5.

Figure 6 presents the results, both for the monotonic and the coherent sample. On average, participants who were publicly recognized received a net-zero payoff from their experience of pride and shame. 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 an upper bound on the net emotional effect 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 emotional payoff. We return to this in Section 7, where we estimate structural models and evaluate the impact of scaling up public recognition to all of YOTA.

4.6 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 Figure 7, 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., 2019).

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).

5 Charitable contribution experiments design

5.1 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.[22] 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.[23] Participants from all subject pools were informed they could only complete the experiment on a laptop or personal computer with a working webcam.

5.2 Experimental protocol

Except where noted below, the experimental protocol was identical for each of the three samples.[24] 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. Thus while the incentives are significantly lower in the Prolific sample in absolute terms, they are approximately the same in relative terms.

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. Figure 9 contains a screenshot of the visual provided to participants. 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

[22] 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.

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

[24] 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.

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 bonus 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.[25] Figure 8 contains a screenshot of the example given to participants. 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.

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."

Importantly, 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 "rational" 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 op-

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

tion.[26] 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 and independent of, for example, the order in which participants completed the experiment. The participant-level randomization implies that there cannot be within-group correlation in performance and preferences, and thus 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

[26]The ex-ante probability of a publicized contribution being based on the public recognition round was greater than 0.8. We explained to participants that most publicly recognized scores would be based on the Publicly-Shared Effort round, and that that is what they and others should infer.

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.

5.3 Discussion of the design

Within-person variation We chose to have participants complete all three possible rounds for two reasons. First, and most importantly, 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 shame.

Second, our design maximizes statistical power for comparisons of performance across the three rounds, and allows for some additional analyses of individual differences. We show in the next section that within- and between-participant estimates of performance differences between the three rounds are very similar, and thus that there is no evidence that the within-subject nature of the design biases our estimates.

Relation to the YMCA experiment The charitable contribution experiments complement the YMCA experiments in six 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.[27]

Fourth, the charitable contribution experimental protocol is relatively easy to implement and extend in a number of different directions. This allows for a number of interesting extensions of

[27] 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.

our design, which can accelerate the testing and refinement of social signaling models. We discuss these in Section 8.

Fifth, 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, 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.

Sixth, the charitable contribution experimental design has a number of 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 not just about the average performance but also about, e.g., the distribution at the very top or bottom; (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. This last feature implies that participants' WTP for public recognition cannot be affected by a demand for additional information.

6 Reduced-form results from the charitable contribution experiments

6.1 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. These exclusions yield 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 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 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 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.

6.2 The effects of public recognition on behavior

Figure 10 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 5 quantifies the effects depicted in Figure 10. 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 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 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 An important aspect of our design is that all participants completed all three rounds, which mitigated potential selective attrition and increased statistical power. Because all rounds were presented in random order, our design allows a between-subjects comparison of the money raised in the three rounds by simply limiting to the first round the participants encountered. We present this analysis in Table A8 in Appendix D. The table shows that 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. The confidence intervals around the between-subject estimates in the BU sample are wide, and include the within-subject estimates. Thus, on net we find no evidence that within-subject estimates differ from between-subject estimates.

6.3 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 11 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 shameful versus pride-worthy behavior. As discussed in Section 2, the net effect of shame and pride 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 11 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 A6 in Appendix D presents confidences intervals for the average WTP in each interval.

Table 6 quantifies the descriptive results in Figure 11 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 and all other regression analyses of the WTP data, we exclude the ≥ 1700 points interval, as it is ambiguous what number best represents that interval.

Consistent with Figure 11, 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. We find

significant concavity in the Prolific experiment, and find 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, $-R''/R'(0)$, also include the point estimate from the Prolific sample.

We can also compare our unitless measures of curvature, $-R''/R'(0)$ 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 3 and 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 3 and 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, Table 7 shows that there is no interaction between group size and WTP for public recognition in the Prolific sample. The table presents results from regressions in which WTP for public recognition is regressed on the publicized level of effort, with all covariates interacted with group size dummies. Column (1) includes only a linear term for publicized effort, whereas column (2) includes a quadratic term for publicized effort. 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 analysis 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. We investigate this possibility in the charitable contribution experiments in Table 8. This table presents results from regressions analogous to those in Table 6, except that we restrict to data points where the intervals for which WTP is elicited are within 500 points of participants' average performance in the three rounds. The average standard deviation of the difference in scores between any two rounds is just above 500 points, and thus the performance intervals studied in Table 8 are likely to be within the range of what participants consider plausible. Interestingly, the estimates in Table 8 are almost identical to those in Table 6. Thus, in contrast to the YMCA experiment, we find no evidence for attenuation in the charitable contribution experiments. This is perhaps due to the fact that participants faced greater uncertainty about their scores in this experiment, or 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

A9 in Appendix D 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.

Despite some evidence of heterogeneity, Table A10 in Appendix D shows that there is no relationship between the PRU and participants' intrinsic motivation. This result is consistent with the graphical evidence in Figure 11.

6.4 Realized payoffs from shame and pride

Finally, we estimate the net effect of shame and pride induced by public recognition. For each participant, we compute the utility that they would receive from having their Publicly-Shared Effort round score publicized. 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 12 presents the results. The net emotional effect of public recognition is positive in the Prolific sample, approximately 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).

6.5 Consistency with financial incentive effects

Before turning to our structural estimation, we provide simple 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 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. Consider, for example, the Prolific sample. Column (1) of Table 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 5. Analogous arguments imply that our Table 6 estimates imply that the financial incentive should increase performance by 216 and 150 points in the Berkeley and BU samples, respectively. Empirically, Table 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.

7 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. Second, we validate our experimental and structural methodology by more formally implementing the consistency tests from Section 6.5. Third, we study the welfare effects of scaling up the public recognition intervention to the full YOTA population.

We focus 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. In fact, our intervention was of potential interest to the YMCA administrators because it was regarded as a "cheap" means of increasing attendance, which spurred enthusiasm for our study. Our structural estimates allow us to study the full costs of scaling up this intervention. We focus on the YMCA setting because there is widespread interest in encouraging more exercise, and because interventions similar to ours were in fact considered by YOTA and spurred enthusiasm for running our study.

7.1 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 is quadratic. Letting $\bar{a}$ denote the average action, and $\bar{\theta}$ denote the average type, we assume that

$$\nu S^a(a - \rho\bar{a}) = \gamma_1^a(a - \rho\bar{a}) + \gamma_2^a(a - \rho\bar{a})^2 \tag{5}$$

$$\nu S^\theta(\mathbb{E}[\theta|a] - \rho\bar{\theta}) = \gamma_1^\theta(\mathbb{E}[\theta|a] - \rho\bar{\theta}) + \gamma_2^\theta(\mathbb{E}[\theta|a] - \rho\bar{\theta})^2 \tag{6}$$

for the action-signaling and characteristics-signaling models, respectively.[28] As shown in Appendix E, the resulting reduced-form PRU, $R(a)$, will be quadratic with both microfoundations.

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.[29] 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.[30]

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 4b for the YMCA sample, and in columns (2), (4), (6) of Table 6 for the Prolific, Berkeley, and BU samples.[31] As shown in Appendix E, estimates of the structural parameters γ_i^j and ρ from the structural PRUs in (5) and (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 $a = \theta/c$, and average performance in the absence of public recognition is

$$\mathbb{E}[a|PR = 0] = \mathbb{E}[\theta]/c. \tag{7}$$

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

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

[29] 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.

[30] 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.

[31] As discussed in the reduced-form results, the specification in column (4) of Table 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.

$$\begin{aligned}\mathbb{E}[a|PR=1] &= \mathbb{E}[\theta]/(c-2r_2) + r_1/(c-2r_2) \\ &= \mathbb{E}[a|PR=0]\cdot c/(c-2r_2) + r_1/(c-2r_2) \end{aligned} \tag{8}$$

Given an estimated average treatment effect $\bar{\tau}$ of public recognition on performance, the cost parameter c is identified by setting the difference between (8) and (7) equal to $\bar{\tau}$. We use the treatment effect estimates from column 5 of Table 2 for the YMCA sample, and estimates from columns (1)-(3) of Table 5 for the Prolific, Berkeley, and BU samples.

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 $a = (\theta + p)/c$, 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 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. (2019) 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.[32]

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 E.4 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.

[32]Moreover, our participants' forecasts about the effects of financial incentives are of similar magnitude as the estimates in Carrera et al. (2019): participants in their experiment forecasted that a $2 per attendance incentive would increase health club attendance by 2.3 visits over four weeks, and the incentive increased their attendance by 2.9 visits. The respective 95 percent confidence intervals are [2.1, 2.6] and [2.0, 3.9].

7.2 Estimation results

Table 9 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 shameful behavior transitions to pride-worthy behavior 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. 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.

7.3 Welfare effects of scaling up public recognition to all YOTA members

We use our structural estimates to assess the effects of scaling up the public recognition intervention to all members of the YOTA population. 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. Formal derivations of the equilibrium predictions are in Appendix E.2.3 and E.3.4.

We present the results in Table 10. Column (1) shows the net welfare effect from feelings of shame and pride, column (2) presents the predicted change in behavior, and column (3) presents the per-attendance emotional effect of public recognition, which is the ratio of columns (1) and (2). 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 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)-(5) explore the importance of varying ρ by considering the point estimates from the Prolific, Berkeley, and BU samples, as well as simply setting $\rho = 1$. Rows (6)-(10) consider the same values of ρ as rows (1)-(5) but assume less concavity, motivated by the smaller point estimates in the Berkeley and BU samples. Rows (11) and (12) hold constant the ρ estimated in the YMCA sample and consider concavity that is at the upper and lower end of the 95 percent confidence interval in the YMCA sample. Row (13) scales up the PRU estimate for the YMCA sample by a factor of two.

The estimated models generally predict that the net effect of shame and pride would be negative. This negative effect is particularly pronounced in the action-signaling model estimated off of the YMCA sample, where it is estimated to be $-\$3.41$ for a 1.75-visit increase in attendance, or approximately $-\$2$ per person per visit.

We show in Appendix B that revenue-neutral financial incentive schemes have higher welfare effects than public recognition when the net effect of shame and pride is negative, and are in fact pareto-dominating under some assumptions. Thus, public recognition generates a deadweight loss relative to financial incentives when the net emotional effect is negative. For example, in the action-signaling case where the net emotional effect is -$3.41 per person, there exists a revenue-neutral financial incentive scheme that generates the same change in behavior as does public recognition, but makes each individual better off by $3.41. In Appendix B.2 we discuss the case in which incentives must be in the form of a subsidy that must be funded with distortionary taxation, and argue that incentives are still likely to be preferred to public recognition in at least some cases.

While the welfare implications are sensitive to both the value of ρ and the degree of concavity, there are only two cases in which the net effect of shame and pride is positive. The first case corresponds to the particularly small value of ρ that is estimated in the Prolific sample. The second case corresponds to the value of ρ estimated in the Berkeley sample, under the assumption that concavity is only half as big as the estimate in the YMCA sample.

Although our conclusions about the consequences of scaling up public recognition rely on a number of strong assumptions, the estimates in the table illustrate that even when setting aside the ethical objections to exposing individuals to shame (Massaro, 1991; Nussbaum, 2009), the costs of public recognition can be high from a standard economic efficiency perspective. The estimates in the table show that the aggregate emotional effects can be substantially negative under empirically realistic assumptions the structure of the PRU. At the same time, our results also illustrate realistic scenarios in which the net effect could positive. This highlights the potential pitfalls of both ignoring the emotional effects of public recognition, and of over-generalizing about its effects without domain-specific measurement and careful modeling.

8 Discussion

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 the emotional effects of public recognition are significant and highly unequal: a large share of individuals experience strong feelings of shame and a large share of individuals experience strong feelings of 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 do not speak to the benefits of the behavior change itself. Finding prevalent feelings of shame does not by itself imply that public recognition decreases welfare. Rather, as formalized in Appendix B, our results about utility from public recognition speak most directly to whether a different policy lever, such as financial incentives, might be more efficient in creating the same behavior change. We also 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., 2009).

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. Although repeated application of our methodology across a variety of different contexts is likely to uncover predictable patterns in how utility from public recognition varies across different settings, we suspect that there will always be some ex-ante difficult-to-predict context dependence. Ultimately, the most accurate evaluation of a public recognition program will involve an application of our methods to the specific context in which the public recognition program is being implemented.

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), 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 easily augmented to further study how the effects of public recognition vary with the signal structure, which would provide additional tests of models of social signaling.

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' payoffs from public recognition. By providing a direct estimate of utility from public recognition, our methodology can thus enable more direct tests of phenomena such as the overjustification effect and motivation crowding (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).

Despite the open questions and the weaknesses of our approach that we hope future work will evaluate and address, our approach nevertheless provides a tractable toolkit for evaluating public recognition interventions and, with some extension, other social influence levers. Although non-financial policy instruments such as these 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* (Bernheim and Taubinsky, 2018). 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 wellbeing is intensely sensitive to the experience of shame and pride.

References

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

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," *working paper*, 2020.

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.

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.

_ and Matthew Rabin, "An experiment on time preference and misprediction in unpleasant tasks," *Review of Economic Studies*, 2019, *86* (3), 941–975.

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.

_ , 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.

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.

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

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," *Review of Economic Studies*, 2019. Forthcoming.

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

Carrera, Mariana, Heather Royer, Mark Stehr, Justin Sydnor, and Dmitry Taubinsky, "How are Preferences For Commitment Revealed?," 2019. Working Paper.

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.

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

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.

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

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

DellaVigna, Stefano, "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.

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

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.

Fedyk, Anastassia, "Asymmetric Naivete: Beliefs About Self-Control," 2018. Working Paper.

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

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.

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

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

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.

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

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.

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.

Laibson, David, "Why Don't Present-Baised Agents Make Commitments?," *American Economic Review*, 2015, *105* (5), 267–272.

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

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.

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

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

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.

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.

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.

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.

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.

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.

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.

Figures and Tables

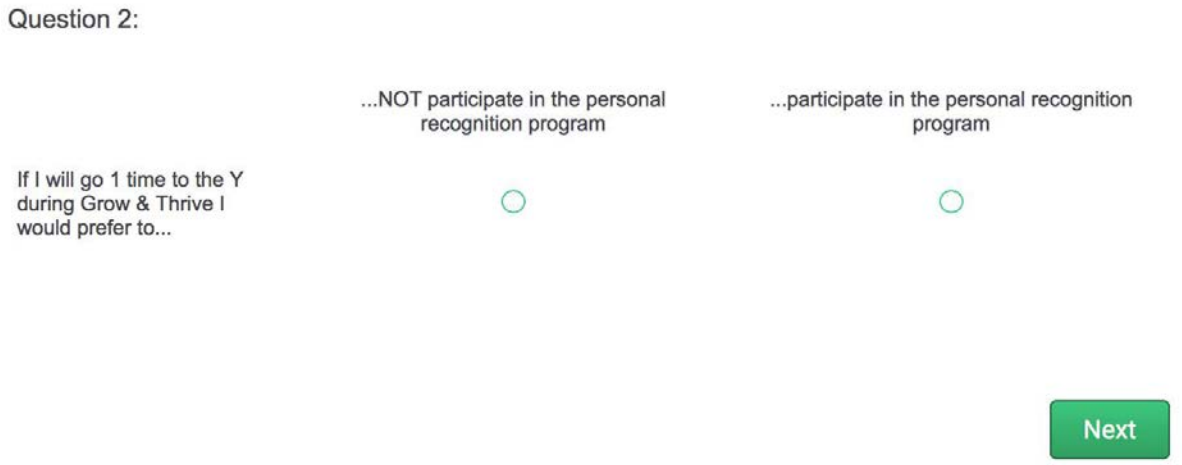
Figure 1: Illustration of public recognition information

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
..		

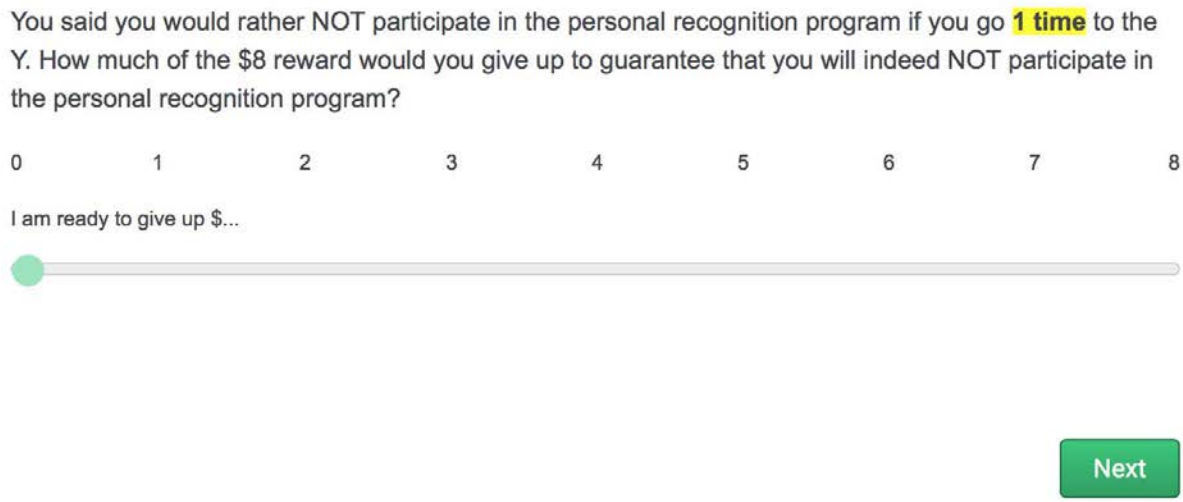
Notes: This figure shows an illustration of how individuals' attendance was publicized in the YMCA experiment.

Figure 2: An example of WTP for public recognition in the YMCA experiment

(a) First step of elicitation

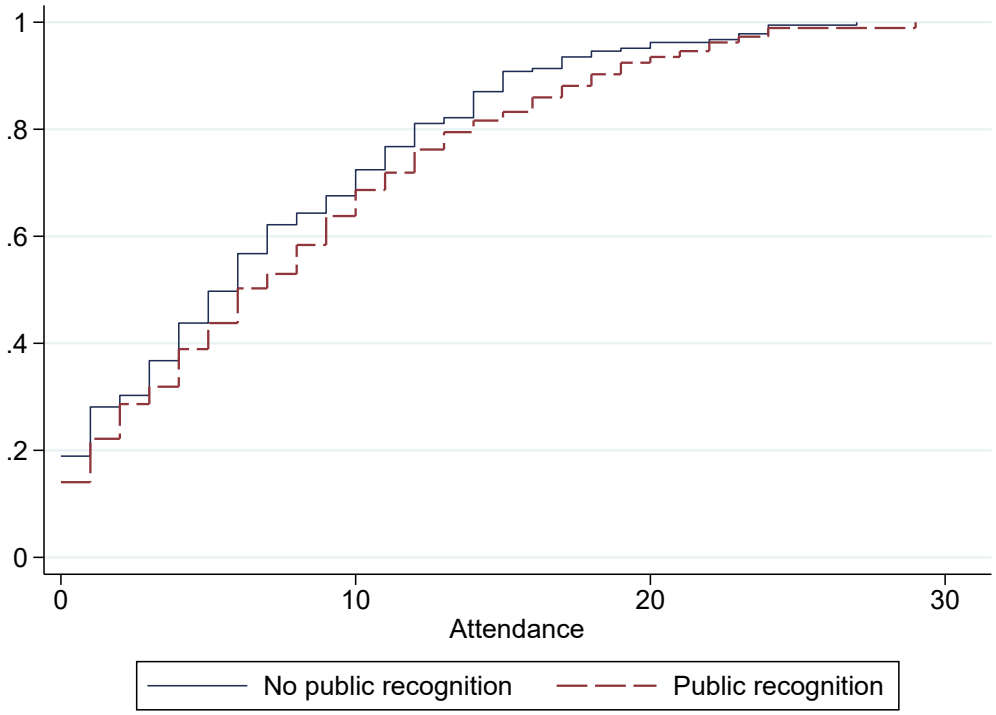


(b) Second step of elicitation



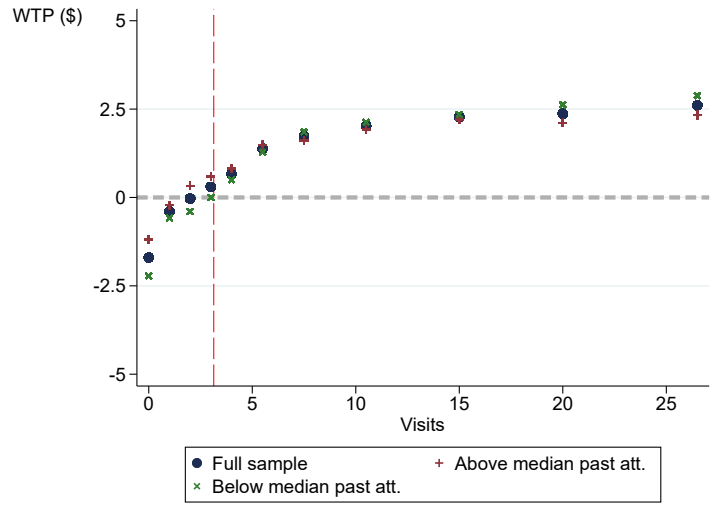
Notes: These figures present screenshots of the procedure for elicitation of WTP for public recognition. The example above shows the elicitation of WTP for attending the YMCA once during Grow & Thrive. The top panel presents the first step of the elicitation, where participants are asked whether they 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 3: Cumulative distributions of attendance during the YMCA experiment, by treatment

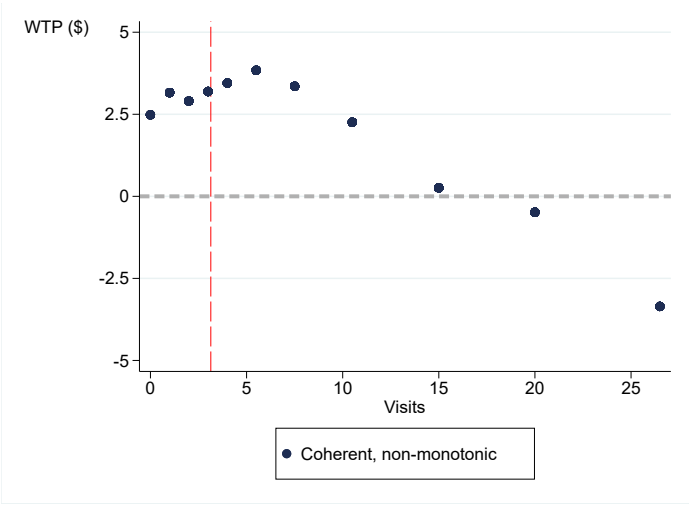


Notes: This figure plots the cumulative distribution functions of attendance during the experiment, by whether participants were in the public recognition group. The analysis excludes 15 participants with "incoherent" preferences for public recognition.

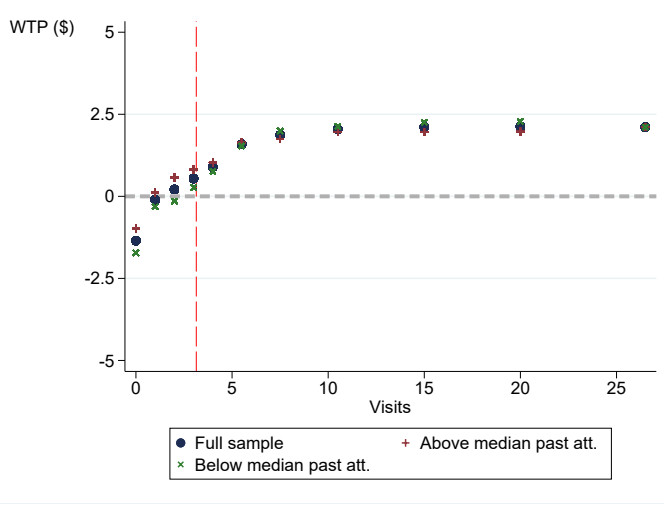
Figure 4: WTP for public recognition, by YMCA attendance



(a) Monotonic sample



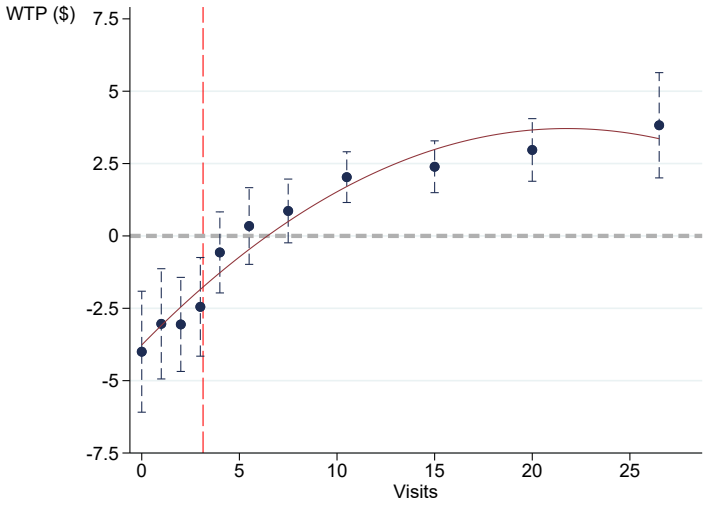
(b) Coherent but non-monotonic participants



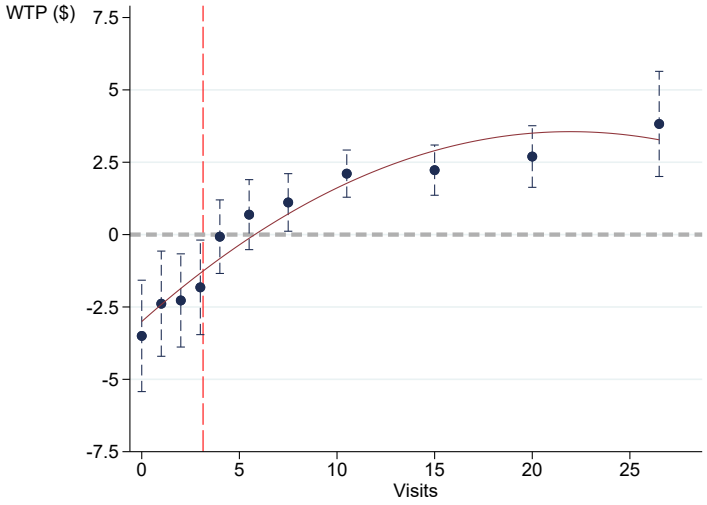
(c) Coherent sample

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 the full monotonic sample, as well as for the monotonic sample split by median past attendance. Panel (b) reports the average WTP for participants included in the coherent sample, but with non-monotonic preferences for public recognition. Panel (c) reports the average WTP for the full coherent sample, as well as for the coherent sample split by median past attendance. The average YOTA attendance during Grow & Thrive is indicated by the dashed red line.

Figure 5: WTP for public recognition by YMCA attendance, restricting to questions about visits close to participants' expectations



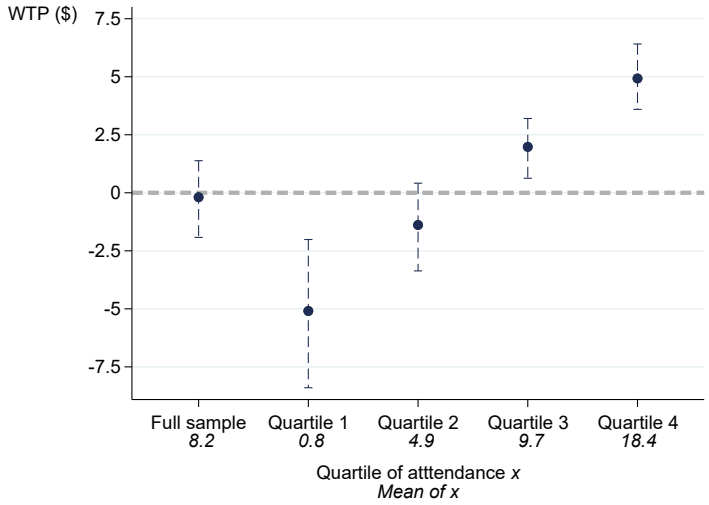
(a) Monotonic sample



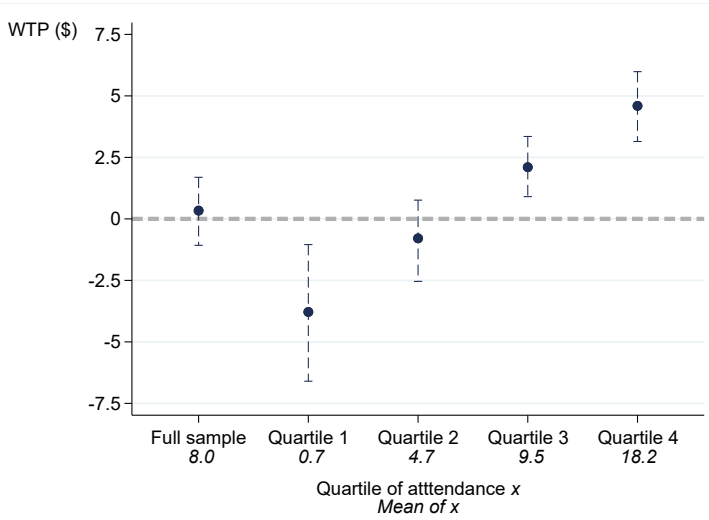
(b) Coherent sample

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. 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 average YOTA attendance is indicated by the dashed red line. Panel (a) restricts to the monotonic sample and panel (b) restricts to the coherent sample. 95 percent confidence intervals are constructed from standard errors clustered by participant. Quadratic fit curves are plotted in red.

Figure 6: The net effect of shame and pride in the YMCA experiment



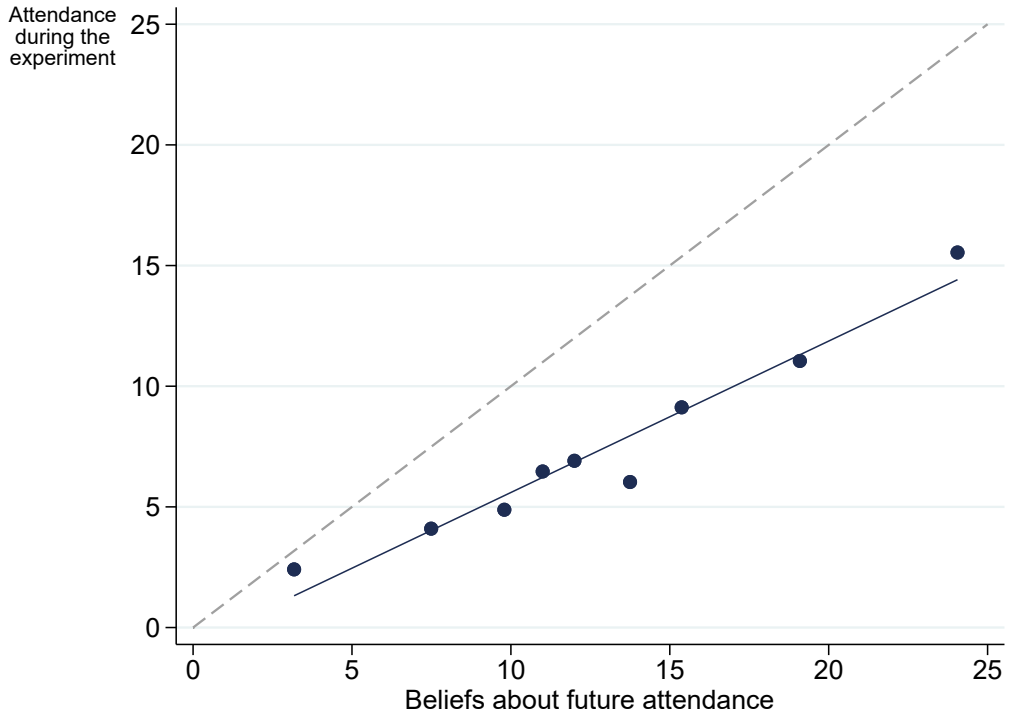
(a) Monotonic sample



(b) Coherent sample

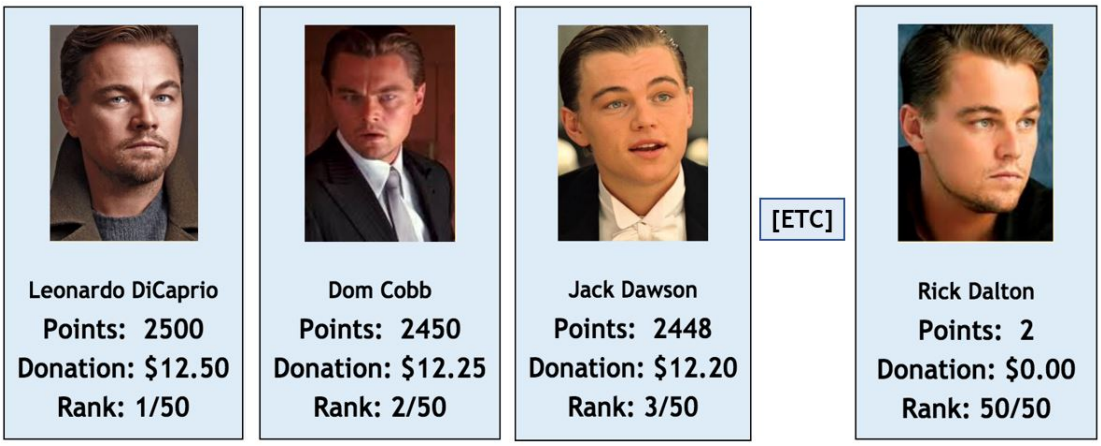
Notes: These figures plot the average realized public recognition payoff of participants assigned public recognition, for both the full sample and each quartile of actual attendance. The average attendance is reported below each subsample label. For panel (a), a participant's payoff is defined as the WTP predicted by the regression in column (4) of Table 4a, given the participant's realized attendance. For panel (b), it is defined as the WTP predicted by the regression in column (4) of Table 4b. Panel (a) restricts to the monotonic sample and panel (b) restricts to the coherent sample. Bootstrapped percentile-based confidence intervals, sampled by participant with 1000 iterations, are displayed.

Figure 7: Actual versus forecasted attendance in the YMCA experiment



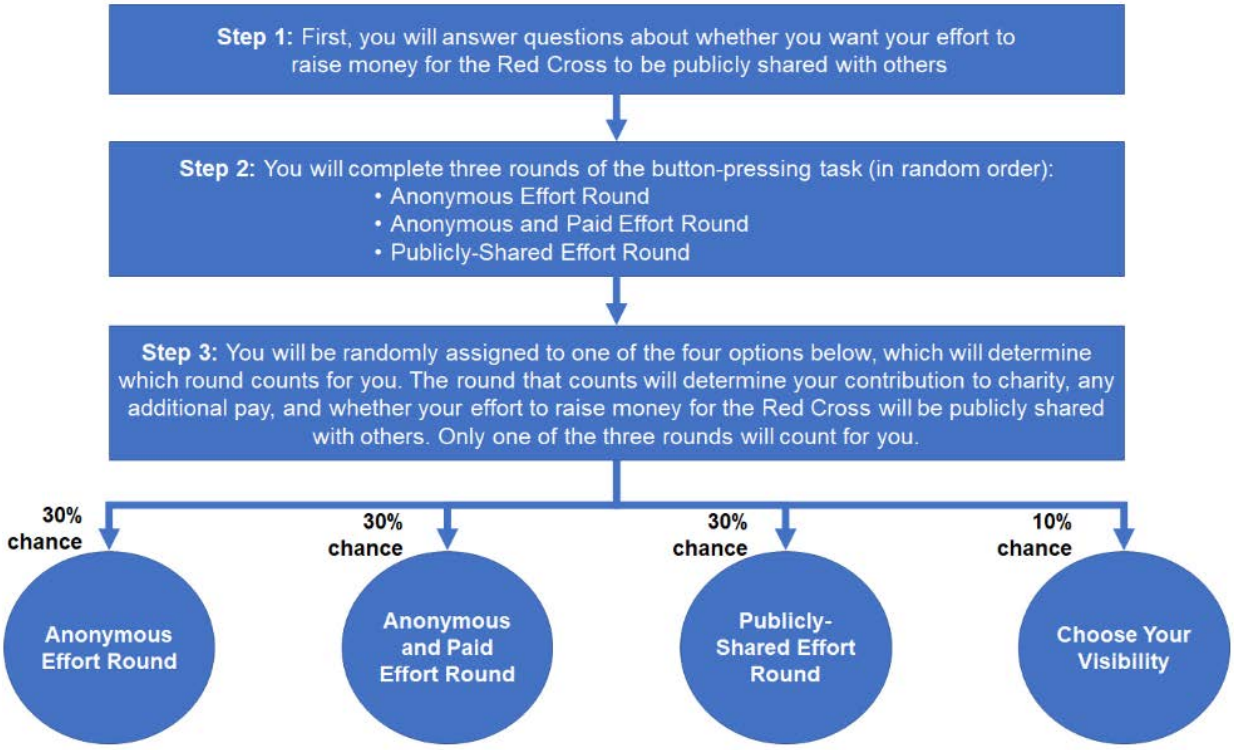
Notes: This figure plots the relationship between participants' forecasted and actual attendance. For participants in the public recognition group, we compare attendance to their beliefs about attendance if they are randomized into the public recognition group. For participants not in the public recognition 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.

Figure 8: Screenshot of public recognition example given to participants in the charitable contribution experiment



Notes: This figure contains a screenshot of the example given to participants of how their performance would be publicly recognized. For the Prolific sample, names were not included. For each sample and experimental group size, donations were scaled and ranks were adjusted accordingly.

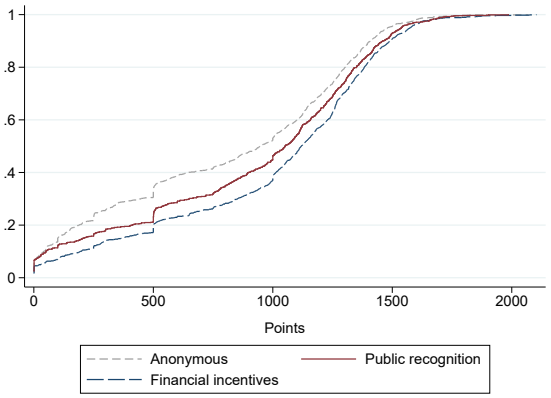
Figure 9: Overview of the charitable contribution experiment



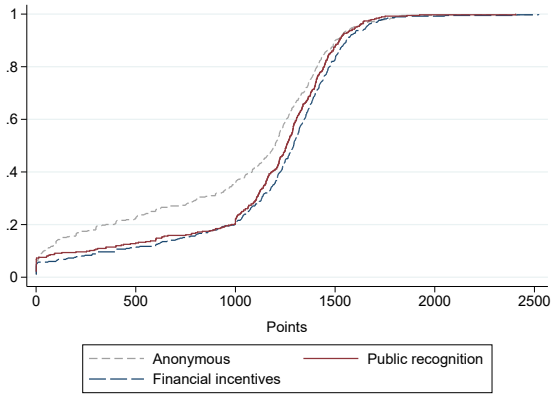
Notes: This figure presents a screenshot of the flowchart that was shown to participants in the instructions of the charitable contribution experiment.

Figure 10: Cumulative distributions of points scored in each of the three rounds of the charitable contribution experiments

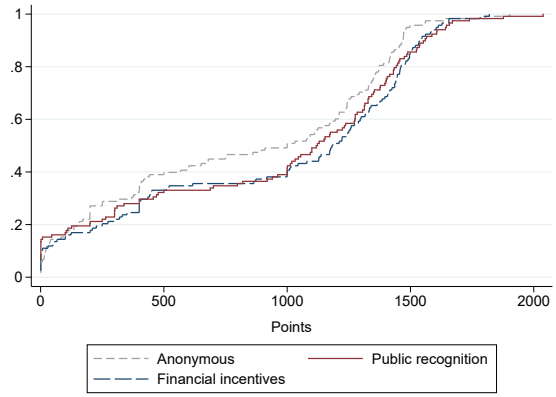
(a) Prolific



(b) Berkeley



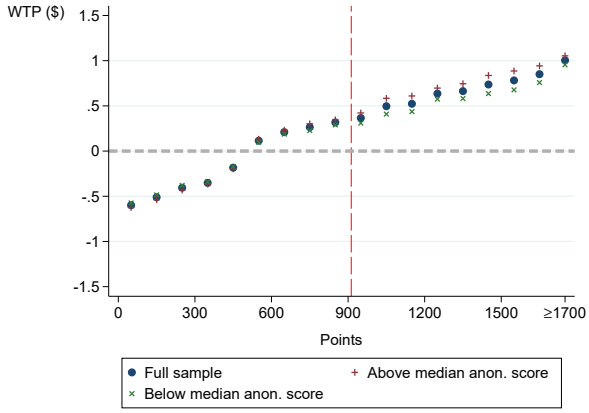
(c) BU



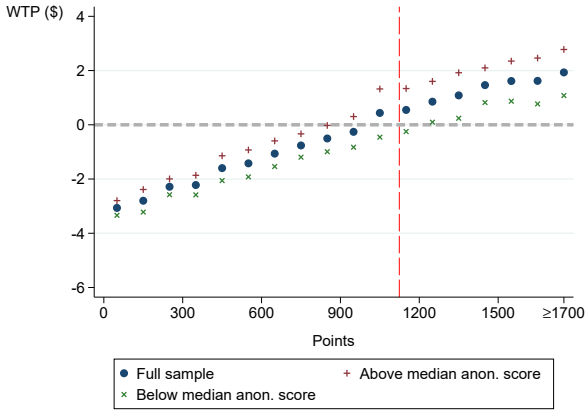
Notes: These figures plot the cumulative distribution functions of points scored in the Anonymous Effort Round, the Anonymous and Paid Effort Round, and the Publicly-Shared Effort Round. Panel (a) presents results for the Prolific sample, panel (b) presents results for the Berkeley sample, and panel (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 11: Willingness to pay for public recognition by effort in the charitable contribution experiments

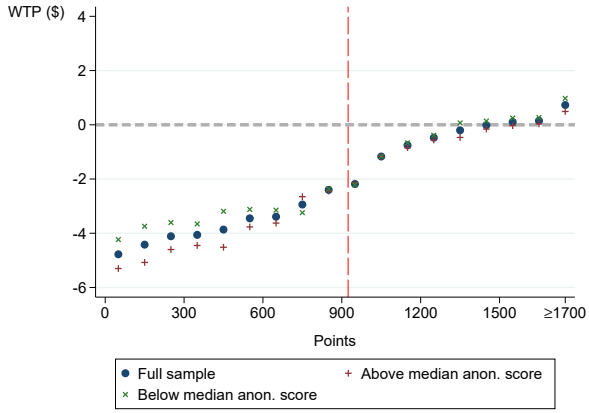
(a) Prolific



(b) Berkeley



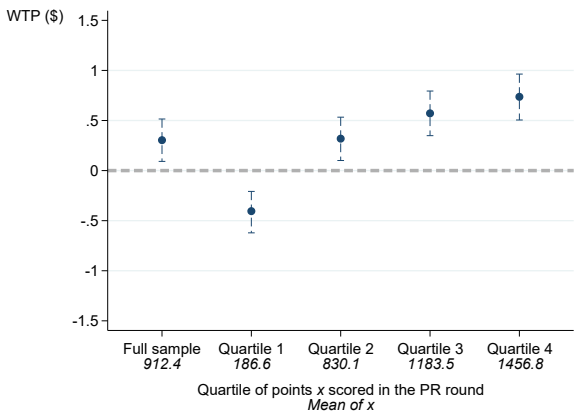
(c) BU



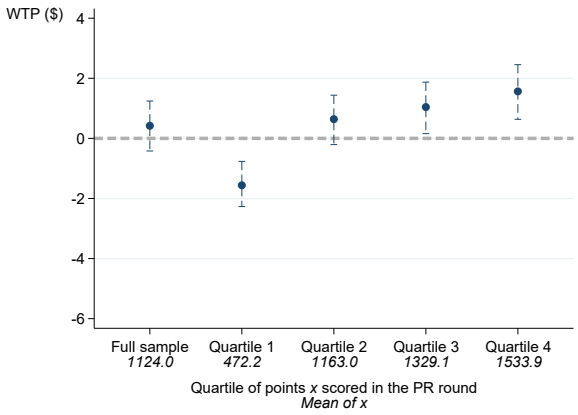
Notes: These figures plot the average WTP for public recognition by each of the 18 possible intervals of points scored. The WTP is plotted at the midpoint of each of the first seventeen intervals and at ≥1700 points for the 1700 or more points interval. Panel (a) presents results for the Prolific sample, 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 12: The net effect of shame and pride in the charitable contribution experiments

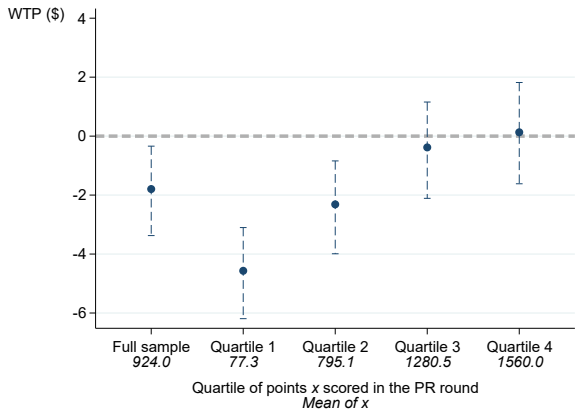
(a) Prolific



(b) Berkeley



(c) BU



Notes: These figures plot the average realized public recognition 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 public recognition 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: 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 2: The impact of public recognition on YMCA attendance

	(1)	(2)	(3)	(4)	(5)	(6)
Public recognition	1.20	1.26***	1.34***	1.10	1.19***	1.27***
	(0.73)	(0.48)	(0.47)	(0.69)	(0.46)	(0.45)
Avg. past att.		0.89***	0.78***		0.88***	0.77***
		(0.04)	(0.05)		(0.04)	(0.05)
Beliefs			0.20***			0.19***
			(0.05)			(0.05)
Control mean	6.95	6.95	6.95	6.91	6.91	6.91
	(0.49)	(0.49)	(0.49)	(0.47)	(0.47)	(0.47)
Sample	Mon	Mon	Mon	Coh	Coh	Coh
N. Subjects	339	339	339	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. Columns (1)-(3) restrict to the monotonic sample, while columns (4)-(6) restrict to the coherent sample. 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 3: WTP for public recognition by YMCA attendance

	(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	0.13***	0.39***	0.25***	0.68***	0.10***	0.36***	0.19***	0.62***
	(0.01)	(0.04)	(0.03)	(0.08)	(0.01)	(0.04)	(0.03)	(0.07)
N. visits sq.		-0.01***		-0.02***		-0.01***		-0.02***
		(0.00)		(0.00)		(0.00)		(0.00)
Constant	-0.14	-0.91***	-0.69	-2.00***	0.20	-0.57*	-0.03	-1.35**
	(0.32)	(0.34)	(0.63)	(0.68)	(0.30)	(0.32)	(0.59)	(0.63)
$-R''/R'$	–	0.052	–	0.051	–	0.057	–	0.056
95% CI	–	[0.049, 0.055]	–	[0.047, 0.054]	–	[0.053, 0.061]	–	[0.052, 0.060]
$-R''/R' \times$ SD	–	0.254	–	0.247	–	0.277	–	0.271
95% CI	–	[0.239, 0.269]	–	[0.230, 0.264]	–	[0.258, 0.295]	–	[0.251, 0.291]
Sample	Mon	Mon	Mon	Mon	Coh	Coh	Coh	Coh
Observations	3729	3729	3729	3729	4070	4070	4070	4070
N. Subjects	339	339	339	339	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 $-R''_{exp}/R'_{exp}(0)$ and $-R''_{exp}/R'_{exp}(0) \times SD$, where $SD = 4.86$ is the standard deviation attendance for the general YOTA population. Columns (1)-(3) restrict to the monotonic sample, while columns (4)-(6) restrict to the coherent sample. 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 4: WTP for public recognition by YMCA attendance, restricting to questions about visits close to participants' expectations

(a) Monotonic sample

	(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)
$-R''/R'$	–	0.048	–	0.045	–	0.045	–	0.043
95% CI	–	[0.036, 0.060]	–	[0.029, 0.061]	–	[0.031, 0.058]	–	[0.027, 0.059]
$-R''/R' \times$ SD	–	0.234	–	0.220	–	0.217	–	0.209
95% CI	–	[0.175, 0.292]	–	[0.143, 0.297]	–	[0.153, 0.280]	–	[0.129, 0.288]
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

(b) Coherent sample

	(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)
$-R''/R'$	–	0.049	–	0.046	–	0.044	–	0.042
95% CI	–	[0.035, 0.062]	–	[0.028, 0.063]	–	[0.029, 0.059]	–	[0.023, 0.061]
$-R''/R' \times$ SD	–	0.237	–	0.223	–	0.213	–	0.205
95% CI	–	[0.171, 0.302]	–	[0.137, 0.308]	–	[0.140, 0.286]	–	[0.113, 0.296]
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 $-R''_{exp}/R'_{exp}(0)$ and $-R''_{exp}/R'_{exp}(0) \times SD$, where $SD = 4.86$ is the standard deviation attendance for the general YOTA population. Panel (a) restricts to the monotonic sample and panel (b) restricts to the coherent sample. 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 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 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)
$-R''/R'$	–	0.047	–	0.021	–	-0.015
95% CI	–	[0.036, 0.059]	–	[-0.009, 0.051]	–	[-0.097, 0.068]
$-R''/R' \times SD$	–	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 $-R''_{exp}/R'_{exp}(0)$ and $-R''_{exp}/R'_{exp}(0) \times SD$, where SD is the standard deviation of points scored in the anonymous round (in units of hundreds of points). 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 7: 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 ≥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.

Table 8: 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)
$-R''/R'$	–	0.033	–	0.001	–	-0.017
95% CI	–	[-0.014, 0.081]	–	[-0.105, 0.107]	–	[-0.206, 0.173]
$-R''/R' \times$ 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 $-R''_{exp}/R'_{exp}(0)$ and $-R''_{exp}/R'_{exp}(0) \times SD$, where SD is the standard deviation of points scored in the anonymous round (in units of hundreds of points). 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 9: 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
	[0.35,0.92]	[-0.038,-0.003]	[0.94,2.52]	[0.20,1.95]
Prolific	0.12	-0.004	0.58	0.09
	[0.09,0.14]	[-0.005,-0.002]	[0.40,0.80]	[0.07,0.11]
Berkeley	0.30	-0.004	0.87	0.23
	[0.23,0.37]	[-0.011,0.003]	[0.66,1.15]	[0.16,0.35]
BU	0.38	0.002	1.61	0.34
	[0.19,0.55]	[-0.009,0.014]	[1.14,2.37]	[0.14,1.78]

(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
	[0.28,2.34]	[-0.292,-0.003]	[0.31,2.19]	[0.20,1.95]
Prolific	1.24	-0.416	0.50	0.09
	[0.94,1.57]	[-0.679,-0.227]	[0.26,0.77]	[0.07,0.11]
Berkeley	1.26	-0.071	0.86	0.23
	[0.85,1.74]	[-0.223,0.058]	[0.61,1.18]	[0.16,0.35]
BU	1.14	0.021	1.68	0.34
	[0.09,2.66]	[-0.100,0.345]	[1.17,2.49]	[0.14,1.78]

(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
	[0.42,5.00]	[1.30,2.26]	[-1.27,3.19]
Prolific	2.29	1.82	0.47
	[1.74,2.97]	[1.56,2.07]	[-0.02,1.08]
Berkeley	2.17	1.78	0.39
	[1.42,3.11]	[1.36,2.22]	[-0.28,1.17]
BU	1.49	1.18	0.31
	[0.16,3.05]	[0.42,1.96]	[-0.90,1.69]

†: 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 7.1, equations (5) and (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 10: Welfare estimates of scaling up public recognition at the YMCA

(a) Action-signaling model

Row	Scenario	(1) Direct emotional effect	(2) Change in attendance	(3) $\frac{(1)}{(2)}$
1.	$\gamma_1^a = 0.64$, $\gamma_2^a = -0.020$, $\rho^a = 1.85$	-3.41	1.75	-1.95
2.	$\gamma_1^a = 0.64$, $\gamma_2^a = -0.020$, $\rho^a = 0.58$	0.70	1.23	0.57
3.	$\gamma_1^a = 0.64$, $\gamma_2^a = -0.020$, $\rho^a = 0.87$	-0.04	1.34	-0.03
4.	$\gamma_1^a = 0.64$, $\gamma_2^a = -0.020$, $\rho^a = 1.61$	-2.46	1.64	-1.49
5.	$\gamma_1^a = 0.64$, $\gamma_2^a = -0.020$, $\rho^a = 1$	-0.40	1.39	-0.29
6.	$\gamma_1^a = 0.64$, $\gamma_2^a = -0.010$, $\rho^a = 1.85$	-2.94	1.56	-1.88
7.	$\gamma_1^a = 0.64$, $\gamma_2^a = -0.010$, $\rho^a = 0.58$	0.94	1.31	0.72
8.	$\gamma_1^a = 0.64$, $\gamma_2^a = -0.010$, $\rho^a = 0.87$	0.15	1.37	0.11
9.	$\gamma_1^a = 0.64$, $\gamma_2^a = -0.010$, $\rho^a = 1.61$	-2.13	1.51	-1.41
10.	$\gamma_1^a = 0.64$, $\gamma_2^a = -0.010$, $\rho^a = 1$	-0.22	1.39	-0.16
11.	$\gamma_1^a = 0.64$, $\gamma_2^a = -0.038$, $\rho^a = 1.85$	-4.26	2.11	-2.02
12.	$\gamma_1^a = 0.64$, $\gamma_2^a = -0.003$, $\rho^a = 1.85$	-2.60	1.44	-1.81
13.	$\gamma_1^a = 1.29$, $\gamma_2^a = -0.040$, $\rho^a = 1.85$	-9.64	3.80	-2.54

(b) Characteristics-signaling model

Row	Scenario	(1) Direct emotional effect	(2) Change in attendance	(3) $\frac{(1)}{(2)}$
1.	$\gamma_1^\theta = 1.28$, $\gamma_2^\theta = -0.079$, $\rho^\theta = 1.40$	-1.18	1.49	-0.79
2.	$\gamma_1^\theta = 1.28$, $\gamma_2^\theta = -0.079$, $\rho^\theta = 0.50$	0.49	1.26	0.39
3.	$\gamma_1^\theta = 1.28$, $\gamma_2^\theta = -0.079$, $\rho^\theta = 0.86$	-0.14	1.36	-0.10
4.	$\gamma_1^\theta = 1.28$, $\gamma_2^\theta = -0.079$, $\rho^\theta = 1.68$	-1.74	1.56	-1.12
5.	$\gamma_1^\theta = 1.28$, $\gamma_2^\theta = -0.079$, $\rho^\theta = 1$	-0.40	1.39	-0.29
6.	$\gamma_1^\theta = 1.28$, $\gamma_2^\theta = -0.040$, $\rho^\theta = 1.40$	-1.01	1.44	-0.70
7.	$\gamma_1^\theta = 1.28$, $\gamma_2^\theta = -0.040$, $\rho^\theta = 0.50$	0.73	1.32	0.55
8.	$\gamma_1^\theta = 1.28$, $\gamma_2^\theta = -0.040$, $\rho^\theta = 0.86$	0.05	1.37	0.04
9.	$\gamma_1^\theta = 1.28$, $\gamma_2^\theta = -0.040$, $\rho^\theta = 1.68$	-1.57	1.48	-1.06
10.	$\gamma_1^\theta = 1.28$, $\gamma_2^\theta = -0.040$, $\rho^\theta = 1$	-0.22	1.39	-0.16
11.	$\gamma_1^\theta = 1.28$, $\gamma_2^\theta = -0.292$, $\rho^\theta = 1.40$	-1.40	1.57	-0.90
12.	$\gamma_1^\theta = 1.28$, $\gamma_2^\theta = -0.003$, $\rho^\theta = 1.40$	-0.87	1.41	-0.62
13.	$\gamma_1^\theta = 2.56$, $\gamma_2^\theta = -0.159$, $\rho^\theta = 1.40$	-2.13	2.97	-0.72

Notes: These tables report welfare estimates based on the structural estimates of the action-signaling and characteristics-signaling models described in Section 7.1. In row (1), γ_1^j, γ_2^j, and ρ^j are set equal to the parameter estimates from the YMCA sample. In rows (2)-(4), ρ^j is set equal to ρ^j for the Prolific, Berkeley, and BU samples, respectively. Rows (6)-(10) repeat rows (1)-(5) with γ_2^j set equal to one-half of the estimate from the YMCA sample. In rows (11) and (12), γ_2^j is set equal to the lower-bound and upper-bound, respectively, of the confidence interval for γ_2^j estimated in Table 9. In row (13), γ_1^j and γ_2^j are set equal to twice the parameter estimates from the YMCA sample.

Online Appendix

Measuring the Welfare Effects of Shame and Pride

Luigi Butera, Robert Metcalfe, William Morrison, and Dmitry Taubinsky

Table of Contents

A 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 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. 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 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.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.3 The net effect of shame and pride

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:

Proposition 1. *Assume that S is increasing. If S is concave and w is increasing, then the net effect of shame and pride is negative. If S is convex and w is decreasing, then the net effect of shame and pride is positive.*

Proof. Suppose that S is concave and that w is increasing. Then Jensen's inequality implies that

$$\begin{aligned}
&\int_{\theta'\in\Theta}\int_{\sigma} S\left(\mathbb{E}[\theta|\sigma] - \frac{\int_{x\in\Theta} xw(x)dH^*(x)}{\int_{x\in\Theta} w(x)dH^*(x)}\right) dG(\sigma|\boldsymbol{a}(\theta'))dF(\theta') \\
\leq& S\left(\mathbb{E}[\theta|\sigma]dG(\sigma|\boldsymbol{a}(\theta'))dF(\theta') - \frac{\int_{x\in\Theta} xw(x)dH^*(x)}{\int_{x\in\Theta} w(x)dH^*(x)}\right) \\
=& S\left(\int_{x\in\Theta} xdH^*(x) - \frac{\int_{x\in\Theta} xw(x)dH^*(x)}{\int_{x\in\Theta} w(x)dH^*(x)}\right) \qquad (9) \\
\leq& S(0) = 0. \qquad (10)
\end{aligned}$$

Line (10) follows from line (9) 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. □

B Deadweight loss relative to financial incentives

B.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. 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 emotional effect of 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 emotional consequences of 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.

B.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 effect of shame and pride 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 effect of shame and pride.

As an example, suppose that $p(a)$ is required to be non-negative, and return to the welfare estimate in column (1) of Table 10a, where the net effect of shame and pride 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).[33]

B.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 effect of shame and pride. 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

$$\begin{aligned}\Delta W^R &= \int_{\theta\in\Theta} \left(\Delta(\theta)e(\theta) + r(\theta)\right) dF(\theta) \\ &= \bar{\Delta}\bar{e} + \bar{r} + Cov[\Delta(\theta), e(\theta)]. \end{aligned} \tag{11}$$

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.

[33] 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 $3.41/4.89 \approx 0.7$.

The incremental effect of these financial incentives is given by

$$\begin{aligned}\Delta W^p &= \int_{\theta\in\Theta} \left(\Delta_p(\theta)e(\theta) + p(\boldsymbol{a}(\theta))\right) dF(\theta) - \lambda \int_{\theta\in\Theta} p(\boldsymbol{a}(\theta))dF(\theta) \\ &= \bar{\Delta}\bar{e} + Cov[\Delta_p(\theta), e(\theta)] + (1-\lambda)\bar{p}. \end{aligned} \tag{12}$$

Equations (11) and (12) imply that the difference between the welfare effect of public recognition and financial incentives is given by

$$\underbrace{-\bar{r}}_{\text{net emotional effect}} + \underbrace{Cov[(\Delta_p(\theta) - \Delta(\theta), e(\theta)]}_{\text{relative targeting}} + \underbrace{(1-\lambda)\bar{p}}_{\text{cost of public funds}} . \tag{13}$$

Equation (13) shows that in addition to the net effect of shame and pride, 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 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 B.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.

C Supplementary empirical results for YMCA experiment

C.1 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 4, but considers visits within 4 of past attendance, rather than expectations.

Table A1: WTP for public recognition by YMCA attendance: heterogeneity along average past attendance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model	OLS	Tobit	OLS	Tobit	OLS	Tobit	OLS	Tobit
Dependent var.	WTP	WTP	WTP	WTP	WTP	WTP	WTP	WTP
N. visits	0.42***	0.73***	0.39***	0.68***	0.69***	1.04***	0.63***	0.94***
	(0.05)	(0.11)	(0.05)	(0.10)	(0.19)	(0.37)	(0.18)	(0.33)
N. visits sq.	-0.01***	-0.02***	-0.01***	-0.02***	-0.02**	-0.02	-0.01**	-0.02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Avg. past att.	0.03	0.04	0.02	0.02	-0.03	-0.11	0.14	0.24
	(0.06)	(0.12)	(0.06)	(0.11)	(0.39)	(0.77)	(0.35)	(0.66)
N. visits × Past att.	-0.00	-0.01	-0.01	-0.01	-0.01	-0.01	-0.03	-0.04
	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.08)	(0.04)	(0.07)
N. visits sq. × Past att.	0.00	0.00	0.00	0.00	0.00	-0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	-1.11**	-2.26**	-0.70	-1.48*	-3.42**	-5.65**	-2.99**	-4.93**
	(0.48)	(0.95)	(0.45)	(0.87)	(1.35)	(2.57)	(1.20)	(2.18)
Sample	Mon	Mon	Coh	Coh	Mon	Mon	Coh	Coh
Restriction	All	All	All	All	≤ 4	≤ 4	≤ 4	≤ 4
Observations	3729	3729	4070	4070	830	830	923	923
N. Subjects	339	339	370	370	339	339	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) use all 11 intervals of future attendance, while columns (5)-(8) restrict to intervals with a midpoint within 4 of a participant's predicted attendance if assigned public recognition. Columns (1), (2), (5), and (6) exclude 46 participants with non-monotonic preferences for public recognition. Columns (3), (4), (7), and (8) exclude 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 A2: WTP for public recognition by YMCA attendance: using number of visits within 4 of past attendance

(a) Monotonic sample

	(1)	(2)	(3)	(4)
Model	OLS	OLS	Tobit	Tobit
Dependent var.	WTP	WTP	WTP	WTP
N. visits	0.26***	0.54***	0.46***	0.92***
	(0.05)	(0.11)	(0.11)	(0.23)
N. visits sq.		-0.02***		-0.03**
		(0.01)		(0.01)
Constant	-0.79*	-1.40***	-1.73*	-2.72***
	(0.44)	(0.49)	(0.90)	(1.00)
$-R''/R'$	–	0.063	–	0.061
95% CI	–	[0.042, 0.084]	–	[0.035, 0.087]
$-R''/R' \times$ SD	–	0.306	–	0.297
95% CI	–	[0.203, 0.408]	–	[0.169, 0.425]
Observations	1503	1503	1503	1503
N. Subjects	339	339	339	339

(b) Coherent sample

	(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)
$-R''/R'$	–	0.067	–	0.066
95% CI	–	[0.046, 0.088]	–	[0.041, 0.092]
$-R''/R' \times$ SD	–	0.326	–	0.322
95% CI	–	[0.224, 0.427]	–	[0.198, 0.445]
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 4.51 for the monotonic sample control group, 4.42 for the coherent sample control group, and is 3.19 for the general YOTA population. 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. Panel (a) excludes 46 participants with non-monotonic preferences for public recognition, Panel (b) 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.

C.2 Excluding high visits intervals

Table A3: WTP for public recognition by YMCA attendance in monotonic sample, 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	0.18***	0.50***	0.32***	0.88***	0.24***	0.59***	0.43***	1.03***
	(0.02)	(0.05)	(0.04)	(0.10)	(0.02)	(0.06)	(0.05)	(0.12)
N. visits sq.		-0.02***		-0.03***		-0.02***		-0.04***
		(0.00)		(0.00)		(0.00)		(0.01)
Constant	-0.35	-1.13***	-1.05	-2.41***	-0.59*	-1.28***	-1.49**	-2.67***
	(0.32)	(0.34)	(0.64)	(0.69)	(0.33)	(0.34)	(0.67)	(0.70)
$-R''/R'$	–	0.067	–	0.066	–	0.082	–	0.080
95% CI	–	[0.063, 0.071]	–	[0.061, 0.071]	–	[0.074, 0.089]	–	[0.071, 0.088]
$-R''/R' \times$ SD	–	0.326	–	0.321	–	0.397	–	0.388
95% CI	–	[0.304, 0.347]	–	[0.297, 0.345]	–	[0.361, 0.433]	–	[0.346, 0.429]
Sample	Mon	Mon	Mon	Mon	Mon	Mon	Mon	Mon
Excl. int.	Top	Top	Top	Top	Top 2	Top 2	Top 2	Top 2
Observations	3390	3390	3390	3390	3051	3051	3051	3051
N. Subjects	339	339	339	339	339	339	339	339

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 $-R''_{exp}/R'_{exp}(0)$ and $-R''_{exp}/R'_{exp}(0) \times SD$, where $SD = 4.86$ is the standard deviation attendance for the general YOTA population. This analysis excludes 46 participants with non-monotonic 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 A4: WTP for public recognition by YMCA attendance in coherent sample, excluding high number of visits questions

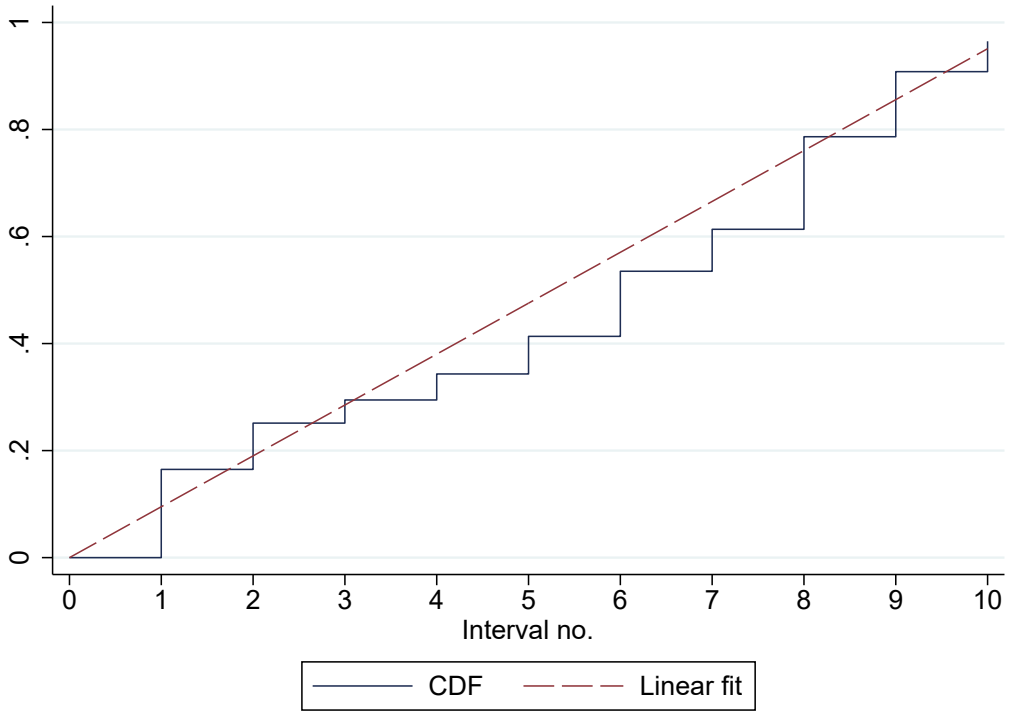
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model	OLS	OLS	Tobit	Tobit	OLS	OLS	Tobit	Tobit
Dependent var.	WTP	WTP	WTP	WTP	WTP	WTP	WTP	WTP
N. visits	0.15***	0.47***	0.26***	0.81***	0.21***	0.57***	0.37***	0.99***
	(0.02)	(0.05)	(0.04)	(0.09)	(0.02)	(0.06)	(0.05)	(0.12)
N. visits sq.		-0.02***		-0.03***		-0.03***		-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)
$-R''/R'$	–	0.071	–	0.070	–	0.087	–	0.086
95% CI	–	[0.066, 0.076]	–	[0.065, 0.076]	–	[0.080, 0.095]	–	[0.077, 0.095]
$-R''/R' \times$ SD	–	0.347	–	0.342	–	0.425	–	0.419
95% CI	–	[0.322, 0.371]	–	[0.315, 0.369]	–	[0.387, 0.463]	–	[0.376, 0.461]
Sample	Coh	Coh	Coh	Coh	Coh	Coh	Coh	Coh
Excl. int.	Top	Top	Top	Top	Top 2	Top 2	Top 2	Top 2
Observations	3700	3700	3700	3700	3330	3330	3330	3330
N. Subjects	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 $-R''_{exp}/R'_{exp}(0)$ and $-R''_{exp}/R'_{exp}(0) \times SD$, where $SD = 4.86$ is the standard deviation attendance for the general YOTA population. 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.

C.3 Rescaling the visits intervals to have equal width

Figure A1 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. Tables A5 and A6 show that the SRU is still estimated to be highly concave when we index intervals not by their midpoint, but instead by their sequential order. In particular,

Figure A1: 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.

Table A5: WTP for public recognition by index of attendance interval

(a) Monotonic sample

	(1)	(2)	(3)	(4)
Model	OLS	OLS	Tobit	Tobit
Dependent var.	WTP	WTP	WTP	WTP
Interval no.	0.40***	0.70***	0.72***	1.19***
	(0.04)	(0.08)	(0.08)	(0.16)
Interval no. sq.		-0.03***		-0.05***
		(0.01)		(0.01)
Constant	-0.98***	-1.44***	-2.18***	-2.89***
	(0.35)	(0.34)	(0.70)	(0.70)
$-R''/R'$	–	0.086	–	0.079
95% CI	–	[0.063, 0.110]	–	[0.051, 0.106]
$-R''/R' \times$ SD	–	0.420	–	0.384
95% CI	–	[0.306, 0.534]	–	[0.249, 0.518]
Observations	3729	3729	3729	3729
N. Subjects	339	339	339	339

(b) Coherent sample

	(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)
$-R''/R'$	–	0.111	–	0.106
95% CI	–	[0.090, 0.132]	–	[0.082, 0.130]
$-R''/R' \times$ SD	–	0.541	–	0.515
95% CI	–	[0.439, 0.642]	–	[0.398, 0.632]
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 $-R''_{exp}/R'_{exp}(0)$ and $-R''_{exp}/R'_{exp}(0) \times SD$, where $SD = 4.86$ is the standard deviation attendance for the general YOTA population. Panel (a) excludes 46 participants with non-monotonic preferences for public recognition. Panel (b) 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 A6: WTP for public recognition by index of attendance interval, restricting to number of visits questions within 4 of predicted PR attendance

(a) Monotonic sample

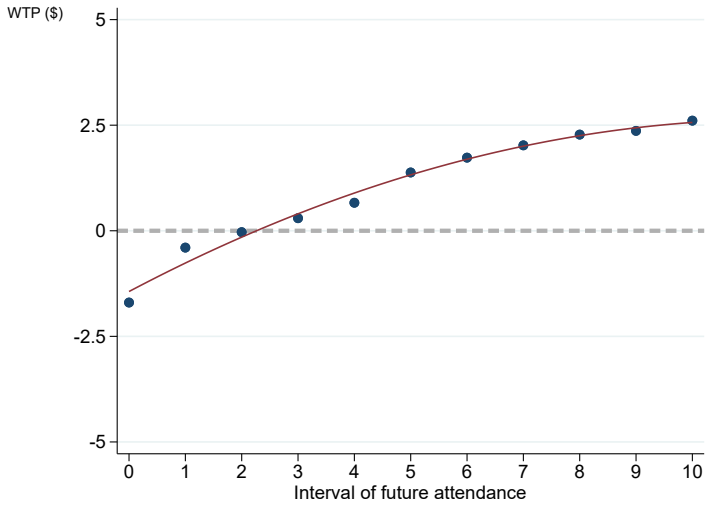
	(1)	(2)	(3)	(4)
Model	OLS	OLS	Tobit	Tobit
Dependent var.	WTP	WTP	WTP	WTP
Interval no.	0.82***	1.04***	1.39***	1.61**
	(0.13)	(0.37)	(0.26)	(0.75)
Interval no. sq.		-0.02		-0.02
		(0.03)		(0.06)
Constant	-4.05***	-4.50***	-7.21***	-7.66***
	(0.96)	(1.14)	(1.87)	(2.39)
$-R''/R'$	–	0.040	–	0.026
95% CI	–	[-0.059, 0.140]	–	[-0.110, 0.162]
$-R''/R' \times$ SD	–	0.195	–	0.125
95% CI	–	[-0.289, 0.678]	–	[-0.536, 0.786]
Observations	830	830	830	830
N. Subjects	339	339	339	339

(b) Coherent sample

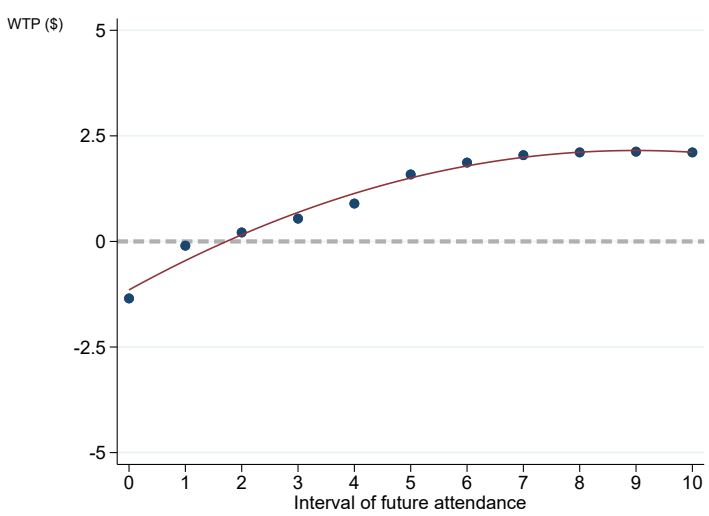
	(1)	(2)	(3)	(4)
Model	OLS	OLS	Tobit	Tobit
Dependent var.	WTP	WTP	WTP	WTP
Interval no.	0.70***	0.99***	1.17***	1.50**
	(0.12)	(0.34)	(0.23)	(0.67)
Interval no. sq.		-0.03		-0.03
		(0.03)		(0.06)
Constant	-3.18***	-3.75***	-5.56***	-6.22***
	(0.90)	(1.06)	(1.67)	(2.09)
$-R''/R'$	–	0.055	–	0.042
95% CI	–	[-0.033, 0.144]	–	[-0.077, 0.161]
$-R''/R' \times$ SD	–	0.270	–	0.205
95% CI	–	[-0.160, 0.700]	–	[-0.375, 0.785]
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 $-R''_{exp}/R'_{exp}(0)$ and $-R''_{exp}/R'_{exp}(0) \times SD$, where $SD = 4.86$ is the standard deviation attendance for the general YOTA population. Panel (a) excludes 46 participants with non-monotonic preferences for public recognition. Panel (b) 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 A2: WTP for public recognition by index of interval



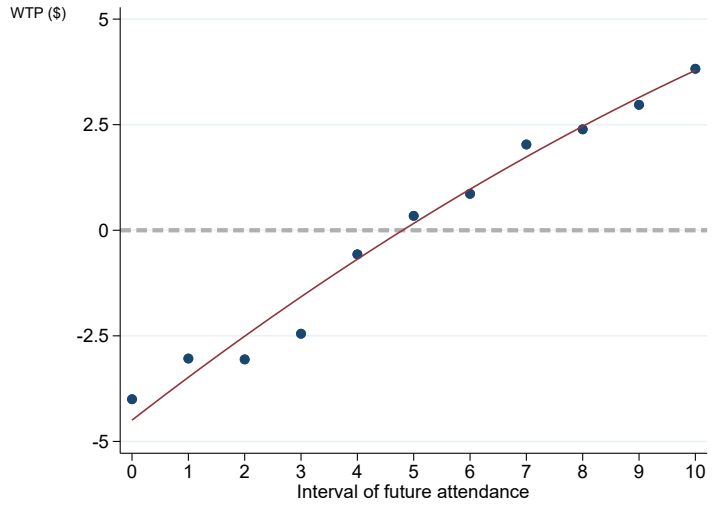
(a) Monotonic sample



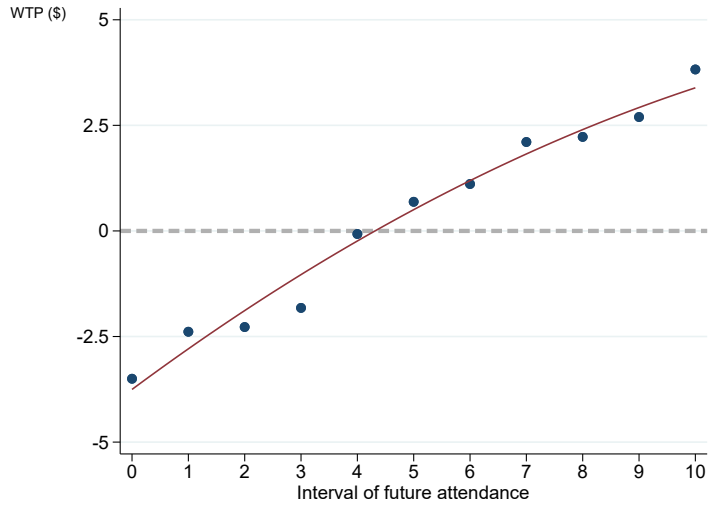
(b) Coherent sample

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. Panel (a) excludes 46 participants with non-monotonic preferences for public recognition. Panel (b) excludes 15 participants with "incoherent" preferences for public recognition.

Figure A3: The reduced-form public recognition function: by interval, restricting to number of visits questions within 4 predicted PR attendance



(a) Monotonic sample



(b) Coherent sample

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. Panel (a) excludes 46 participants with non-monotonic preferences for public recognition. Panel (b) excludes 15 participants with "incoherent" preferences for public recognition.

C.4 Interaction between demand for commitment and WTP for public recognition

To develop our measure of the WTP for motivation, we follow Carrera et al. (2019) and Allcott et al. (2020). 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. (2019) and Allcott et al. (2020) show that

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

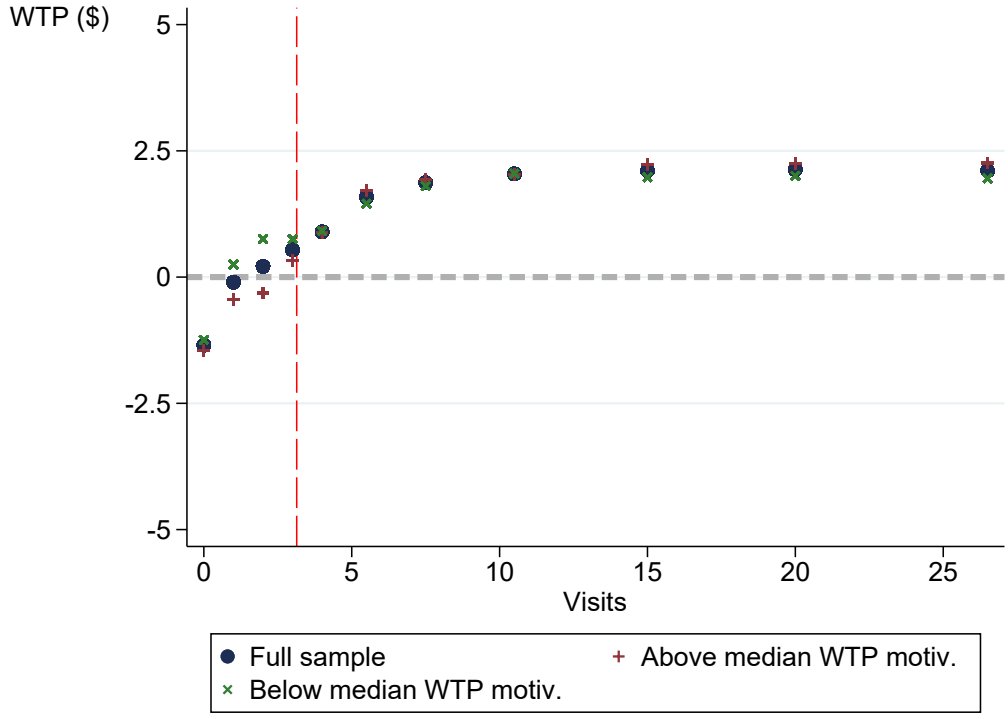
is a measure of individuals' perceived time-inconsistency. This measure equals 0 for individuals who perceive themselves to be time-consistent, is positive for individuals who would like to attend the YMCA more, and is negative for individuals who believe that they attend the YMCA too much. Below, we study whether this measure relates to participants' profile of WTP for public recognition. We present regression results in Table A7 and graphical results in Figure A4.

Table A7: WTP for public recognition by YMCA attendance: heterogeneity along demand for commitment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model	OLS	Tobit	OLS	Tobit	OLS	Tobit	OLS	Tobit
Dependent var.	WTP	WTP	WTP	WTP	WTP	WTP	WTP	WTP
N. visits	0.40***	0.68***	0.36***	0.62***	0.66***	1.05***	0.59***	0.92***
	(0.04)	(0.08)	(0.04)	(0.08)	(0.14)	(0.28)	(0.14)	(0.26)
N. visits sq.	-0.01***	-0.02***	-0.01***	-0.02***	-0.02***	-0.02**	-0.01***	-0.02**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)
WTP motivation	-0.03	-0.05	-0.03	-0.04	0.05	0.12	0.08	0.14
	(0.04)	(0.07)	(0.03)	(0.07)	(0.12)	(0.25)	(0.10)	(0.20)
N. visits × WTP motiv.	0.00	0.00	0.00	-0.00	-0.00	-0.01	-0.01	-0.01
	(0.00)	(0.01)	(0.00)	(0.01)	(0.02)	(0.03)	(0.01)	(0.03)
N. visits sq. × WTP motiv.	-0.00	-0.00	-0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	-1.00***	-2.14***	-0.64*	-1.45**	-3.52***	-6.07***	-2.81***	-4.77***
	(0.34)	(0.68)	(0.33)	(0.63)	(0.95)	(1.84)	(0.91)	(1.69)
Sample	Mon	Mon	Coh	Coh	Mon	Mon	Coh	Coh
Restriction	All	All	All	All	≤ 4	≤ 4	≤ 4	≤ 4
Observations	3729	3729	4070	4070	830	830	923	923
N. Subjects	339	339	370	370	339	339	370	370

Notes: This table reports regression estimates of quadratic models of willingness to pay for public recognition by YMCA attendance. Columns (1)-(4) use all 11 intervals of future attendance, while columns (5)-(8) 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. Columns (1), (2), (5), and (6) exclude 46 participants with non-monotonic preferences for public recognition. Columns (3), (4), (7), and (8) exclude 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 A4: WTP for public recognition by YMCA attendance: heterogeneity along demand for commitment (coherent sample)

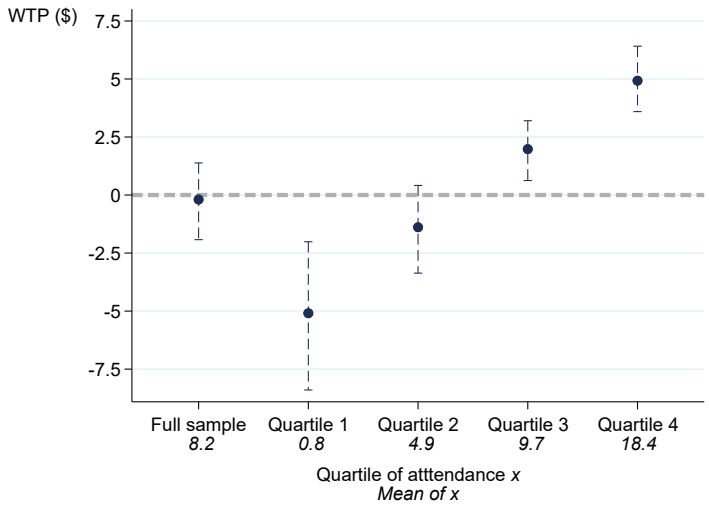


Notes: This figure plots the average WTP for public recognition by YMCA attendance. For intervals including more than one value of visits (e.g., "5 or 6 visits"), the WTP is plotted at the midpoint the interval. The figure separately reports the average WTP for the whole sample of coherent participants, 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.

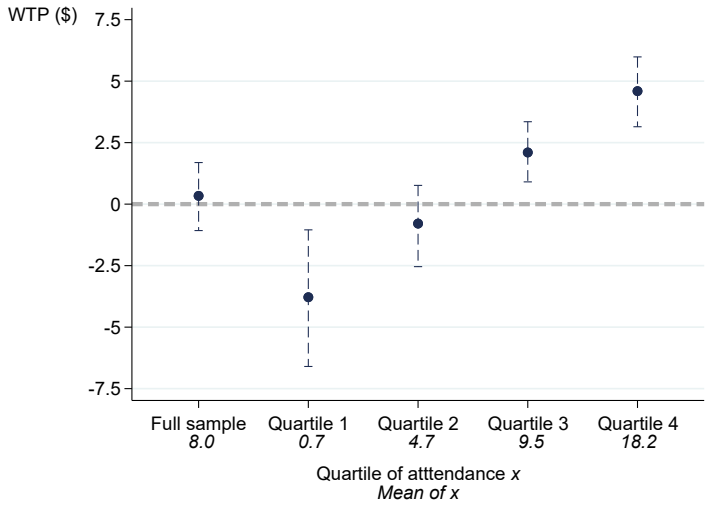
C.5 Additional results on realized payoffs from pride and shame

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.

Figure A5: The net effect of shame and pride in the YMCA experiment



(a) Monotonic sample



(b) Coherent sample

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. Panel (a) excludes 46 participants with non-monotonic preferences for public recognition. Panel (b) excludes 15 participants with "incoherent" preferences for public recognition. Bootstrapped percentile-based confidence intervals, sampled by participant with 1000 iterations, are displayed.

D Supplementary empirical results for charitable contribution experiments

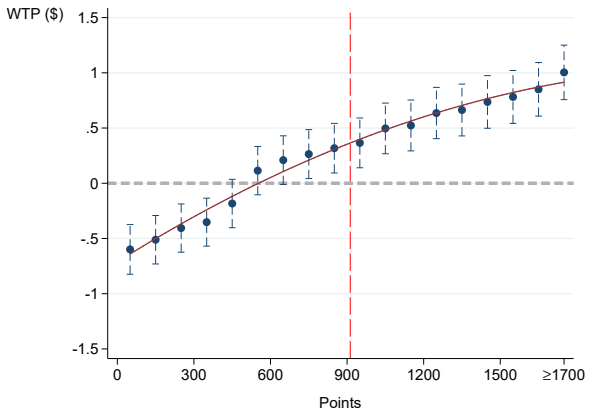
Table A8: The effect of public recognition on points scored, first round only

	(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

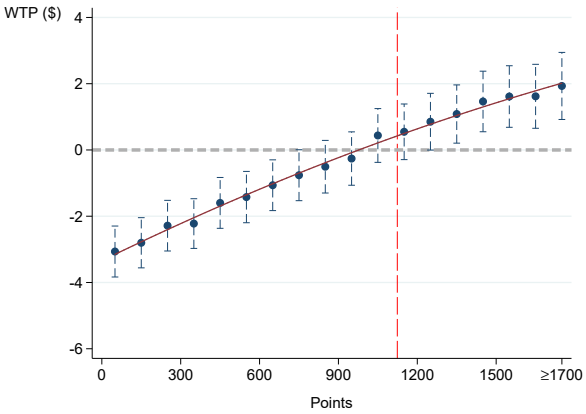
Notes: This table reports regression estimates of the effects of public recognition and financial incentives on points scored and is limited to observations from the first round randomly assigned to be completed by each participant. The control mean is the mean points scored in the Anonymous Effort Round. 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 A6: WTP for public recognition by effort in the charitable contribution experiments

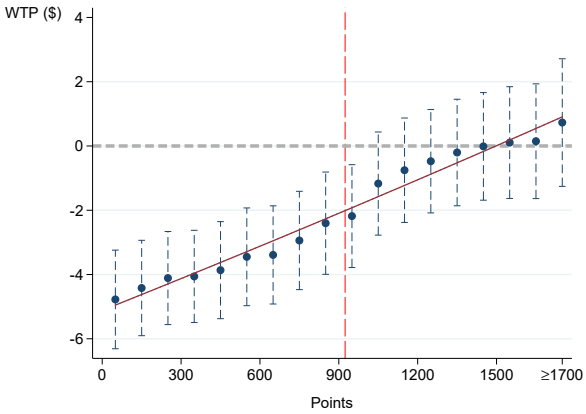
(a) Prolific



(b) Berkeley



(c) BU



Notes: These figures plot the average WTP for public recognition 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 A9: 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 A10: WTP for public recognition by effort in the charitable contribution experiments: heterogeneity in 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.

E Structural estimation details

E.1 Mapping to estimates from the reduced-form results

E.1.1 The reduced-form public recognition function

Motivated by the reduced-form results, we assume the reduced-form public recognition function takes the quadratic form $R(a) = r_0 + r_1 a + r_2 a^2$. For the YMCA sample, $R(a)$ corresponds to the quadratic Tobit regression of WTP on visits in column (2) of Table 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 6.

E.1.2 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, $\mathbb{E}[a|PR = 0]$ is directly observable as the average

YMCA attendance from the no PR treatment, or as the average performance in the Anonymous Effort Round.

E.2 Action-signaling model

In the action-signaling model, participants compare their action to $\rho^a \bar{a}_{pop}$,where $\bar{a}_{pop}$ is either the average attendance in the YOTA population, or the average performance in the Publicly-Shared Effort Round of the charitable contribution experiment. Public recognition utility has the form $\nu S^a(a - \rho^a \bar{a}_{pop}) = \gamma_1^a(a - \rho^a \bar{a}_{pop}) + \gamma_2^a(a - \rho^a \bar{a}_{pop})^2$. Total utility $U(a;\theta)$ is thus:

$$U(a;\theta) = \theta a - \frac{c}{2}a^2 + y + pa + \gamma_1^a(a - \rho^a \bar{a}_{pop}) + \gamma_2^a(a - \rho^a \bar{a}_{pop})^2 \tag{14}$$

E.2.1 Estimating the model parameters

Equation (14) has four unknown parameters γ_1^a, γ_2^a, ρ^a, and c. We estimate these parameters as functions of the reduced-form parameters r_0, r_1, r_2:

$$\gamma_1^a = \sqrt{r_1^2 - 4r_0 r_2} \tag{15}$$

$$\gamma_2^a = r_2 \tag{16}$$

$$\rho^a = \frac{\sqrt{r_1^2 - 4\gamma_2^a r_0} - r_1}{2\bar{a}_{pop} r_2} \tag{17}$$

$$c = \frac{r_1 + 2r_2\left(\mathbb{E}[a|PR = 0] + \bar{\tau}\right)}{\bar{\tau}} \tag{18}$$

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 using percentile-based bootstrap blocked at the individual level.

To see why equations (15)-(18) hold, we begin by regrouping the terms in $\nu S(a - \bar{a}_{pop})$:

$$\nu S(a - \rho^a \bar{a}_{pop}) = \left[\gamma_2^a(\rho^a \bar{a}_{pop})^2 - \gamma_1^a(\rho^a \bar{a}_{pop})\right] + \left[\gamma_1^a - 2\gamma_2^a(\rho^a \bar{a}_{pop})\right] a + \gamma_2^a \cdot a^2$$

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:

$$\gamma_2^a(\rho^a \bar{a}_{pop})^2 - \gamma_1^a(\rho^a \bar{a}_{pop}) = r_0 \tag{19}$$

$$\gamma_1^a - 2\gamma_2^a(\rho^a \bar{a}_{pop}) = r_1 \tag{20}$$

$$\gamma_2^a = r_2 \tag{21}$$

From this we immediately verify equation (16). Using $\gamma_2^a = r_2$ and the quadratic formula, we solve equation (19) for ρ^a in terms of γ_1^a :

$$\rho^a = \frac{\gamma_1^a - \sqrt{(\gamma_1^a)^2 + 4r_0 r_2}}{2r_2 \bar{a}_{pop}} \tag{22}$$

We verify equation (15) by substituting equation (22) and $\gamma_2^a = r_2$ into equation (20):

$$\gamma_1^a = \sqrt{r_1^2 - 4r_0 r_2} \tag{23}$$

By substituting equation (23) into equation (22), we verify equation (17):

$$\rho^a = \frac{\sqrt{r_1^2 - 4r_0 r_2} - r_1}{2\bar{a}_{pop} r_2}$$

To verify equation (18), we first note that, absent public recognition and financial incentives, $U(a;\theta) = \theta a - \frac{c}{2}a^2 + y$, and thus the optimal solution for each agent is $a = \theta/c$. From this, we have $\mathbb{E}[a|SR = 0] = \mathbb{E}[\theta/c]$.

We next use the first-order condition of $U(a;\theta)$, assuming there is public recognition, to solve for a:

$$\begin{aligned} 0 &= \theta - ca + \gamma_1^a + 2\gamma_2^a(a - \rho^a \bar{a}_{pop}) \\ a &= \frac{\theta/c + \gamma_1^a/c - 2\gamma_2^a \rho^a \bar{a}_{pop}/c}{1 - 2\gamma_2^a/c} \end{aligned}$$

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 $\mathbb{E}[\theta/c] = \mathbb{E}[a|PR = 0]$ and $\mathbb{E}[a|PR = 1] = \mathbb{E}[a|PR = 0] + \bar{\tau}$ into the expression above, and solve for c:

$$\begin{aligned} \mathbb{E}[a|PR = 0] + \bar{\tau} &= \frac{\mathbb{E}[a|PR = 0] + \gamma_1^a/c - 2\gamma_2^a \rho^a \bar{a}_{pop}/c}{1 - 2\gamma_2^a/c} \\ c &= \frac{\gamma_1^a - 2\gamma_2^a \rho^a \bar{a}_{pop} + 2\gamma_2^a \left(\mathbb{E}[a|PR = 0] + \bar{\tau}\right)}{\bar{\tau}} \end{aligned} \tag{24}$$

Finally, we substitute in for γ_1^a, γ_2^a, and ρ^a to verify equation (18):[34]

$$\begin{aligned} c &= \frac{\sqrt{r_1^2 - 4r_0r_2} - 2r_2 \cdot \frac{\sqrt{r_1^2-4r_0r_2}-r_1}{2r_2} + 2r_2\left(\mathbb{E}[a|PR=0] + \bar{\tau}\right)}{\bar{\tau}} \\ &= \frac{r_1 + 2r_2\left(\mathbb{E}[a|PR=0] + \bar{\tau}\right)}{\bar{\tau}} \end{aligned}$$

E.2.2 Estimating 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:

$$a(p) = \theta/c + p/c$$

The impact of financial incentives on attendance, $a(p) - a(0)$, is thus equal to p/c. We compute 95 percent confidence intervals for the predicted impact of financial incentives using percentile-based bootstrap blocked at the individual level.

E.2.3 Estimating 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. Here, the average attendance will increase until it reaches an equilibrium value $\bar{a}_{eq}$, and the reference point will become $\rho\bar{a}_{eq}$. We use a_0 to denote an individual's attendance absent public recognition and $\bar{a}^0_{pop}$ to denoted average population attendance absent public recognition. We also restrict $p = 0$. Here $\nu S(a|\bar{a}_{eq}) = \gamma_1^a(a - \rho^a\bar{a}_{eq}) + \gamma_2^a(a - \rho^a\bar{a}_{eq})^2$ and total utility takes the form:

$$U(a;\theta) = \theta a - \frac{c}{2}a^2 + y + \gamma_1^a(a - \rho^a\bar{a}_{eq}) + \gamma_2^a(a - \rho^a\bar{a}_{eq})^2$$

We first estimate $\bar{a}_{eq}$, and derive an expression for a, or the predicted attendance, as follows:

$$\bar{a}_{eq} = \frac{\bar{a}^0_{pop} + \gamma_1^a/c}{(1 - 2(1-\rho^a)\gamma_2^a/c)} \tag{25}$$

$$a = \frac{a_0 + \gamma_1^a/c - 2\gamma_2^a\rho\bar{a}^a_{eq}/c}{1 - 2\gamma_2^a/c} \tag{26}$$

We use these estimates to compute the change in average attendance and the net welfare effect from feeling pride and shame. We compute 95 percent confidence intervals using percentile-based

[34] While c can already be estimated from equation (24), it is useful to write c in terms of equation (18) to see that c is the same in the characteristic-signaling model. We estimate c using equation (18) rather than equation (24).

bootstrap blocked at the individual level.

To obtain equations (25) and (26), we again use the first order condition of total utility to solve for a:

$$a = \frac{\theta/c + \gamma_1^a/c - 2\gamma_2^a \rho^a \bar{a}_{eq}/c}{1 - 2\gamma_2^a/c}$$

We now substitute $a_0 = \theta/c$ into the above expression, which verifies equation (26):

$$a = \frac{a_0 + \gamma_1^a/c - 2\gamma_2^a \rho^a \bar{a}_{eq}/c}{1 - 2\gamma_2^a/c}$$

We next take the expectation of both sides, recalling that $\mathbb{E}[a] = \bar{a}_{eq}$, and that $\mathbb{E}[a_0] = \bar{a}^0_{pop}$ in equilibrium:

$$\bar{a}_{eq} = \frac{\bar{a}^0_{pop} + \gamma_1^a/c - 2\gamma_2^a \rho^a \bar{a}_{eq}/c}{1 - 2\gamma_2^a/c}$$

Finally, we verify equation (25) by solving the above expression for $\bar{a}_{eq}$:

$$\bar{a}_{eq} = \frac{\bar{a}^0_{pop} + \gamma_1^a/c}{(1 - 2(1-\rho^a)\gamma_2^a/c)}$$

E.3 Characteristic-signaling model

In the characteristic-signaling model, participants compare the signal of their type, $\mathbb{E}[\theta|a]$, to $\rho^\theta \bar{\theta}$, where $\bar{\theta} = c\bar{a}_{pop}$ and $\bar{a}_{pop}$ is either the average attendance in the YOTA population, or the average performance in the anonymous round for the charitable contribution experiment. Public recognition utility has the form $\nu S^\theta(\theta|\bar{\theta}) = \gamma_1^\theta(\mathbb{E}[\theta|a] - \rho^\theta\bar{\theta}) + \gamma_2^\theta(\mathbb{E}[\theta|a] - \rho^\theta\bar{\theta})^2$. Total utility $U(a;\theta)$ is thus:

$$U(a;\theta) = \theta a - \frac{c}{2}a^2 + y + pa + \gamma_1^\theta(\mathbb{E}[\theta|a] - \rho^\theta\bar{\theta}) + \gamma_2^\theta(\mathbb{E}[\theta|a] - \rho^\theta\bar{\theta})^2 \tag{27}$$

E.3.1 Signaling model microfoundations

We first provide a formal proof that $\nu S^\theta(\theta|\bar{\theta})$ can be mapped to a reduced-form function $R(a) = r_0 + r_1 a + r_2 a^2$ with $r_2 - \frac{c}{2} < 0$ and $R\left(a(\rho^\theta\bar{\theta})\right) = 0$.[35] Specifically, we show that S is quadratic, and derive the unique separating equilibrium.

To see that S is quadratic in θ, define $\phi(\theta) = \frac{\theta}{c-2r_2} + \frac{r_1}{c-2r_2}$. The public recognition function that leads to the quadratic reduced-form public recognition function $R(a)$ is thus:

[35]The condition $r_2 - \frac{c}{2} < 0$ ensures that S is quadratic, and that our solutions are well-defined.

$$\nu S(\theta - \rho^\theta \bar{\theta}) = r_0 + r_1 \cdot \phi(\theta) + r_2 \cdot \phi(\theta)^2$$

Since $\phi(\theta)$ is a linear function, S is quadratic in θ.

We now show that the unique equilibrium action function is given by $a(\theta) = \phi(\theta)$. To see this, note that if it were the case, then the reduced-form public recognition function would 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:

$$U(a;\theta) = \theta a - \frac{c}{2}a^2 + y + pa + r_0 + r_1 a + r_2 a^2 \tag{28}$$

We now verify that each type's optimal response is then $a(\theta) = \frac{\theta}{c-2r_2} + \frac{r_1}{c-2r_2}$. We do so by using the first order condition of equation (28) to solve for a:

$$\begin{aligned} 0 &= \theta - ca + r_1 + 2r_2 a \\ a &= \frac{\theta/c + r_1/c}{1 - 2r_2/c} \\ &= \phi(\theta) \end{aligned} \tag{29}$$

Finally, because the material utility function $\theta a - \frac{c}{2}a^2$ satisfies the single-crossing property, i.e., the derivative with respect to a, $\theta - ca$, is increasing in θ, the results of Mailath (1987) imply that this separating equilibrium must be a unique separating equilibrium.

E.3.2 Estimating the model parameters

Equation (27) has four unknown parameters γ_1^θ, γ_2^θ, ρ^θ, and c. As with the action-signaling model, we estimate these parameters as functions of the reduced-form parameters r_0, r_1, r_2:

$$\gamma_1^\theta = \frac{\sqrt{r_1^2 - 4r_0 r_2}}{c - 2r_2}. \tag{30}$$

$$\gamma_2^a = \frac{r_2}{(c-2r_2)^2} \tag{31}$$

$$\rho^\theta = \frac{\sqrt{r_1^2 - 4r_0 r_2} - r_1}{2\bar{a}_{pop} r_2} - \frac{\sqrt{r_1^2 - 4r_0 r_2}}{c\bar{a}_{pop}} \tag{32}$$

$$c = \frac{r_1 + 2r_2\left(\mathbb{E}[a|PR=0] + \bar{\tau}\right)}{\bar{\tau}} \tag{33}$$

As with the action-signaling model, the parameters are highly nonlinear functions of these empirical moments. We thus compute 95 percent confidence intervals without relying on asymptotic normality approximations using percentile-based bootstrap blocked at the individual level.

To see why equations (30)-(33) hold, we first note from equation (29) that the action of type

$\rho^\theta\bar{\theta}$ is given by :

$$a(\rho^\theta\bar{\theta}) = \frac{\rho^\theta\bar{\theta}/c + r_1/c}{1 - 2r_2/c}$$

Using $\bar{\theta}/c = \bar{a}_{pop}$, we rewrite this as:

$$a(\rho^\theta\bar{\theta}) = \frac{\rho^\theta\bar{a}_{pop} + r_1/c}{1 - 2r_2/c}$$

We next substitute the above expression into $R\left(a(\rho^\theta\bar{\theta})\right) = 0$:

$$0 = r_0 + r_1\frac{\rho^\theta\bar{a}_{pop} + r_1/c}{1 - 2r_2/c} + r_2\left(\frac{\rho^\theta\bar{a}_{pop} + r_1/c}{1 - 2r_2/c}\right)^2$$

We next solve this equation for ρ^θ via the quadratic formula, and verify equation (32):

$$\rho^\theta = \frac{\sqrt{r_1^2 - 4r_0r_2} - r_1}{2\bar{a}_{pop}r_2} - \frac{\sqrt{r_1^2 - 4r_0r_2}}{c\bar{a}_{pop}}$$

To verify equations (30) and (31), we use equation (29) to write $R(a) = r_0 + r_1a + r_2a^2$ as:

$$R\left(a(\theta)\right) = r_0 + r_1\frac{\theta/c + r_1/c}{1 - 2r_2/c} + r_2\left[\frac{\theta/c + r_1/c}{1 - 2r_2/c}\right]^2$$

The above expression is algebraically equivalent to the following:

$$\begin{aligned} R\left(a(\theta)\right) = r_0 + r_1\frac{\rho^\theta\bar{\theta}/c + r_1/c}{1 - 2r_2/c} + r_2\left(\frac{\rho^\theta\bar{\theta}/c + r_1/c}{1 - 2r_2/c}\right)^2 \\ + \frac{r_1 + \frac{2r_1r_2 + 2r_2\rho^\theta\bar{\theta}}{c - 2r_2}}{c - 2r_2}(\theta - \rho^\theta\bar{\theta}) + \frac{r_2}{(c - 2r_2)^2}(\theta - \rho^\theta\bar{\theta})^2 \end{aligned}$$

The first three terms in the equation above sum to $R\left(a(\rho^\theta\bar{\theta})\right) = 0$. Using equation (32), we simplify the coefficient on $(\theta - \rho^\theta\bar{\theta})$:

$$R\left(a(\theta)\right) = \frac{\sqrt{r_1^2 - 4r_0r_2}}{c - 2r_2}(\theta - \rho^\theta\bar{\theta}) + \frac{r_2}{(c - 2r_2)^2}(\theta - \rho^\theta\bar{\theta})^2$$

Because we have a unique separating equilibrium, each agent's action reveals their true type. Thus $\mathbb{E}[\theta|a] = \theta$. Using this and the above expression, we match γ_1^θ and γ_2^θ to the reduced-form public recognition function via the following equations:

$$\gamma_1^\theta = \frac{\sqrt{r_1^2 - 4r_0 r_2}}{c - 2r_2}$$
$$\gamma_2^\theta = \frac{r_2}{(c - 2r_2)^2}$$

To verify equation (33), we next take the expectation of both sides of equation (29), recalling that we are in the case where $PR = 1$:

$$\mathbb{E}[a|PR = 1] = \frac{\mathbb{E}[\theta/c] + r_1/c}{1 - 2r_2/c}$$

We substitute $\mathbb{E}[\theta/c] = \mathbb{E}[a|PR = 0]$ and $\mathbb{E}[a|PR = 1] = \mathbb{E}[a|PR = 0] + \bar{\tau}$ into the expression above, and solve for c:

$$c = \frac{r_1 + 2r_2 \left(\mathbb{E}[a|PR = 0] + \bar{\tau}\right)}{\bar{\tau}}$$

E.3.3 Estimating the predicted impact of financial incentives

Since (i) c is the same here as in the action model, and (ii) the derivation for $a(p) - a(0) = p/c$ did not depend on the public recognition function, the predicted impact of financial incentives in the characteristic-signaling model is the same as in the action-signaling model.

E.3.4 Estimating 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 (29) to predict equilibrium attendance $\bar{a}_{eq}$:

$$a = \frac{\theta/c + r_1/c}{1 - 2r_2/c}$$
$$\bar{a}_{eq} = \frac{\bar{a}^0_{pop} + r_1/c}{1 - 2r_2/c}$$

To estimate the individual attendance in the counterfactual, we substitute $\theta/c = a_0$ into the optimal attendance rule:

$$a = \frac{a_0 + r_1/c}{1 - 2r_2/c}$$

We use these estimates to compute the change in average attendance and the net welfare effect from feeling pride and shame. We compute confidence intervals using percentile-based bootstrap blocked at the individual level.

E.4 Incorporating heterogeneity and uncertainty

E.4.1 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 recognition function $R(a) = r_0 + r_1 a + r_2 a^2$ is

$$a = \frac{(\theta - \xi)/c}{1 - 2r_2/c} + \frac{r_1/c}{1 - 2r_2/c} \tag{34}$$

and thus

$$\mathbb{E}[a|\theta] = \frac{\theta/c}{1 - 2r_2/c} + \frac{r_1/c}{1 - 2r_2/c} \tag{35}$$

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.

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 (34), 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 $\nu S(\theta - \rho^\theta \bar{\theta}) = r_0 + r_1 \cdot \phi(\theta) + r_2 \cdot \phi(\theta)^2$, where $\phi(\theta) = \frac{\theta}{c-2r_2} + \frac{r_1}{c-2r_2}$.

E.4.2 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 tedious they'll find the Click for Charity task, or how much time they will have to attend the YMCA. Plainly, this does not affect our analysis in any way 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.