American Economic Review 2015, 105(8): 2501–2538
http://dx.doi.org/10.1257/aer.20131564

Evaluating Behaviorally Motivated Policy: Experimental Evidence from the Lightbulb Market†

By Hunt Allcott and Dmitry Taubinsky*

Imperfect information and inattention to energy costs are important potential motivations for energy efficiency standards and subsidies. We evaluate these motivations in the lightbulb market using a theoretical model and two randomized experiments. We derive welfare effects as functions of reduced-form sufficient statistics capturing economic and psychological parameters, which we estimate using a novel within-subject information disclosure experiment. The main results suggest that moderate subsidies for energy-efficient lightbulbs may increase welfare, but informational and attentional biases alone do not justify a ban on incandescent lightbulbs. Our results and techniques generate broader methodological insights into welfare analysis with misoptimizing consumers. (*JEL* D12, D83, H21, H31, L67, Q41, Q48)

A fundamental assumption in traditional policy analysis is that people's choices identify their true preferences. In practice, however, many policies are at least partially predicated on the idea that consumers' choices may not maximize their own welfare. Examples include consumer financial protection, taxes and bans on drugs, alcohol, cigarettes, and unhealthy foods, and subsidies and mandates for energy-efficient products. To evaluate such policies, it is necessary to extend traditional public finance analysis to allow for the possibility of consumer mistakes and to design empirical strategies that identify the necessary economic and psychological parameters. This paper carries this out in the context of energy efficiency policy.

*Allcott: Department of Economics, New York University, 19 West 4th Street, New York, NY 10012, and NBER (e-mail: hunt.allcott@nyu.edu); Taubinsky: Harvard University, Littauer M-35, 1805 Cambridge Street, Cambridge, MA 02138, and UC-Berkeley (e-mail: taubinsk@fas.harvard.edu). This paper previously circulated under the title "The Lightbulb Paradox: Evidence from Two Randomized Experiments." We are grateful to Raj Chetty, Lucas Davis, Stefano DellaVigna, Marc Kaufmann, Mushfiq Mubarak, Sendhil Mullainathan, Emmanuel Saez, Josh Schwartzstein, and other colleagues, as well as seminar audiences at the ASSA Annual Meeting, Bonn, Boston University, the Behavioral Economics Annual Meeting, Berkeley, CESifo Behavioral Economics Conference, Cologne, Dartmouth, the European Summer Symposium for Economic Theory, Frankfurt, Harvard, the NBER Public Economics Meetings, Regensburg, Resources for the Future, Stanford, the Stanford Institute for Theoretical Economics, UCLA, the University of California Energy Institute, the University of Chicago, and Wharton for constructive feedback. We thank our research assistants—Jeremiah Hair, Nina Yang, and Tiffany Yee—as well as management at the partner company, for their work on the in-store experiment. Thanks to Stefan Subias, Benjamin DiPaola, and Poom Nukulkij at GfK for their work on the TESS experiment. We are grateful to the National Science Foundation and the Sloan Foundation for financial support. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.
† Go to http://dx.doi.org/10.1257/aer.20131564 to visit the article page for additional materials and author disclosure statements.

Energy efficiency subsidies and standards are important examples of policies partially motivated by addressing consumer bias. It has long been suggested that consumers may be imperfectly informed about or inattentive to energy costs when they buy energy-using durables such as cars, air conditioners, and lightbulbs.[1] This suggestion is supported by recent empirical evidence that people are inattentive to other ancillary product costs such as sales taxes (Chetty, Looney, and Kroft 2009), shipping and handling charges (Hossain and Morgan 2006), and out-of-pocket insurance costs (Abaluck and Gruber 2011). Because energy is costly—American households spent $325 billion on gasoline and another $245 billion on electricity, natural gas, and heating oil in 2011 (Bureau of Labor Statistics (BLS) 2013)—even small inefficiencies can aggregate to substantial losses in the absence of corrective policies.

This paper focuses on the lightbulb market, a particularly compelling case study of what Jaffe and Stavins (1994) call the "energy paradox": the low adoption of energy-efficient technologies despite apparently large cost savings. Compared to standard incandescents, compact fluorescent lightbulbs (CFLs) last much longer and use four times less electricity, so a 60-watt equivalent CFL saves about $5 per year on average.[2] In 2010, however, only 28 percent of residential sockets that could hold CFLs actually had them (US Department of Energy (DOE) 2010), and in that year, using incandescents instead of CFLs cost US households a total of $15 billion.[3] Of course, CFLs and incandescents are far from perfect substitutes, and many consumers dislike CFLs for various reasons. Is the CFL's low market share simply an expression of well-informed preferences, or are consumers unaware of or inattentive to how much money they could save?

This question matters for policy. Electric utilities in the United States spent $252 million promoting CFLs in 2010, largely through subsidies (DOE 2010). Furthermore, the Energy Independence and Security Act of 2007 sets minimum efficiency standards that ban traditional incandescent lightbulbs, and Argentina, Australia, Brazil, Canada, China, Cuba, the European Union, Israel, Malaysia, Russia, and Switzerland have similar bans. Although externalities and other market failures also play a role, many advocates argue that the incandescent lightbulb ban acts in consumers' best interests by preventing them from buying a product with large "shrouded" costs. A rancorous debate has evolved in a void of relevant evidence, despite a simple testable hypothesis: fully informed and attentive consumers would have higher willingness-to-pay for a CFL.

We use two randomized experiments to answer two questions. (i) How much does information provision affect demand for CFLs? (ii) If powerful information provision is costly or infeasible, does a CFL subsidy or a ban on incandescents increase welfare as a second-best solution to imperfect information and inattention? The first

[1] Allcott (forthcoming) includes an extended series of quotes from policymakers and policy analyses that document this argument. See also Anderson and Claxton (1982); Blumstein et al. (1980); Jaffe and Stavins (1994); Sanstad and Howarth (1994); Gillingham and Palmer (2013); and many others.

[2] The $5 estimate reflects $4.50 in electricity savings, based on an average usage of 1,000 hours per year (DOE 2010) and a national average electricity price is $0.10 per kilowatt-hour (DOE 2014), plus $0.50 in bulb replacement savings at typical prices. Throughout the paper, we assume that incandescents and CFLs last an average of 1,000 and 8,000 hours, respectively. (To receive the Energy Star rating, a CFL model must last a median of 8,000 hours in official tests.)

[3] This $15 billion estimate is equal to 5.8 billion residential sockets (DOE 2012), times the 80 percent of sockets that can accommodate CFLs (DOE 2010) minus the actual "socket share" of 28 percent (DOE 2010), times $5 per socket per year.

is a positive question that can be answered by estimating the effects of information, with no additional structure or assumptions. To answer the second question, we use an optimal policy framework to derive "sufficient statistic" formulas for welfare effects and carry out an experiment specifically designed to estimate those sufficient statistics.

Our optimal policy framework follows Allcott, Mullainathan, and Taubinsky (2014); Baicker, Mullainathan, and Schwartzstein (2015); DellaVigna (2009); Mullainathan, Schwartzstein, and Congdon (2012); and others,[4] and provides a simple extension of classical optimal tax formulas. Just as Diamond (1973) shows that the optimal externality tax equals the average marginal externality, the optimal internality tax (or subsidy) equals what we call the *average marginal bias*—the average valuation mistake of consumers whose choices are marginal to the policy change. The net welfare effect of a ban on incandescents is the loss of perceived surplus for consumers who had purchased the incandescent plus any gain from internality reduction. Two functions are sufficient statistics to evaluate a subsidy or ban: the market demand curve and the average marginal bias at each point on that demand curve.

Our first experiment is an "artefactual field experiment" (Harrison and List 2004) using a nationally representative online platform called Time-Sharing Experiments for the Social Sciences (TESS). Two specific features allow it to identify the two sufficient statistics. First, it is a within-subject design: consumers make baseline choices between CFLs and incandescents at different relative prices using a multiple price list format, then there is a randomly assigned information treatment, and then consumers make endline choices using another multiple price list. We thus observe the baseline market demand curve and the conditional average treatment effect (CATE) on willingness-to-pay (WTP) at each point on that curve. Second, the information treatment was specifically designed to provide only hard information, ensure comprehension, and minimize demand effects and other potential confounds. It is thus not unreasonable to assume that our information treatment is what we call a *pure nudge*: it informs all previously uninformed consumers and draws full attention to energy costs, with no other effects. Under this pure nudge assumption, the CATEs on WTP from our information treatments equal the average marginal bias from imperfect information and inattention. Although this assumption is also made in Chetty, Looney, and Kroft (2009) and other work, it is perhaps the greatest weakness of this approach, and we view it only as a useful approximation. We evaluate it throughout the paper and provide robustness checks under plausible alternative assumptions.

In the TESS experiment, information increases the CFL's market share at market prices by about 12 percentage points. The treatment effects on willingness-to-pay for a 60-watt-equivalent CFL differ across points on the baseline demand curve, with an average treatment effect of \$2.30. While this effect is small compared to the average rated lifetime cost savings from a CFL (about \$40), it is larger compared to the market prices of our lightbulb packages (about \$4) or compared to the baseline average WTP for the CFL relative to the incandescent (\$2.90). Under the pure

[4] Also closely related is Spinnewijn (2014), as well as earlier work by O'Donoghue and Rabin (2006) and Gruber and Kőszegi (2004) on optimal sin taxes with present-biased preferences.

nudge assumption that the effects of information measure consumer bias, our optimal policy framework suggests that the optimal CFL subsidy is approximately \$3. This is slightly larger than typical CFL subsidies offered by many electric utilities in the United States.

However, a large group of consumers purchase incandescents at baseline and are still willing to pay substantially more for incandescents after the informational intervention. Banning incandescents imposes welfare losses on this population that outweigh the gains to uninformed or inattentive consumers. This implies that in our model, imperfect information and inattention alone do not justify a ban on traditional incandescents. This qualitative conclusion holds in most, although not all, of our welfare analyses under alternative assumptions. For simplicity, our quantitative analysis assumes zero distortions from other factors such as uninternalized environmental externalities. We discuss these issues further in Section I, and we provide formulas that can easily extend the empirical welfare analysis to include such additional distortions.

Our second experiment is a natural field experiment with a 2-by-2 design that randomly assigned subsidies and information provision across shoppers at a large home improvement retailer. It is a useful complement to the TESS experiment: while this in-store setting imposes design constraints that limit the parameters that can be identified, the results provide evidence from a more realistic shopping environment. In this experiment, information did not statistically significantly affect CFL market share, and we bound the effect at around 5 percentage points with 90 percent confidence. We discuss factors that could explain why the in-store market share effects were smaller than the TESS effects, including that there was additional information available to the control group in stores or that the more complex in-store environment attenuated effects on the treatment group. While we show formally that market share effects are not informative about the average marginal bias, the smaller in-store market share effects do suggest smaller bias. This would strengthen the TESS result that imperfect information and inattention do not justify the incandescent lightbulb ban. The two experiments are also qualitatively consistent in showing that meaningful shares of consumers still purchase incandescents even after substantial effort to inform and draw attention to energy costs.

The paper makes three main contributions. In answer to our first research question, we are the first (to our knowledge) to use real-stakes randomized experiments to study how energy cost information affects choices of energy-using durables.[5] While there has been extensive work on other aspects of information and energy demand, including Jessoe and Rapson (2014) and many others, only experiments like ours that provide durable good energy cost information are directly relevant to

[5] There are some related studies that differ from our experiments on one or more dimensions. Kallbekken, Sælen, and Hermansen (2013) study energy information disclosure in Norway using a nonrandom control group. Anderson and Claxton (1982) study energy information labels but has only 18 units of randomization. Newell and Siikamäki (2013); Ward et al. (2011); and many other papers study effects of information on hypothetical choices. Deutsch (2010a,b) study information disclosure with online shoppers, measuring what products they click on and what products they put in online shopping carts, but he does not observe actual purchases. Houde (2012) uses quasi-experimental variation with a structural demand model to estimate how the Energy Star label affects consumer welfare, while Herberich, List, and Price (2011) studies how prices and social norm information affect CFL purchases.

the important policy debates around multibillion dollar subsidies, standards, and information disclosure for energy-using durables.

In answer to our second research question, we provide a theoretically grounded empirical analysis of the "behavioral" motivation for lighting energy efficiency standards. This is especially important because while consumer misoptimization has become an important rationale for energy efficiency policy, there is confusion and disagreement about how to formalize and test this rationale. Our analysis also advances a broader empirical literature on whether durable good buyers "undervalue" energy costs relative to purchase prices.[6] In this literature, our approach is innovative in that we test for undervaluation using randomized experiments instead of using observational data to compare how markets respond to prices versus energy costs.

Our third contribution is methodological: as the more general framework in online Appendix D.C clarifies, the average marginal bias is a key statistic not just for energy policy, but also for a broader set of questions in behavioral public finance. Our TESS experimental design is the first (to our knowledge) to directly measure the average marginal bias. Existing empirical analyses, including the influential work of Chetty, Looney, and Kroft (2009), instead estimate a statistic that we call the *equivalent price metric* (EPM), which equals the average marginal bias only under special homogeneity and linearity assumptions. We find that the EPM is a poor approximation in our data—at market prices, for example, the EPM is only about one-half as large as the average marginal bias. The fact that the EPM and other commonly estimated statistics differ meaningfully from the average marginal bias implies that most existing empirical estimates of consumer misoptimization are not applicable to welfare analysis.

Section I gives more background on lightbulbs and related policies. Section II lays out our theoretical framework and defines the sufficient statistics that must be estimated. Section III presents the TESS experiment, and Section IV carries out welfare evaluation. Section V presents the in-store experiment, and Section VI concludes.

I. Background: Reasons for Subsidies and Standards

Why subsidize CFLs or ban[7] traditional incandescents? One potential reason to subsidize or mandate energy efficiency would be if retail energy prices were below social marginal cost and could not be raised due to political constraints. But while

[6]This literature includes Allcott (2013); Allcott and Wozny (2014); Busse, Knittel, and Zettelmeyer (2013); Dubin and McFadden (1984); Goldberg (1998); Hassett and Metcalf (1995); Hausman (1979); Sallee, West, and Fan (2015); and many others. There are also several theoretical and simulation analyses of energy taxes, energy efficiency standards, or subsidies for energy-efficient goods when consumers misoptimize, including Allcott, Mullainathan, and Taubinsky (2014); Heutel (forthcoming); Parry, Fischer, and Harrington (2007); and Parry, Evans, and Oates (2010).

[7]The US lighting efficiency standards do not ban all incandescents. Instead, they set a maximum energy use per unit of light output. Along with CFLs, light-emitting diodes (LEDs) and high-efficiency halogen bulbs also comply. We focus on the choice between CFLs and incandescents because these are *by far* the most important current technologies. In 2012, about 1.5 billion incandescents and 300 million CFLs were purchased, compared to only 23 million LEDs (Energy Star 2013). While our quantitative welfare calculations would certainly change in the future if LEDs become a relevant part of the choice set, the basic questions about imperfect information and inattention from CFLs might also apply to LEDs, given that LEDs also have high purchase prices, long lifetimes, and large energy cost savings relative to both incandescents and CFLs.

the lack of a carbon price artificially depresses electricity prices, two other distortions imply that electricity prices could actually be *above* social marginal cost. First, retailers typically include much of fixed distribution costs in marginal prices, as Borenstein and Davis (2012) and Davis and Muehlegger (2010) show for natural gas. Second, most residential customers are charged time-invariant prices instead of real-time market prices, which are lower at night and higher during the day. Thus, if lightbulbs are relatively more likely to be used at night, they use underpriced electricity. This suggests that if the primary distortion is mispriced residential electricity, it could actually be optimal to *subsidize* incandescents.[8]

Alternatively, subsidies for new or emerging products might help correct for uninternalized spillovers from research and development or consumer learning. However, the CFL is an established technology, and the vast majority of consumers already have experience with it: 70 percent of consumers report having at least one CFL in their home, compared to 80 percent who report having at least one incandescent (Sylvania 2012).

Asymmetric information in real estate markets could also justify subsidies and standards. For example, prospective renters cannot costlessly observe energy efficiency, which reduces the incentive of landlords to invest in energy-efficient capital stock (Davis 2012; Gillingham, Harding, and Rapson 2012). Similarly, renters or owners who expect to move before the end of the investment life have reduced incentive to invest. Davis (2012) estimates that renters in the United States are 5 percent less likely to use CFLs, but this would explain only a small fraction of the CFL's smaller market share given that only one-fourth of US households are renters.

A final set of inefficiencies are "internalities," or choices that don't maximize the consumer's own welfare. Informational and attentional internalities play an important role in the policy debate. For example, the Regulatory Impact Statement for Australia's ban on energy-inefficient lightbulbs argues:

> [*Incandescent lightbulbs*] *continue to sell remarkably well because, if their energy costs are ignored, they appear cheap … There are significant information failures and split incentive problems in the market for energy efficient lamps. Energy bills are aggregated and periodic and therefore do not provide immediate feedback on the effectiveness of individual energy saving investments. Consumers must therefore gather information and perform a reasonably sophisticated calculation to compare the life-cycle costs of* [*incandescents*] *and CFLs. But many lack the skills. For others, the amounts saved are too small to justify the effort…* (DEWHA 2008, p. vii)

The official US government Regulatory Impact Analysis (RIA) of the Energy Independence and Security Act of 2007 (EISA) argues that after accounting for incremental production costs, the lighting efficiency standards will save consumers a net present value of \$27–\$64 billion over 30 years (DOE 2009). Of course, for

[8] California is a particularly stark example. Regulations encouraging low-carbon electricity generation mean that the carbon content of electricity consumed there is extremely low relative to other states, so the downward distortion to electricity prices from the lack of a carbon tax is particularly small. Meanwhile, residential electricity tariffs with sharply increasing block prices distort marginal prices upward. Despite the fact that these two forces significantly weaken or reverse the argument that underpriced electricity justifies energy efficiency policies, California implemented the federal lighting efficiency standards early.

market forces to not generate these private benefits independently, there must be some market inefficiency—or some additional utility cost that the RIA has ignored.[9] Private benefits are central to many energy efficiency regulations: for example, the RIA for the 2012–2016 Corporate Average Fuel Economy (CAFE) standards projects $111 billion in net private savings from inducing consumers to buy higher-fuel economy vehicles (NHTSA 2010). While the EISA documents do not take a stand on what market inefficiency makes possible the apparent private savings, the CAFE final rule states that "the problem is that consumers appear not to purchase products that are in their economic self-interest," and proposes several explanations, including that consumers "lack information" and that "the benefits of energy-efficient vehicles may not be sufficiently salient" (EPA and DOT 2010, p. 25511).[10]

An overview article by Gayer (2011, p. 17) summarizes the argument. "Private net benefits represent the bulk of the benefits of the energy-efficiency standards," according to the official cost-benefit analyses (CBAs). "Energy efficiency regulations and fuel economy regulations are therefore justified by such CBAs only by presuming that consumers are unable to make market decisions that yield personal savings, that the regulator is able to identify these consumer mistakes, and that the regulator should correct economic harm that people do to themselves."

In the absence of our results, this policy argument could be quite plausible, as empirical estimates from other contexts suggest large attentional biases. A CFL saves $36 (undiscounted) in energy costs over its expected life relative to an incandescent. This dwarfs the typical relative price difference, suggesting that a consumer who is inattentive to these savings would be much more likely to buy a CFL. If 20 percent of consumers don't think about energy costs, which is a seemingly conservative estimate relative to estimates for other energy-using durables, sales taxes, and health insurance plans,[11] the average bias would be around $7 and our informational interventions could have massive impacts on demand.

In summary, while there are other market failures that could justify lightbulb subsidies and standards, we focus on imperfect information and inattention because results from other literatures suggested that these two distortions could be large, while other market failures appear to be less relevant in this context.

[9] The EISA RIA is not the only analysis to focus on private cost savings from the lighting efficiency standards: the Environmental Protection Agency (2011) non-technical summary and advocates such as the NRDC (2011) do so as well. Opponents focus on the internality argument, suggesting that the ban is "over-reaching government intrusion into our lives" (Formisano 2008). US senator Rand Paul says that he supports energy conservation but objects to the idea that Department of Energy regulators "know what's best for me" (ABC News 2011).

[10] In justifying the lighting energy efficiency standards, private net benefits are considerably more important than the carbon externality reduction, which the EISA RIA values at no more than $16 billion over 30 years. The importance of private benefits relative to externalities is not unique to the lighting efficiency standards: summing across all durable good energy efficiency standards in the EISA, the net private cost savings outweigh the value of carbon externality reductions by 34–194 percent. In the CAFE standard RIA (NHTSA 2010), net private benefits outweigh all externality benefits (from carbon, local air pollution, safety, noise, and congestion) by a factor of 5.7.

[11] Forty percent of Americans report that they "did not think about fuel costs at all" when buying their most recent vehicle (Allcott 2011). In their two empirical studies, Chetty, Looney, and Kroft (2009) estimate that consumers are only 35 percent and 6 percent as attentive to sales taxes as they are to product prices. Abaluck and Gruber (2011) find that consumers are five times more responsive to insurance plan premiums than to out-of-pocket costs.

II. Theoretical Framework

A. *Consumer Choice and Optimal Policy*

Consumer Choice.—We consider consumers that make one of two choices, labeled E and I. In our empirical application, E represents the purchase of an energy-efficient good (the CFL), while I is an energy-inefficient good (the incandescent). We let p_j denote the price of good $j \in \{E, I\}$ and let $p = p_E - p_I$ denote the relative price of E.

We define v_j as the consumer's true utility from consuming product j and let $v = v_E - v_I$ denote the relative true utility from E. In our empirical application, v can depend on any and all differences between CFLs and incandescents, such as electricity costs, lifetimes, mercury content, brightness, and "warm glow" utility from reduced environmental impact.

A consumer's utility from purchasing product j at price p_j is $v_j + (Z - p_j)$, where Z is the consumer's budget and $Z - p_j$ is utility from consuming the numeraire good. A fully optimizing consumer thus chooses E if and only if $v > p$. A misoptimizing consumer chooses E if and only if $v - b > p$, where b is a bias that affects choice but not true utility.

We let F denote the cumulative density function (CDF) of v, let $G(b\,|\,v)$ denote the CDF of b conditional on a true valuation v, and let H denote the CDF of perceived valuations $\hat{v} = v - b$. We let $D_B(p) = 1 - H(p)$ and $D_N(p) = 1 - F(p)$, respectively, denote the demand curves corresponding to consumers' actual choices and to the choices consumers would make if they were unbiased. We assume that D_B and D_N are both smooth and strictly decreasing.

Our utility function is quasilinear, which is reasonable for purchases such as lightbulbs where p is small relative to Z. This simplifies the results, although the analysis can easily be generalized. We also assume that there are no externalities, although we will show in Section IIB that the welfare formulas are easily generalized to incorporate externalities.

Optimal Policy.—The policymaker seeks to maximize social welfare and can set two policies: a subsidy of amount s for good E and a ban on either choice. We assume that the policymaker maintains a balanced budget through lump-sum recycling (taxes or transfers), and we let $Z(s)$ denote consumers' after-tax income when the policymaker sets a subsidy s.[12] Under the lump-sum recycling and quasilinear utility assumptions, the subsidy does not distort other consumption and is thus purely corrective. Because of lump-sum recycling and no outside option (consumers choose either E or I), a subsidy for E is equivalent to a tax on I, and a ban on one choice is equivalent to a mandate for the other.

Goods E and I are produced in a competitive economy at constant marginal costs c_j, with relative cost $c = c_E - c_I$. Good E's relative price after subsidy s is $p = c - s$.

[12] Formally, to fund a subsidy s, the government must raise revenue $R(s) = \int 1_{v-b \geq c-s}\, s\, dF(v)\, dG(b\,|\,v)$. Thus consumers' after tax income is given by $Z(s) = Z - R(s)$.

Generalizing the classic analysis of Harberger (1964a, b), we now derive a simple formula for the welfare impact of a subsidy. $W(s) = Z(s) + v_I - p_I + \int_{v-b\geq p} (v - p)\, dF\, dG$ denotes social welfare as a function of the post-subsidy price $p = c - s$.

PROPOSITION 1:

$$W'(s) = (s - B(p))D_B'(p) \tag{1}$$

and

$$W(s + \Delta s) - W(s) \approx \overbrace{s\Delta s D_B'(p) + \frac{(\Delta s)^2}{2} D_B'(p)}^{\text{“Harberger distortion”}} + \overbrace{\underbrace{-\Delta s D_B'(p)}_{\text{Change in demand}} \underbrace{\big(E_H[B(x)\,|\,p - \Delta s \leq x \leq p]\big)}_{\text{Average marginal bias}}}^{\text{“Internality reduction”}}, \tag{2}$$

where $B(p) = E_G(b\,|\,v - b = p)$ *is the average marginal bias at price* $p = c - s$.[13]

Equation (2) follows from equation (1) by considering nonmarginal changes in the subsidy.[14] Both equations show that a subsidy has two effects. First, a subsidy distorts the market away from consumers' perceived private optimum. That is, it induces consumers to buy goods that they think they value at less than production cost. We call this the "Harberger distortion," and when the average bias of consumers marginal to the subsidy change is zero, equation (2) reduces to the standard Harberger formula. Second, when the average bias of marginal consumers is positive, the subsidy reduces internalities. That is, it induces consumers to buy products that are more valuable to them than they realize.

In our framework with no outside option, lump-sum taxation, and quasilinear utility, a ban on good *I* is equivalent to an infinite subsidy, so the welfare impact of a ban is simply $\int_0^\infty W'(s)ds = \int_0^\infty (s - B(c - s))D_B'(c - s)\, ds$. This can be approximated empirically by applying equation (2) over increasing subsidy levels.

At the social optimum, equation (1) must equal zero, which leads to a simple characterization of the optimal subsidy:

COROLLARY 1: *If* s^* *is an optimal subsidy, then* $s^* = B(c - s^*)$.

The corollary is analogous to a result obtained by Allcott, Mullainathan, and Taubinsky (2014) in a richer framework in which consumers both choose a product (e.g., a car) and then choose how much to utilize it (e.g., miles driven). Corollary 1

[13] As usual, we subscript the expectation operator with the distribution over which the expectation is taken.

[14] Under the additional assumption that $B''(p) \approx 0$, we can also derive the additional approximation $W(s + \Delta s) - W(s) = \Delta s(s - B(p))D_B'(p) + \frac{\Delta s^2}{2}(1 + B'(p))D_B'(p)$. See online Appendix D.A for details.

extends the sin tax logic of O'Donoghue and Rabin (2006) to the case of general biases[15] and also extends Mullainathan, Schwartzstein, and Congdon (2012) to the case of arbitrarily heterogeneous biases.

In this analysis, there is a close analogy between internalities and externalities. Analogous to Proposition 1, the welfare impact of an externality tax can similarly be decomposed into (i) the negative impact of distorting the market away from the private optimum, and (ii) the positive impact of externality reduction. Analogous to Corollary 1, Diamond (1973) shows that the optimal externality tax equals the average marginal externality.

Proposition 1 and Corollary 1 show that the average marginal bias $B(p)$ and the market demand curve $D_B(p)$ are sufficient statistics for computing the welfare effects of a subsidy or ban. One powerful implication is that while different consumers might be biased for different reasons (for example, biased beliefs, inattention, or present bias), the underlying behavioral model of the bias does not matter conditional on $B(p)$. This is also important because some models have consumers that are either fully unbiased or fully biased with some probability, while other models might have all consumers with a partial bias, and it may be difficult to empirically distinguish between these models. Notice also that even if many consumers are biased, the standard Harberger (1964a, b) formulas still hold exactly and the optimal corrective subsidy is still zero if the bias is not *systematic*, i.e., if the bias has mean zero at all values of $\hat{v}$.

B. *Estimating the Average Marginal Bias*

Chetty, Looney, and Kroft (2009); DellaVigna (2009); and Mullainathan, Schwartzstein, and Congdon (2012) categorize several approaches to estimating $B(p)$. One is to experimentally deliver what we call a *pure nudge*: a nonprice lever that does not change true values v but causes biased types to choose optimally—i.e., ensures that perceived values $\hat{v}$ equal true values v. For example, if biases arise from inattention or biased beliefs, then carefully designed information disclosure can address those biases without changing actual payoffs. Assuming that the researcher has access to a pure nudge, what strategies can identify $B(p)$ in our model?

One strategy is to directly compute $B(p)$ by evaluating $E[v - \hat{v} \mid \hat{v}]$ at each level of $\hat{v}$ using the following steps. First, elicit each consumer's perceived value $\hat{v}$, which gives the distribution $H(\hat{v})$. Second, apply the pure nudge. Third, observe each consumer's new valuation v (which is the true value by assumption), and then estimate the average change in valuation induced by the nudge for each level of initial valuation $\hat{v}$. This gives $E_G[v - \hat{v} \mid \hat{v} = p] = E_G[b \mid \hat{v} = p] = B(p)$. The TESS experimental design follows this strategy.

Such a strategy, however, requires within-subject identification that is difficult to implement in a natural field experiment. One potential alternative strategy might be to calibrate the change in price that has the same effect on market share as the nudge. Intuitively, if the effect of a nudge is twice the effect of a \$1 price change, then the

[15] When bias is nonnegative for all consumers, but also positive for some (with at least some of those types on the margin), $B(p) > 0$ for all p, and thus the optimal subsidy is positive. This generalizes the O'Donoghue and Rabin (2006) result that the optimal sin tax must be positive when at least some consumers are present-biased.

nudge might be increasing valuations by approximately \$2, thus giving $B(p) = \$2$. Chetty, Looney, and Kroft (2009) implement this by estimating how labels with total tax-inclusive prices affect market shares and comparing this to the price elasticity of demand. We call this measure the *equivalent price metric*:[16]

$$EPM(p) = \frac{D_B(p) - D_N(p)}{D'_N(p)}. \tag{3}$$

The benefit of the EPM strategy is that it can be implemented with a much simpler 2-by-2 experimental design that varies nudges and prices, like our in-store experiment. Mullainathan, Schwartzstein, and Congdon (2012) show that the EPM approximates $B(p)$ under a restrictive homogeneous bias assumption which in our notation corresponds to when $G(\cdot\,|\,v, p)$ is degenerate for all v, p.[17]

Unfortunately, $EPM(p) \neq B(p)$ in the general case with more realistic heterogeneity in bias. More broadly, any strategy such as the EPM that utilizes only the biased and unbiased demand curves D_B and D_N cannot identify $B(p)$, except under special conditions. Intuitively, this is because the EPM is a coarse statistic that cannot identify whether the most biased consumers are relatively more or less elastic to the subsidy. For example, if all consumers who undervalue E are so strongly biased against it that they all prefer I over E by at least \$2, then none of them will be marginal to a \$1 subsidy (implying $B(p) = 0$ for that subsidy level), even while a debiasing nudge would increase the demand at both baseline and subsidized prices (implying $EPM(p) > 0$).

For a stark mathematical example illustrating that both $B(p) \gg EPM(p)$ and $B(p) \ll EPM(p)$ are possible, suppose that $v \sim N(\mu, \sigma_v^2)$, $b \sim N(0, \sigma_b^2)$, and $\mathrm{cov}(v, b) = \sigma_{v,b}$. Using standard convolution and signal extraction formulas, $v - b \sim N(\mu, \sigma_v^2 + \sigma_b^2 - 2\sigma_{v,b})$ and $E_G(b\,|\,v - b = p) = \frac{-\sigma_b^2 + \sigma_{v,b}}{\sigma_v^2 + \sigma_b^2 - 2\sigma_{v,b}} \times (p - \mu)$. If we let $\sigma_{v,b} = \sigma_b^2/2$, then $v - b \sim N(\mu, \sigma_v^2)$, and thus $D_B(p) = D_N(p)$ for all p. However, $B(p) = E_G(b\,|\,v - b = p) = \frac{-\sigma_b^2/2}{\sigma_v^2}(p - \mu)$, which is positive for $p < \mu$ and negative for $p > \mu$. Thus, the average marginal bias can be arbitrarily large or small *even when the biased and unbiased demand curves are identical*, meaning that the nudge has no effect on market share at any price. Proposition 3 in online Appendix D.B extends this example and shows that even under strong restrictions including linear demand curves and tightly bounded support for the bias, it is still possible to have $B(p) \gg 0$ or $B(p) \ll 0$ when $EPM(p) = 0$.

To provide further intuition, online Appendix D.B includes an example with two bias types that illustrates the mechanisms causing the divergence between $EPM(p)$ and $B(p)$. A key statistic for understanding this divergence is each bias type's "elasticity ratio": the price elasticity in the biased state divided by the price elasticity in

[16] As in Mullainathan, Schwartzstein, and Congdon (2012); Baicker, Mullainathan, and Schwartzstein (2015); and Chetty, Looney, and Kroft (2009), we divide by the slope of the unbiased demand curve, though in practice one could instead normalize by D'_B or the average of the slopes. We focus on this normalization because it approximates $B(p)$ under the broadest range of assumptions.

[17] Formally, if $b(p)$ is the bias of all consumers marginal at price p, then $D_B(p) = D_N(p + b(p)) \approx D_N(p) + D'_N(p)b(p)$, from which it follows that $b(p) \approx EPM(p)$.

the unbiased state. We show that an approximate condition for $B(p) > EPM(p)$ is that the high type's elasticity ratio is greater than the low type's, and conversely for $B(p) < EPM(p)$.

Chetty, Looney, and Kroft (2009); DellaVigna (2009); and Mullainathan, Schwartzstein, and Congdon (2012) also discuss a second approach to estimating $B(p)$, which we call *comparing demand responses*. This exploits the fact that optimizing consumers should care only about a good's total costs, so demand should be equally responsive to changes in purchase prices versus changes in potentially less-salient add-on costs such as sales taxes, shipping and handling charges, or energy costs. A large literature uses this approach, including Abaluck and Gruber (2011); Allcott and Wozny (2014); Busse, Knittel, and Zettelmeyer (2013); the alcohol tax analysis in Chetty, Looney, and Kroft (2009); Hossain and Morgan (2006); and others. Using a general model that encompasses many settings including energy efficiency and tax salience, we show in online Appendix D.C that the comparing demand responses approach approximates $B(p)$ under an even stronger set of conditions than those that are required for the EPM. Thus, our theoretical and empirical results on how the EPM poorly approximates $B(p)$ also suggest that the comparing demand responses approach poorly approximates $B(p)$.

C. *Biases Eliminated by Information Provision*

In practice, a given nudge addresses some biases and not others. In the context of lightbulbs, we are interested in identifying the effects of imperfect information and inattention. To do this, we use informational interventions that fully inform consumers about energy costs and bulb lifetimes and aggregate upfront and future costs into one total user cost.

The idea that information provision could identify bias is inspired by Chetty, Looney, and Kroft (2009), who identify inattention to sales taxes by informing consumers of tax-inclusive purchase prices in a supermarket. In justifying their approach, they write that "when tax-inclusive prices are posted, consumers presumably optimize relative to the tax-inclusive price." Similarly, it seems reasonable to assume that consumers optimize relative to lightbulb lifetimes and energy costs after we provide them with information about these attributes. Providing information plausibly eliminates the following types of biases:

(i) Biased beliefs, as tested by Allcott (2013); Attari et al. (2010); Bollinger, Leslie, and Sorensen (2011); and others. In our context, consumers may know that CFLs use less energy but mis-estimate the cost savings.

(ii) Exogenous inattention to energy as a "shrouded" add-on cost, related to Gabaix and Laibson (2006) or Heidhues, Kőszegi, and Murooka (2014).

(iii) Costly information acquisition, as in Gabaix et al. (2006) and Sallee (2014). This includes many standard models of imperfect information in which the consumer incurs a cost to learn about energy efficiency and, in the absence of that information, assumes that different goods have the same energy efficiency.

(iv) "Noisy" and costly thinking models, as in Gabaix (2014); Sims (2003); Caplin and Dean (2014); and others. In these models, consumers might at first have only a noisy representation of the true value of energy efficiency, but thinking allows a more precise representation, subject to either a cognitive constraint or an explicit thinking cost.

Information interventions would not affect all biases that could affect lightbulb demand. For example, "bias toward concentration" (Kőszegi and Szeidl 2013) could cause consumers to undervalue electricity costs because they occur in a stream of small future payments. Kőszegi and Szeidl (2013) point out that reframing the stream of payments as one net present value, as our interventions do, does not necessarily address this possible bias. Furthermore, consumers could be imperfectly informed about or inattentive to other attributes not discussed in our informational interventions.

Denoting $A(p)$ as the average marginal bias from other biases not addressed by information provision and $\phi(p)$ as the average marginal uninternalized externality, equation (4) generalizes to

$$W'(s) = (s - B(p) - A(p) - \phi(p))D'_B(p). \tag{4}$$

The generalization of (2) would follow similarly. This equation illustrates that estimates of average marginal bias from imperfect information and inattention can be easily extended into a more comprehensive welfare analysis when combined with complementary estimates of $A(p)$ or $\phi(p)$. We illustrate this approach in Section IVB.

This section has clarified that an experiment to identify the welfare effects of a subsidy or ban to address imperfect information and inattention must have two features. First, the treatment must plausibly approximate a "pure nudge": it should provide clear information while minimizing demand effects and confounds. Second, the design must identify the sufficient statistics for welfare analysis: average marginal bias $B(p)$ and market demand curve $D_B(p)$.

III. TESS Experiment

A. *Survey Platform and Population*

We implemented the artefactual field experiment through Time-Sharing Experiments for the Social Sciences (TESS), which provides a nationwide sample of more than 50,000 consumers for computer-based experiments. Many economists have used this platform, including Allcott (2013); Fong and Luttmer (2009); Heiss, McFadden, and Winter (2007); Newell and Siikamäki (2013); and Rabin and Weizsacker (2009). One key feature of TESS is that the recruitment process generates a sample that is as close as practically possible to nationally representative on unobservable characteristics, which allows more credible generalization to the US population. Unrecruited volunteers are not allowed to opt in. Instead, potential TESS participants are randomly selected from the US Postal Service Delivery Sequence File and recruited through an extensive series of mailings and telephone calls. About 10 percent of invitees actually become participants. Households without computers

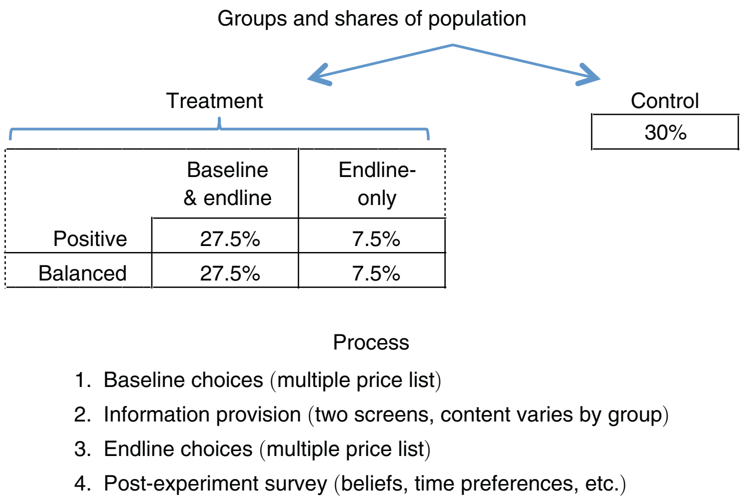

Figure 1. TESS Experimental Design

are given computers in order to complete the studies. We reweight all TESS results to be nationally representative on observables.

Participants take an average of two studies per month, and no more than one per week. Of the qualified participants who began our survey, about three-fourths completed it, giving a final sample size of 1,533. Per TESS rules, we could not force participants to answer all questions, although we successfully negotiated to require responses to the most important ones.

B. *Experimental Design*

Overview.—Figure 1 gives a synopsis of the TESS experimental design. The study had four parts: baseline lightbulb choices, information provision screens, endline lightbulb choices, and a post-experiment survey. This design is both within-subject (we have both pre-information and post-information choices) and between-subject (consumers received different information screens).

Each consumer was randomly assigned to Treatment or Control, and within Treatment to a matrix of four subtreatments. These group assignments determined which two information screens the consumer would receive. As we discuss in more detail below, the "Positive" subtreatment included information about the cost savings from CFLs, while the "Balanced" subtreatment included information about cost savings *and* the CFL's negative attributes. The right column in the matrix of subtreatments is the Endline-only treatment, in which consumers skipped the baseline choices and began directly with the information provision. Except when specified, we pool these four subtreatments together and refer to them as the "Treatment" group; we show in Section IIIE that effects of these four subtreatments are not statistically distinguishable.

Choices were incentive compatible. Consumers were given a $10 "shopping budget" that they could use to purchase packages of incandescents or CFLs at varying prices. Each consumer made 15 baseline choices and 15 endline choices via standard multiple price lists, and 1 of those 30 was randomly selected as the "official purchase." TESS staff shipped consumers the lightbulb package they had chosen in that official purchase, charged the price of the package, and added the remainder of the $10 to consumers' TESS bonus accounts. Online Appendix A contains screen shots from the experiment.

Baseline and Endline Lightbulb Choices.—Consumers chose between two lightbulb packages, one containing one Philips 60-watt-equivalent compact fluorescent lightbulb, and the other containing four Philips 60-watt incandescent lightbulbs. The two lightbulb packages were chosen to be as comparable as possible, except for the CFL versus incandescent technology. While the choice screen had only pictures of the bulbs, consumers could click to a "Detailed Product Information" window, which included light output in lumens, a quantitative measure of light color, energy use in watts, and other information. About 19 percent of consumers opened this window. Both packages typically sell online for about $4, so the market relative price is $p = 0$. We did not tell consumers these typical prices.

One-half of consumers were randomly assigned to see the incandescent on the left, labeled as "Choice A," while the other half were assigned to see the incandescent on the right, labeled as "Choice B." The choice screen included 15 decisions in which the relative price of Choice A increased monotonically in decision number. For example, Decision Number 1 offered Choice A for free and Choice B for $10, Decision Number 8 had equal prices of $4, and Decision Number 15 offered Choice A for $10 and Choice B for free. Consumers spent a median of 3 minutes on the baseline choice screen and 1 minute, 20 seconds on the endline choice screen.

Information Provision.—After the baseline lightbulb choices, each group received two information screens in random order. The screens were designed to closely parallel each other, to minimize the chance that idiosyncratic factors other than the information content could affect purchases. Each screen included about 10–15 lines of text, plus a graph to illustrate the key concept. The text was read verbatim on an audio recording, which is available as part of the online materials. At the bottom of the information screen, there was a "quiz" on a key fact.

Two design features help to ensure that consumers processed and understood the information. First, using multiple channels to convey information (text, graphical, and audio) means that people who learn in different ways had a higher chance of internalizing the information. Second, the quiz forced respondents to internalize the information if they had not done so initially.

The different groups received some combination of the following four screens:

(i) **Treatment Information:** As described below, this screen compared electricity costs and replacement costs for CFLs and incandescents.

(ii) **Negative Information:** This screen was designed to present information about disposal and warm-up time, two ways in which CFLs might not be preferred to incandescents. It explained that "because CFLs contain mercury, it is recommended that they be properly recycled instead of disposed of in regular household trash." It also explained that "after the light switch is turned on, CFLs take longer to warm up than incandescents," and graphed a typical CFL's warm-up time.

(iii) **Number of Bulbs:** This screen presented information on the number of lightbulbs installed in residential, commercial, and industrial buildings in the United States.

(iv) **Sales Trends:** This screen detailed trends in total US lightbulb sales between 2000 and 2009.

Control group consumers received the Number of Bulbs and Sales Trends screens. We designed these screens to have no impact on relative WTP, and neither screen mentioned energy costs or distinguished between CFLs and incandescents. The Positive Treatment group received the Treatment Information screen plus a randomly selected one of the two Control screens. The Balanced Treatment group received the Treatment Information and Negative Information screens. We included the Balanced Treatment to both test whether consumers might be inattentive to or misinformed about product attributes other than energy costs and also help test for experimenter demand effects, as it is especially unlikely that this group would assume that the experimenter wanted them to purchase CFLs.

The Treatment Information screen began by explaining that CFLs both last longer and use less electricity compared to incandescents, and it translated these differences into dollar amounts using simple calculations at typical prices. The bottom line was:

> *Thus*, *for eight years of light*, *the total costs to purchase bulbs and electricity would be:*
>
> - $56 *for incandescents:* $8 *for the bulbs plus* $48 *for electricity.*
> - $16 *for a CFL:* $4 *for the bulbs plus* $12 *for electricity.*

The graph was a simple bar graph illustrating these bullets.

The quiz question at the bottom of the screen was: "For eight years of light, how much larger are the total costs (for bulbs plus electricity) for 60-watt incandescents as compared to their CFL equivalents?" The correct answer could be inferred from the information on the screen: $56 for incandescents − $16 for CFLs = $40. Sixty-four percent of consumers correctly typed $40. Those who did not were prompted to try again. After this point, 73 percent of consumers had typed $40. The remaining consumers were told that the correct answer was $40. After this point, 89 percent of consumers had typed $40. The 11 percent failure rate is higher than we expected. However, results in online Appendix Table A.7 show that the average treatment effect (ATE) on WTP is only 4 percent higher and statistically indistinguishable when excluding consumers who failed, suggesting that the failures do not meaningfully affect our results.

Consumers spent a median of 2 minutes, 12 seconds reading the Cost Info Screen and completing the quiz question. This substantial time, along with the quiz, show that the vast majority of Treatment group consumers engaged with and understood the information.

C. *Data*

The multiple price list allows us to observe choices at relative prices of $p \in \{-10, -8, -6, -4, -3, -2, -1, 0, 1, 2, 3, 4, 6, 8, 10\}$. Consumers' relative willingness-to-pay (WTP) for the CFL, denoted w, must lie between the highest relative price when they choose the incandescent and the lowest relative price when they choose the CFL. We assume that consumers who switch choices between any two relative prices have relative WTP w equal to the mean of those prices. For example, consumers who choose CFLs at relative price $p = 0$ (i.e., when both packages cost \$4) but choose incandescents when incandescents are \$1 cheaper are assumed to have $w = \$0.50$.[18]

Some consumers had censored WTPs: they preferred either Choice A or Choice B at all relative prices. These consumers were asked to report a hypothetical relative price at which they would prefer the other choice. Across all censored consumers, the median absolute value of self-reported relative WTP was \$15, and we impute a relative WTP of \$15 and $-\$15$ for top-coded and bottom-coded consumers, respectively. We will demonstrate the sensitivity of the results to this assumed value.[19]

Online Appendix B presents sample characteristics and shows that the treatment groups are balanced. That Appendix also reports correlations between baseline WTP and observable characteristics. Men, democrats, environmentalists, consumers who had previously reported taking steps to conserve energy, and those with higher discount factors have higher demand for CFLs. (The discount factors are the δ parameter in a β, δ model of present bias, as calibrated from hypothetical intertemporal tradeoffs in the post-experiment survey.) These correlations conform to our intuition and build further confidence that the differences in WTP are meaningful. Renters and more present-biased (lower β) consumers do not have lower WTP for CFLs conditional on other observables. This provides no support for the hypotheses that real estate market failures or present bias affect the lightbulb market, which reinforces the importance of our focus on imperfect information and inattention.[20]

[18] Eight percent and 4.3 percent of consumers in the baseline and endline choices, respectively, did not choose monotonically: they chose Choice A at a higher relative price than another decision at which they chose Choice B. These consumers were prompted with the following message: "The Decision Numbers below are organized such that Choice A costs more and more relative to Choice B as you read from top to bottom. Thus, most people will be more likely to purchase Choice A for decisions at the top of the list, and Choice B for decisions at the bottom of the list. Feel free to review your choices and make any changes." After this prompt, 5.3 percent and 3.6 percent of consumers still chose nonmonotonically in baseline and endline choices, and we code their WTP as missing.

[19] For treatment effects and welfare analysis, we technically should impute using mean censored WTP, not median. However, the distribution of self-reports is skewed, with a small number of consumers reporting very large values. Because these self-reports were not incentive compatible, we wish to be cautious about using them in the primary analysis.

[20] This may not be surprising. Present bias over *cash flows* might cause consumers to buy an incandescent to reduce current expenditures, but as Andreoni and Sprenger (2012) and many others have pointed out, agents in most models would be present-biased over *consumption*, and most US consumers have enough liquidity that paying the incremental few dollars for a CFL does not immediately affect consumption. Present bias could induce people to procrastinate in buying and installing CFLs, but this would play no role in the TESS experiment because we forced consumers to make an active choice.

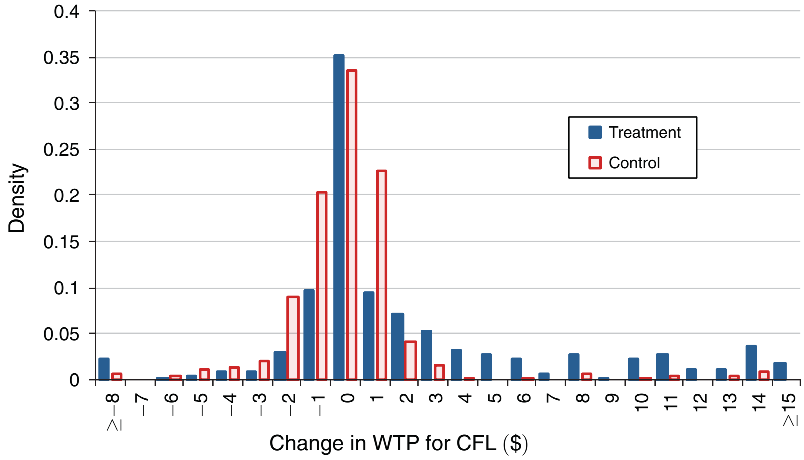

Figure 2. Histogram of Relative WTP Changes

Notes: This figure plots the histogram of changes from baseline to endline in relative willingness-to-pay for the information Treatment and Control groups. Treatment pools both Positive and Balanced Treatment groups, although the Endline-only Treatment group is excluded because there is no baseline WTP from which to calculate a change. Observations are weighted for national representativeness.

D. *Results*

The results in this section begin to answer our first research question: how does information affect demand? Unlike the policy analysis in Section IV, these results do not require the assumption that the treatment is a pure nudge that only eliminates biases.

Quantity Effects, Demand Slopes, and the Equivalent Price Metric.—Figure 2 presents a histogram of the within-subject changes in WTP between baseline and endline. About 90 percent of Control group consumers either have exactly the same WTP or change by $2 or less. In Treatment, there is a mass to the right of the figure, with 36 percent of people increasing WTP by $1–$10. This figure illustrates that the Control information screens were successful in the sense that they did not affect average WTP. It also shows that the Treatment information both increased average WTP and had very heterogeneous effects.

Figure 3 presents endline demand curves. If some Treatment group consumers want to be internally consistent between baseline and endline, endline choices would be biased toward the baseline compared to a design without baseline choices (Falk and Zimmermann 2013). The Endline-only Treatment demand curve lies directly on top of the Baseline and Endline Treatment curve, illustrating that internal consistency does not bias the information effects.[21] Both Treatment curves are shifted out relative to Control. At market prices ($p = 0$), Treatment group CFL market share

[21] Formal tests confirm that Endline-only demand is not statistically different than Baseline and Endline demand: the share of consumers with endline WTP $w^1 > w^+$ does not differ statistically between these two demand curves at any level of w^+.

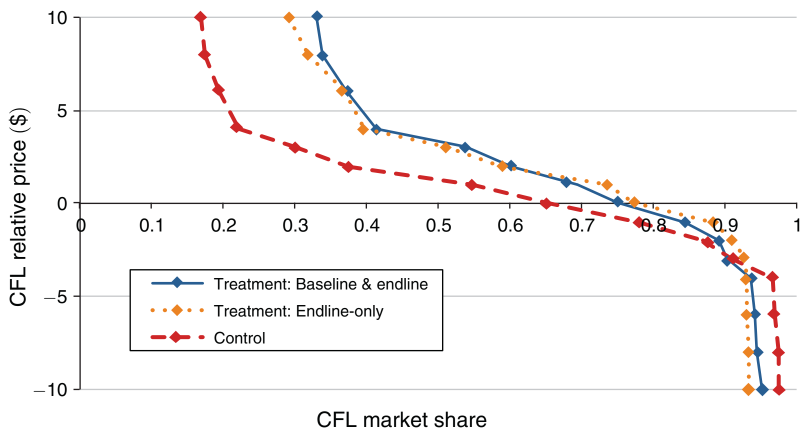

Figure 3. Endline CFL Demand Curves

Notes: This figure plots the endline demand curves from the TESS experiment. Observations are weighted for national representativeness.

is about 77 percent, a 12-percentage-point increase relative to control. The market share effect differs substantially at different relative price levels. For example, at relative price $p = -1$ (i.e., after a \$1 CFL subsidy), the effect is 7 percentage points. A key result is that at the market price, a meaningful share of consumers still prefer incandescents even after being informed about the CFL's cost advantage.

Figure 3 shows that demand is highly price-responsive near market prices. For example, between relative prices of 0 and -1, Treatment group demand has slope of 10 percent per \$1. This is not just an idiosyncratic feature of the TESS experimental setting:[22] we estimate in Section V that demand is equally or perhaps even more price-responsive at market prices in the in-store experiment. We return to this issue in Section IVB.

If demand is fairly inelastic because many consumers have strong preferences, then it is more likely for information to have small effects on market share. We thus use the equivalent price metric to benchmark the effect of information against the effect of prices. Defining $\Delta p = p_h - p_l$, we approximate the average EPM over price interval $p \in [p_l, p_h]$ as

$$(5) \qquad EPM[p_l, p_h] \approx \frac{(\Delta Q(p_l) + \Delta Q(p_h)) \cdot \frac{1}{2}}{(D_N(p_l) - D_N(p_h))/\Delta p}.$$

This equation approximates the average EPM over an interval by the average of the quantity effects at the endpoints divided by the slope between the endpoints. The

[22] Two features of the TESS experiment could in theory have caused highly elastic demand. First, if lightbulbs were perishable and consumers did not immediately need one, consumers would buy the cheapest package instead of revealing the WTP they would have if they did need one. In practice, lightbulbs are easily stored, and we reminded consumers of this fact in the introductory text. Second, if it were costless to resell the experimental purchase and replace it with a different purchase outside the experiment, consumers who know that the typical retail prices are approximately equal would always buy the cheaper package. In order to avoid making this salient, the experiment did not include information about the bulbs' typical retail prices. In practice, it seems unlikely that consumers resold the packages that they received.

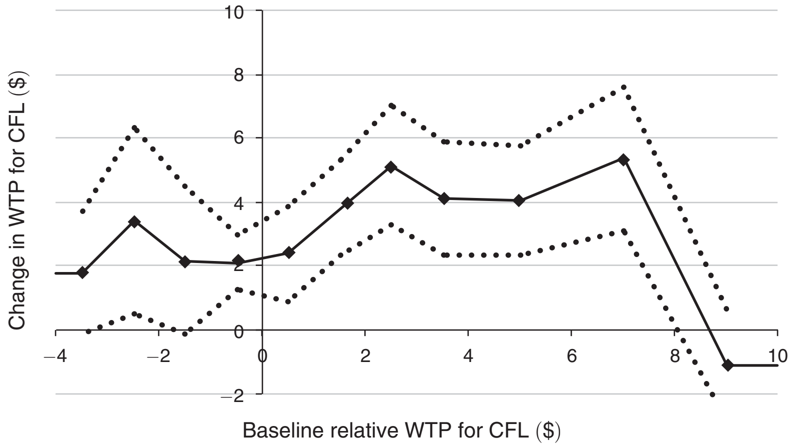

Figure 4. Conditional Average Treatment Effects by Level of Baseline WTP

Notes: This graph presents the conditional average treatment effects of information provision for consumers at each level of baseline relative WTP. Due to limited sample size, baseline WTPs less than −\$3 are grouped together, as are baseline WTPs greater than \$8. Dotted lines are 90 percent confidence intervals. Observations are weighted for national representativeness.

average EPM just below the market price, denoted $EPM[-1, 0]$, can be calculated using numbers in the past few paragraphs: $EPM[-1, 0] \approx \frac{(0.07 + 0.12)/2}{0.1} \approx \$0.94$. On this interval, information affects CFL market share about as much as a \$0.94 price reduction.

Conditional Average Treatment Effects.—Figure 4 presents the conditional average treatment effects on WTP. Each diamond on the figure represents the average treatment effect for consumers with baseline relative WTP w^0 in an interval $w^0 \in [p_l, p_h]$, where p_l and p_h are adjacent points on the multiple price list. There is a thinner density of consumers with outlying high or low values of w^0, so for precision, we group all $w^0 < -\$3$ and all $w^0 > \$8$. Most of these CATEs are in the range of \$2–\$4, except at the highest baseline WTP, where the CATE is statistically zero. This is simply due to top-coding: consumers who start at the top WTP in the multiple price list cannot increase their WTP further. Because these inframarginal consumers are unaffected by the subsidy and the ban, this will not affect the welfare calculations. After excluding consumers with top-coded and bottom-coded baseline WTP, the CATEs are statistically significantly increasing in w^0. Given that the effects at different price levels are important for our policy analysis, this slope highlights the importance of an experimental design like this one that identifies a CATE function instead of approximating it with a single average treatment effect.

Comparing Conditional Average Treatment Effects to the Equivalent Price Metric.—How closely does the EPM approximate the CATE on WTP in these data? Our theoretical arguments in Section IIB and online Appendix D.B show that these two statistics can be very different. Our TESS experimental design enables us to empirically evaluate these theoretical differences.

Figure 4 shows that the CATE on the interval $p \in [-1, 0]$ is \$2.11, which is more than twice the \$0.94 EPM. In Section B.B of the online Appendix, we compare the EPM and CATE at all nine price intervals where both can be calculated. Four of the nine differ with more than 90 percent confidence, and on average, the EPM differs from the CATE by 49 percent.

Given that the CATE on WTP is what we need for policy analysis, our findings of a substantial difference between the EPM and the CATE highlight the importance of a design like the TESS experiment that directly identifies it, instead of the standard 2-by-2 designs that can only approximate it with the EPM. In Section IVB, we return to this point and quantify implications of this divergence for welfare estimates.

E. *Average Treatment Effects, Robustness Checks, and Alternative Estimates*

To more formally assess robustness, we now calculate average treatment effects on relative WTP and discuss alternative estimates. Let T_i be an indicator for whether the consumer is in the Treatment group, denote $\mathbf{X}_i$ as consumer i's vector of individual characteristics, and denote μ_i as a vector of indicator variables for each level of baseline WTP. We estimate the average treatment effects of information provision on endline WTP w_i^1 using ordinary least squares (OLS) with robust standard errors:

$$w_i^1 = \tau T_i + \gamma \mathbf{X}_i + \mu_i + \varepsilon_i. \tag{6}$$

Table 1 presents the results. Because baseline WTP, and thus μ_i, are not available for the Endline-only group, all columns except column 5 exclude that group. Column 1 presents the unconditional difference in means. Column 2 adds the μ_i controls, while column 3 further adds $\mathbf{X}_i$ to give the exact specification from equation (6). The sample size decreases in column 3 because at least one characteristic is missing for 15 consumers. In column 3, information increased relative WTP for the CFL by an average of \$2.30, and the estimates in columns 1 and 2 are economically and statistically identical.

Top-coding and bottom-coding of WTP mechanically influence the treatment effect. Consumers with baseline WTP equal to the maximum could not reveal a post-treatment increase in WTP, and any consumers with baseline WTP equal to the minimum could not reveal a decrease in WTP. Because the treatment tends to increase WTP, the former effect should dominate, and the average treatment effect should be understated. This connects to the result in Figure 4 that consumers with the highest baseline WTP have statistically zero treatment effect. Column 4 excludes consumers with top-coded or bottom-coded baseline WTP of \$15 or $-$\$15, and the estimated effect increases to \$3.10.[23]

[23] Relatedly, the assumed mean censored value of \$15 caps the increase in WTP that any consumer can reveal. Since a larger share of endline WTP is top-coded in treatment relative to control (29 percent versus 16 percent), increasing this assumed value should increase the treatment effect. Regressions in online Appendix Table A.7 show that when we alternatively assume mean censored values of \$12 (\$20) instead of \$15, the ATE changes to \$1.98 (\$2.83).

Table 1—Effects of TESS Information Treatment

	(1)	(2)	(3)	(4)	(5)	(6)
1(Treatment)	2.54	2.28	2.30	3.16	2.29	2.14
	(0.55)***	(0.36)***	(0.37)***	(0.37)***	(0.54)***	(0.50)***
1(Endline-only)					−0.44	
					(0.76)	
1(Positive treatment)						0.35
						(0.56)
R^2	0.03	0.57	0.58	0.33	0.04	0.58
Observations	1,203	1,203	1,188	919	1,449	1,188
Baseline WTP dummies μ	No	Yes	Yes	Yes	No	Yes
Individual characteristics	No	No	Yes	Yes	Yes	Yes
Exclude max./min. baseline WTP	No	No	No	Yes	No	No
Include Endline-only group	No	No	No	No	Yes	No

Notes: This table presents estimates of equation (6). The outcome variable is endline willingness-to-pay for the CFL. 1(Treatment) pools all information subtreatments. Robust standard errors in parentheses. Observations are weighted for national representativeness.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Column 5 adds the Endline-only Treatment group, while excluding the μ_i indicators. Estimates show that the Endline-only group's WTP is not statistically different from the Control group's endline WTP. This confirms the graphical result in Figure 3 that internal consistency does not bias the estimates.

With any experiment other than a natural field experiment, demand effects might arise: participants might change their actions to comply with, or perhaps defy, the perceived intent of the study. We address demand effects in three ways. First, because the Balanced treatment disclosed both positive and negative information about CFLs, these consumers should be less likely to perceive that the experimenters intended to persuade them to purchase CFLs. If demand effects typically cause consumers to comply with the study's perceived intent, the Positive treatment would have larger effects on WTP than the Balanced treatment. Column 6 of Table 1 includes an indicator for the Positive Treatment group, showing that the effects do not differ statistically, and the point estimates are similar. Because the effects do not differ between the Balanced and Positive Treatment groups, we have combined these groups in other parts of the analysis.

Second, demand effects are less likely if participants cannot identify the intent of the study. The post-experiment survey asked consumers what they thought the intent of the study was. Results available in Table A.4 of online Appendix B show that there is substantial dispersion in perceived intent within groups, which suggests that there is no one clear way in which demand effects might act.

Third, if demand effects are present, they should differentially affect people who are more able to detect the intent of the study and are more willing to change their choices given the experimenter's intent. We proxy for this ability using the self-monitoring scale (Snyder 1974), and we find no evidence that self-monitoring ability moderates the treatment effect. Details are available in online Appendix B.C.

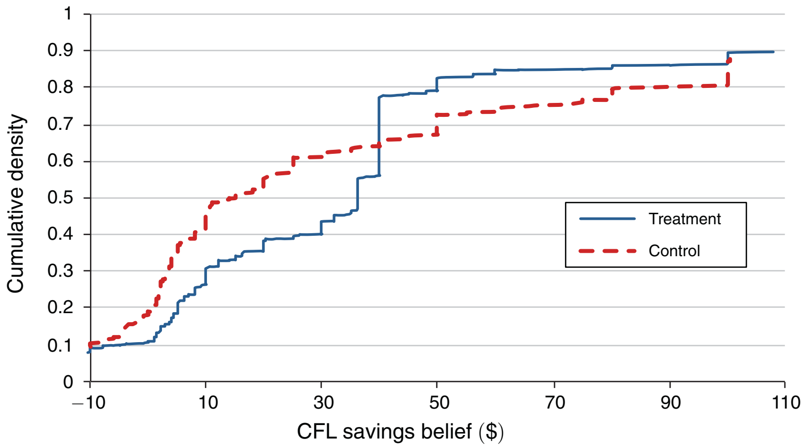

Figure 5. Cumulative Density Function of Cost Savings Beliefs

Notes: This figure plots the cumulative density of beliefs about the electricity cost savings from CFLs compared to incandescents for 8,000 hours of light, from the TESS post-experiment survey. The Treatment group pools all information subtreatments. Observations are weighted for national representativeness.

F. *Effects on Beliefs*

How much did the information treatment affect choices through increased attention versus updated beliefs? The post-experiment survey elicited beliefs over how much less it costs to buy electricity for a CFL versus incandescents over the CFL's 8,000-hour rated life, at national average electricity prices. Figure 5 presents the cumulative density functions (CDFs) of responses in Treatment and Control. The figure has three key features. First, beliefs are highly dispersed. Second, the information treatment substantially reduces this dispersion, and about 30 percent of Treatment group consumers have beliefs that are "correct" in the sense that they correspond to lifetime cost savings provided in the Information Treatment screen.[24] Results in online Appendix Table A.7 show that these consumers have statistically significantly larger ATE on WTP, and the ATE is 34 percent higher when estimated only off of these consumers. We return to this group in alternative welfare analyses in Section IV.

Third, the treatment increases perceived savings at the median of the distribution and all percentiles below the sixty-fifth. These data suggest that the information treatment may act at least partially through belief updating—unlike in Chetty, Looney, and Kroft (2009), whose belief surveys suggests that sales tax information acts through increasing salience. Because of the wide dispersion, however, the average treatment effect on beliefs is statistically indistinguishable from zero and

[24] According to information on the Information Treatment screen, the correct answer to this question was $36 ($48 for the incandescent minus $12 for the CFL). While some Treatment group consumers put $36, many others put $40, apparently misreading the question and also including the $4 in bulb replacement cost savings. Thirty percent of consumers' beliefs were between $36 and $40, inclusive.

Table 2—Welfare Analysis Using TESS Results

CFL subsidy ($/package) (1)	Average relative WTP of marginal consumers ($/package) (2)	Average marginal bias ($/package) (3)	Incremental demand change (share of packages) (4)	Incremental welfare effect ($/package) (5)	Cumulative welfare effect ($/package) (6)
1	−0.5	2.11	0.126	0.20	0.20
2	−1.5	2.16	0.052	0.03	0.24
3	−2.5	3.41	0.028	0.03	0.26
4	−3.5	1.77	0.030	−0.05	0.21
6	−5	1.77	0.006	−0.02	0.19
8	−7	1.77	0.008	−0.04	0.15
10	−9	1.77	0.003	−0.02	0.13
∞	−15	1.77	0.043	−0.57	−0.44

Notes: This table uses the TESS experiment results to calculate the welfare effects at different levels of the CFL subsidy. Observations are weighted for national representativeness. Average Marginal Bias is the point estimates from Figure 4. Incremental Welfare Effect is from equation (2).

very imprecisely estimated, making it impossible to precisely compare it to ATEs on WTP.

IV. Welfare Analysis of Subsidies and Bans

The theoretically ideal way to address imperfect information and inattention would be a powerful and costless nationwide information disclosure technology. Subsidies and standards have been proposed as second-best policies with the idea that practically feasible information disclosure programs either do not fully remove bias or are too costly to scale. We now combine the model in Section II with the TESS results in Section III to evaluate these policies in the lightbulb market.

Our base scenario assumes that bias from imperfect information and inattention is the only market distortion. We view this as a reasonable simplification given the discussion in Section I, and equation (4) shows how the welfare analysis can be easily generalized with credible estimates of additional bias $A(p)$ and uninternalized externality $\phi(p)$. Separately, our base scenario also assumes that the conditional average treatment effect on WTP equals the average marginal bias from imperfect information and inattention.[25] Formally, this is

ASSUMPTION 1: $\tau(p) = B(p)$.

Assumption 1 holds if the information treatment is a pure nudge, although it is slightly weaker: by identifying off of the CATE, it allows the information treatment to have additional idiosyncratic effects on WTP that are mean-zero at each price level.

[25] In the language of Bernheim and Rangel (2009), we define Control group choices as *provisionally suspect* due to the possibility of imperfect information processing. If choices differ between Treatment and Control, we delete Control group choices from the welfare-relevant domain.

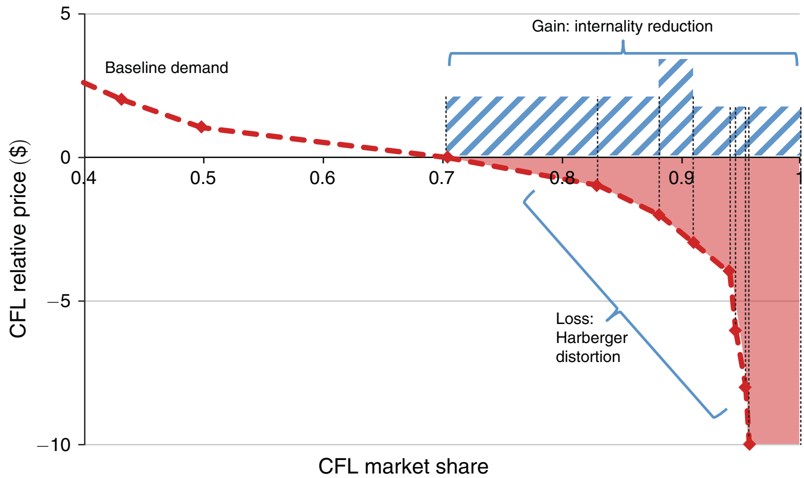

Figure 6. Welfare Calculation Using TESS Experiment

Notes: This figure illustrates the welfare effects of increases in the CFL subsidy using the TESS experiment results. Internality Reduction and Harberger Distortion are as defined in equation (2). Observations are weighted for national representativeness.

A. *Base Scenario Results*

Table 2 evaluates the welfare impacts of incremental increases in the CFL subsidy using equation (2) with the TESS data. We begin at the market price $p = 0$ and then increment the subsidy by amounts corresponding to the price differences in the multiple price list. Column 2 contains the average relative WTP for the CFL for consumers marginal to each subsidy increment. Column 3 presents the CATE on WTP $\tau(p)$ for consumers at that level of baseline WTP, which is just the point estimates from Figure 4. Under Assumption 1, this equals the average marginal bias. Column 4 contains the change in market share from the subsidy increase. Column 5 presents the welfare effect of the subsidy increase, using equation (2). This column is simply column 4 multiplied by the sum of columns 2 and 3. When the average marginal bias is larger than the absolute value of average marginal WTP, internality reduction outweighs the Harberger distortion, and the subsidy increment increases welfare. Once the subsidy is so large that average marginal WTP is highly negative, however, further increases in the subsidy decrease welfare. Column 6 presents the cumulative welfare effect of changing the subsidy from zero to the amount listed in that row. Columns 3–6 are measured with sampling error, which we consider below.

Figure 6 illustrates the calculations in Table 2. The dashed curve is demand from the baseline choices, $D_B(p)$. The vertical black lines are guides to illustrate the changes in market share from each increment in the subsidy: because $p = 0$ at market prices, the first vertical line is drawn at a market share value obtained at $s = 0$, the second vertical line is drawn at a market share value corresponding to $s = 1$, and so forth. The shaded rectangles above the x-axis reflect the internality reduction in equation (2). Their height is the average bias of all consumers marginal to the corresponding change in the subsidy, while their width corresponds the change in market shares. Thus the area of each rectangle corresponds to the "internality reduction" term $-\Delta s D'(p) E_H[B(x) \mid p - \Delta s \leq x \leq p]$. The (triangle and) trapezoids below the x-axis are the familiar Harberger (triangle and) trapezoids generated by the distortion to perceived utility. When the area of the internality reduction rectangles exceeds the area of the Harberger trapezoids, the subsidy increment increases welfare.

Table 2 shows that of the discrete values that we can assess given the TESS multiple price list, the globally optimal subsidy is \$3. Correspondingly, Figure 6 shows that once the subsidy exceeds \$3, the area of the incremental Harberger trapezoid exceeds the area of the incremental internality reduction. A subsidy of \$3 is slightly larger than the CFL subsidies offered by many electric utilities over the last decade, which were typically in the range of \$1–2 per bulb.

Under the assumptions of our model, the welfare effects of an incandescent ban equal the effects of an infinite CFL subsidy. In Figure 6, this is the sum of all internality reduction rectangles above the x-axis minus all Harberger trapezoids below the x-axis. As shown in Table 2, this sum is negative: in our model, the ban reduces welfare by \$0.44 per package sold. While a ban is mechanically weakly worse than the optimal subsidy under the model's assumptions, the empirical magnitude is remarkable: in absolute value, the losses from a ban are 65 percent larger than the gains from the optimal subsidy. Thus, while a ban improves welfare for some biased consumers, this is far outweighed by the losses to consumers who strongly prefer incandescents even after being informed of the CFL's benefits.

In practice, we view Assumption 1 as only an approximation: it is possible that the information treatment might have effects other than removing informational and attentional bias. Furthermore, there may be additional market distortions. Additional distortions or mismeasurement of imperfect information and inattention that increased our base case $B(p)$ function by a multiplicative factor of 1.69 or more would cause the ban to increase welfare in our model. Alternatively, any total homogeneous distortion larger than \$3.60 per package would cause the ban to increase welfare. We further explore these issues below.

B. *Alternative Assumptions*

Table 3 presents results under alternative assumptions. The first row restates the base estimates in Table 2, adding a third column that reports the welfare effects of a ban as a share of incandescent lightbulb buyers' total perceived consumer surplus from having the incandescent available. Put differently, column 3 contains the net welfare effects divided by the total area of the shaded Harberger trapezoids.

The top set of alternative results involve alternative assumptions for relative WTPs censored by the ends of the multiple price list. Top-coding and bottom-coding WTP has two opposing effects on the welfare calculation. First, the treatment causes many Treatment group consumers to be willing to pay the maximum for the CFL. Assuming a larger mean WTP for this top-coded group increases the treatment effect, implying a larger bias and thus larger welfare gains from corrective policies. Second, however, the welfare effects of a ban depend importantly on the lower tail of the WTP distribution: if some consumers very strongly prefer incandescents, banning them can cause large welfare losses. Recall that in the base case, we assume that the mean values of top-coded and bottom-coded WTPs are \$15 and $-$\$15, respectively. Rows 2 and 3 assume {\$12, $-$\$12} and {\$20, $-$\$20}, respectively, for

Table 3—Welfare Analysis under Alternative Assumptions

Row	Scenario	Optimal subsidy (\$/package) (1)	Welfare effect of ban (\$/package) (2)	Welfare effect of ban (percent of surplus) (3)
1.	Base	3	−0.44	−41
Panel A. Alternative censoring assumptions: if censored, assume:				
2.	WTP={\$12, −\$12}	3	−0.34	−36
3.	WTP={\$20, −\$20}	3	−0.60	−47
4.	self-reported hypothetical WTP	3	−0.61	−43
Panel B. Alternatives to Assumption 1: scale average marginal bias to match:				
5.	consumers who pass Treatment Info screen "quiz"	3	−0.41	−38
6.	consumers with "correct" post-experiment beliefs	3	−0.22	−21
7.	Balanced Treatment group	3	−0.48	−45
8.	10 percent confidence bound	1	−0.92	−86
9.	90 percent confidence bound	(Ban)	0.05	4
Panel C. Additional distortion				
10.	Excess mass consumers have $v = 7.66$	8	1.22	114
Panel D. Approximate bias with equivalent price metric				
11.	EPM from Appendix Table A.3	1	−0.79	−74

Notes: This table presents welfare results using the TESS data under alternative assumptions. Column 3 divides column 2 by incandescent lightbulb buyers' total perceived consumer surplus from having the incandescent available. Observations are weighted for national representativeness.

top-coded and bottom coded WTPs. Row 4 assumes consumers' self-reported hypothetical WTP, bounded at $+/-100$.[26]

The bottom set of results consider alternatives to Assumption 1. Rows 5 and 6 consider the possibility that not all consumers understood the information treatment, which could cause the CATEs to understate bias. Row 5 increases the base case $B(p)$ estimates by 4 percent to recognize that the 89 percent of consumers who passed the "quiz" on the Treatment Information screen have 4 percent larger ATEs than the Treatment group as a whole. Row 6 analogously increases $B(p)$ by 34 percent, using the result that the 30 percent of consumers with "correct" CFL savings beliefs on the post-experiment survey had 34 percent larger ATE than the Treatment group as a whole.

Row 7 scales down the $B(p)$ function by 7 percent, to reflect the estimate in Table 1 that the Balanced Treatment group had a 7 percent smaller ATE than the Treatment group as a whole. The effects of the Balanced Treatment would be of primary interest if one thought that the Positive Treatment group were more affected by experimenter demand effects, or if one wanted to also incorporate potential imperfect information or inattention related to the disposal and warm-up attributes discussed on the Negative Information screen. As discussed in Section III, the differences used to scale rows 5 and 7 are not statistically significantly different from

[26] While it may seem unsatisfying to need to make these assumptions about censored WTPs, remember that the TESS experiment substantially improves over the standard approach to analyzing the removal of a product from the choice set, which is to assume a logit or otherwise parametric functional form for demand.

zero. Rows 8 and 9 replace the $B(p)$ function with the tenth and ninetieth percent confidence bounds of the CATEs, as graphed in Figure 4.

Equation (4) in Section II showed how the framework can be extended when there are additional distortions other than imperfect information and inattention. In Section IIID, we pointed out that demand curves are much more price-responsive around zero than they are away from zero. Visual inspection of the demand curves suggest that this "excess mass" of valuations in the Treatment group is primarily contained on the interval $w^1 \in [-2, 4]$. Such excess mass of WTP around zero would be unlikely if all consumers fully value a large lifetime cost savings from the CFL. To see this, use a model as in DellaVigna (2009) with $w^1 = \hat{e} + n$, where $\hat{e}$ is perceived total cost savings and n reflects preferences for other nonenergy attributes. If $\hat{e}$ tends to be large and positive, then n must be symmetrically large and negative in a remarkably coincidental way in order to generate a mass of w^1 near zero.[27] Thus, it seems more likely that there is a mass of consumers for whom both $\hat{e}$ and n are close to zero.

One explanation for a large group of consumers with small $\hat{e}$ is that these consumers think that they might break or discard the lightbulbs before the end of their rated lifetimes. This would not be a distortion, as the true social value of lifetime cost savings is also small in this case. A second potential explanation is asymmetric information in rental markets. Additional (unreported) regressions show that renters are no more likely to have WTP in the interval $w^1 \in [-2, 4]$, which seems to rule this out. A third explanation is that this is a behavioral bias not addressed by information provision, such as the Kőszegi and Szeidl (2013) bias toward concentration.[28] If this is true, how would this affect the welfare analysis?

Because the TESS experimental design does not directly identify average marginal bias functions for distortions not addressed by information provision, we provide an illustrative, back-of-the-envelope calculation using a three-step approach. First, we use an excess mass test inspired by Chetty et al. (2011) to show that about 38 percent of the Treatment group is excess mass on $w^1 \in [-2, 4]$ relative to a prediction based on the density on the rest of the demand curve.[29] This excess mass can also be visually approximated on Figure 3 by assessing the additional market share on the specific interval $w^1 \in [-2, 4]$ compared to a prediction based on the

[27] For a stark example, assume away any heterogeneity in $\hat{e}$ that could result from variation in electricity prices and discount rates. If $\hat{e} = \$40$ for all consumers and w is tightly distributed around zero, then n would need to be tightly distributed around $-\$40$. Then, the demand curves in Figure 3 would imply that if $\hat{e}$ were actually \$35 or \$45 instead of \$40, then the CFL market share would move substantially to 0.4 or 0.94, respectively.

[28] We thank one of our referees for drawing our attention to the high elasticity around $p = 0$ and for suggesting that this may be the consequence of some additional bias.

[29] To do this, we index endline WTP intervals by m, denote D_m as the sample-weighted share of Treatment group consumers with endline WTP in interval m, denote Δp_m as the width of interval m (either \$1 or \$2), denote ξ_m as an indicator for interval m, and run the following regression:

$$D_m = \omega_1 \Delta p_m + \omega_2 w^1 \Delta p_m + \sum_{m=[-2,-1)}^{[3,4)} \xi_m + \epsilon_m .$$

Assuming a quadratic approximation to the demand curve outside the interval $w^1 \in [-2, 4]$, the sum of the ξ_m coefficients identify the total excess mass on $w^1 \in [-2, 4]$. (A higher-order approximation is not merited given the limited number of data points, and following Chetty et al. 2011, the interval $w^1 \in [-2, 4]$ was chosen by visual inspection of the demand curve.) The total excess mass $\sum_{m=[-2,-1)}^{[3,4)} \hat{\xi}_m$ represents about 38 percent of the Treatment group.

slope of demand outside that interval. Second, we calculate the share of consumers at each level of baseline WTP w^0 that are part of the excess mass on $w^1 \in [-2, 4]$. Third, we compute the additional bias function $A(p)$ if consumers who are part of the excess mass on $w^1 \in [-2, 4]$ have average true utility $v = 7.66$, which is the Treatment group mean w^1 after reweighting to eliminate the excess mass.

Row 10 of Table 3 presents results. The ban now increases welfare by 114 percent of incandescent lightbulb buyers' total perceived consumer surplus from having the incandescent available. The optimal subsidy increases to \$8, although welfare is only slightly higher than at a subsidy of \$4. This is because, at least under our back-of-the-envelope assumptions, most of the consumers comprising the excess mass also have baseline WTP w^0 close to zero, and they are thus mostly inframarginal to larger subsidies.

Our last calculation does not consider alternative assumptions, but instead builds on our results in Section IIID on the difference between the average marginal bias and the equivalent price metric. Row 11 shows how our welfare conclusions would change if we instead approximated $B(p)$ with the EPM estimates from online Appendix B.B. Because in our data, the EPMs are typically smaller than the true CATEs on WTP, the predicted optimal subsidy is much smaller and the predicted welfare losses from the ban are 80 percent larger than their true values in row 1. This large divergence in welfare estimates underscores the importance of experimental designs that directly identify the average marginal bias.

In summary, the welfare losses from the incandescent lightbulb ban in most scenarios amount to 30–50 percent of incandescent buyers' total perceived consumer surplus from having the incandescent available. In two scenarios, however, the ban is welfare-enhancing.

V. In-Store Experiment

A. *Experimental Design*

Would the effects of information provision be different in a more typical retail setting compared to the TESS platform? To answer this, we partnered with a large home improvement retailer to implement an in-store experiment. Between July 2011 and November 2011, three research assistants (RAs) worked in four large "big box" stores, one in Boston, two in New York, and one in Washington, DC. The RAs approached customers in the stores' "general purpose lighting" areas, which stock incandescents and CFLs that are substitutable for the same uses.[30] They told customers that they were from Harvard University and asked, "Are you interested in answering some quick research questions in exchange for a discount on any lighting you buy today?" Customers who consented were given a brief survey via iPad in which they were asked, among other questions, the most important factors in their lightbulb purchase decision, the wattage and number of bulbs they planned to buy, and the amount of time each day they expected these lightbulbs to be turned on each

[30] This includes standard bulbs used for lamps and overhead room lights. Specialty bulbs like Christmas lights and other decorative bulbs, outdoor floodlights, and lights for vanity mirrors are sold in an adjacent aisle.

day. The survey did not mention electricity costs or discuss any differences between incandescents and CFLs.

In the taxonomy of Harrison and List (2004), this was a "natural field experiment." Participants believed that they were answering a survey, but they did not know that they were in a randomized experiment or that their subsequent purchase behavior would be scrutinized. This experiment has complementary strengths and weaknesses to the TESS experiment: while we observe consumers naturally participating in a standard marketplace, we could not implement the multiple price lists and within-subject design that allow the TESS experiment to identify the average marginal bias. Instead, we randomize information and prices in a standard 2-by-2 design and focus on answering our positive research question about the effects of information on demand.

The iPad randomized customers into information Treatment and Control groups with equal probability. For the Treatment group, the iPad would display the annual energy costs for CFLs versus incandescents, given the customer's estimated daily usage, desired wattage, and desired number of bulbs. The treatment screen also displayed the energy costs and total user costs (energy plus bulbs) for CFLs versus incandescents over the 8,000-hour rated life of a CFL. Online Appendix C presents the information treatment screen. The RAs would interpret and discuss the information with the customer, but they were trained to not advocate for a particular type of bulb and to avoid discussing any other issues unrelated to energy costs, such as mercury content or environmental externalities. The Control group did not receive this informational intervention, and the RAs did not discuss energy costs or compare CFLs and incandescents with these customers.

At the end of the survey and potential informational intervention, the RAs gave customers a coupon in appreciation for their time. The iPad randomized respondents into either the Standard Coupon group, which received a coupon for 10 percent off all lightbulbs purchased, or the Rebate Coupon group, which received the same 10 percent coupon plus a second coupon valid for 30 percent off all CFLs purchased. Thus, the Rebate Coupon group had an additional 20 percent discount on all CFLs. For a consumer buying a typical package of 60-watt bulbs at a cost of \$3.16 per bulb, this maps to an average rebate of \$0.63 per bulb. The coupons had barcodes which were recorded in the retailer's transaction data as the customers submitted them at the register, allowing us to match the iPad data to purchases. We do not observe the possible purchases of the 23 percent of consumers in the iPad whose coupon numbers do not appear in the transaction data; these consumers either purchased lightbulbs without submitting the coupon or did not purchase any lightbulbs.

After giving customers their coupons, the RAs would leave the immediate area in order to avoid any potential external pressure on customers' decisions. The RAs would then record additional visually observable information on the customer, including approximate age, gender, and ethnicity. The RAs also recorded this information for people who refused. Finally, the RA recorded the total duration of the interaction. The difference between Treatment and Control reflects the amount of time spent on the informational intervention. The difference in means (medians) is 3.17 (3.0) minutes, which suggests that Treatment group consumers did engage meaningfully with the information.

Table 4—Descriptive Statistics and Balance for In-Store Experiment

Individual characteristics	Experimental sample mean (1)	Refused − sample difference (2)	Treatment − control difference (3)	Rebate − standard difference (4)
Energy an important factor in purchase decision	0.25		0.009	−0.024
	(0.43)		(0.026)	(0.026)
Expected usage (minutes/day)	333		12.8	2.7
	(280)		(17.0)	(17.0)
Age	43.8	2.3	0.7	−0.3
	(11.4)	(0.6)***	(0.7)	(0.7)
Male	0.66	0.06	0.009	0.003
	(0.47)	(0.03)**	(0.029)	(0.029)
African American	0.16	−0.04	−0.001	−0.008
	(0.37)	(0.02)**	(0.022)	(0.022)
Asian	0.06	0.04	−0.030	0.005
	(0.24)	(0.02)**	(0.014)**	(0.015)
Caucasian	0.66	−0.07	0.037	−0.005
	(0.47)	(0.03)**	(0.029)	(0.029)
Hispanic	0.07	0.06	0.001	0.011
	(0.25)	(0.02)***	(0.015)	(0.015)
Middle Eastern	0.01	0.01	0.002	0.007
	(0.12)	(0.01)	(0.013)	(0.007)
F-test p-value		0.00	0.742	0.896
Purchase decisions				
Purchased any lightbulb	0.77		0.011	0.027
	(0.42)		(0.025)	(0.025)
Purchased substitutable lightbulb	0.73		−0.008	0.011
	(0.44)		(0.027)	(0.027)

Notes: Column 1 presents means of individual characteristics in the in-store experiment sample, with standard deviations in parentheses. Column 2 presents differences in recorded demographic characteristics between those who refused or did not complete the survey and the experimental sample. Column 3 presents differences in means between Treatment and Control groups, while column 4 presents differences in means between the Rebate Coupon and Standard Coupon groups. Columns 2, 3, and 4 have robust standard errors in parentheses.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

B. *Data*

Of the 1,561 people who were approached, 459 refused, while 1,102 began the iPad survey. Of these, 13 broke off after the first question, 2 broke off later, and 1,087 were assigned to a treatment group and given a coupon. Column 1 of Table 4 presents descriptive statistics for the sample of customers who completed the survey and were given a coupon. Column 2 presents differences between the 474 people who refused or did not complete the survey and the 1,087 who completed, using the demographic characteristics recorded for those who refused. People whom the RAs thought were older, male, Asian, and Hispanic were more likely to refuse. Columns 3 and 4 present differences between the information Treatment and Control groups and between the Rebate Coupon and Standard Coupon groups. In 1 of the 18 t-tests, a characteristic is statistically different with 95 percent confidence: we have slightly

Table 5—Effects of In-Store Information Treatment

	(1)	(2)	(3)
1(Treatment)	−0.002	0.004	−0.022
	(0.035)	(0.033)	(0.045)
1(Rebate)	0.094	0.105	0.078
	(0.035)***	(0.033)***	(0.047)*
1(Rebate and Treatment)			0.054
			(0.066)
R^2	0.01	0.16	0.16
Observations	794	793	793
Individual characteristics	No	Yes	Yes

Notes: This table presents estimates of equation (7), a linear probability model with outcome variable 1(Purchased CFL). The dependent variable has mean 0.38. Robust standard errors in parentheses.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

fewer people coded as Asian in the information treatment group. *F*-tests fail to reject that the groups are balanced.

We restrict our regression sample to the set of consumers that purchase a "substitutable lightbulb," by which we mean either a CFL or any incandescent or halogen that can be replaced with a CFL. The bottom panel of Table 4 shows that 77 percent of interview respondents purchased any lightbulb with a coupon, and 73 percent of survey respondents purchased a substitutable lightbulb. While information or rebates theoretically could affect whether or not customers purchase a substitutable lightbulb, *t*-tests show that in practice the percentages are not significantly different between the groups.

C. *Results*

We denote T_i and S_i as indicator variables for whether customer i is in the Treatment and Rebate Coupon groups, respectively. $\mathbf{X}_i$ is the vector of individual-level covariates. We estimate a linear probability model (LPM)[31] with robust standard errors using the following equation:

$$1(\text{Purchase CFL})_i = \tau T_i + \eta S_i + \gamma \mathbf{X}_i + \varepsilon_i. \tag{7}$$

Quantity Effects, Demand Slopes, and the Equivalent Price Metric.—Table 5 presents estimates of equation (7). Column 1 excludes covariates X_i, while column 2 adds them. The estimates are statistically identical, and the point estimates are very similar. The rebate increased CFL market share by about 10 percentage points. Column 3 shows that the interaction between information and rebates is statistically zero. Using column 3, the point estimate of Treatment group demand slope is $\frac{7.8 + 5.4}{\$0.63} \approx$ 21 percentage points per $1.

[31] In typical cases like ours where the true probability model is not known, Angrist and Pischke (2009) advocate for using the LPM instead of an arbitrary nonlinear model such as probit or logit, and we follow their recommendation. In any event, *S* and *T* are indicator variables, and the probit estimates are almost identical.

In column 2, the information treatment increased market share by 0.4 percentage points, which is statistically indistinguishable from zero. The standard errors rule out with 90 percent confidence that the information treatment increased (decreased) market share by more than 5.8 (5.0) percentage points. Using column 3, the point estimates of the information treatment effect are -2.2 percentage points with the Standard Coupon and $-2.2 + 5.4 = 3.2$ percentage points with the Rebate Coupon; both are statistically indistinguishable from zero.

Inelastic demand could cause this statistically small effect on market shares. For example, one might hypothesize that in-store demand would be less elastic because consumers might enter the store having already decided what type of lightbulb to buy, perhaps as instructed by other family members. The fact that in-store demand responds strongly to prices rules out this hypothesis.

To more formally compare information versus price effects, we again calculate the equivalent price metric. Inserting the above coefficients into equation (5), we have the EPM between the Standard Coupon and Rebate Coupon price levels: $EPM[Rebate, Standard] \approx \frac{(0.032 + -0.022)/2}{0.21} \approx \$0.02$. Using the delta method, the 90 percent confidence interval is $[-\$0.24, \$0.28]$. In other words, the standard errors rule out with 90 percent confidence that information provision had more than the effect of a \$0.28 CFL subsidy.

D. *Comparing and Generalizing from the Two Experiments*

We can compare the three main parameters from the in-store experiment (information effects on market share, demand slope, and EPM) to their analogues from the TESS experiment. As Figure 3 illustrates, the effects of information on market share differ substantially by price level. We separately compare the effects near market prices ($p = 0$ in the TESS experiment and with the Standard Coupon in the in-store experiment) and at a small discount ($p = -1$ in TESS and with the \$0.63 average discount of the Rebate Coupon). Near market prices, the effects in the TESS and in-store experiments are statistically different with a p-value of 0.015. At a small discount, however, the 7 percentage point effect at $p = -1$ in TESS is statistically indistinguishable from the effect on Rebate Coupon recipients in the store (p-value $= 0.44$). Point estimates suggest that in-store demand is even more price-responsive than in TESS, although the slopes are statistically indistinguishable (p-value $= 0.15$). Finally, the TESS EPM on $p \in [-1, 0]$ is statistically larger than the in-store EPM on $p \in [Rebate, Standard]$, with a p-value of 0.012.

It is well understood that empirical results can differ across contexts. If the goal is to evaluate a nationwide policy, ideally one would estimate a nationwide parameter using many experiments with consumers and retailers across the country. When attempting to learn from a limited number of experiments, it is particularly useful if they differ on dimensions that could moderate effects out of sample. While we only have two experiments, this pair is relatively useful because they differ markedly on the three key dimensions: consumer populations, choice environments, and treatments.

First, consumer populations differ substantially: the TESS population is nationwide, while the in-store sample is drawn from four stores in three eastern states. There are some differences on observable characteristics, and there are surely differences on unobservables as well. In combination with the substantial treatment effect heterogeneity suggested by Figure 2, these differences could generate different parameter estimates.[32] On this dimension, the TESS estimates are of greater interest because consumers are drawn from a wider geographic area and are weighted for national representativeness on observables.

Second, the choice environments also differ markedly. The TESS experiment has a deliberately simple and controlled choice environment with a small choice set and limited additional stimuli, while the in-store environment includes hundreds of different lightbulb packages and many other stimuli and purchasing needs competing for attention. These factors could make it more difficult for the in-store Treatment group consumers to internalize, recall, and apply the information when they actually choose a package. Furthermore, like most home improvement stores, the stores we worked in have displays in lightbulb aisles that provide information on different lightbulb technologies, including electricity use. If this existing information fully informed the Control group, incremental information could have no effect. If this is the case, the treatment effects are still the relevant parameters for policy analysis in the in-store environment: if existing information provision mechanisms are fully effective, then there is no remaining imperfect information and inattention to justify subsidies and standards.

The choice environments in both experiments are of interest: home improvement retailers are the most common retail channel through which households buy lightbulbs (DOE 2010), and our partner alone sells upward of 50 million lightbulb packages each year, a nontrivial share of national sales. Notwithstanding, more than one-half of lightbulbs are sold at grocery stores, drugstores, and other retail channels that typically have less in-store energy cost information, and an increasing number of consumers buy online. The TESS Control group's informational environment may be more representative of these channels.

Third, the treatments mechanically differ: the TESS treatments were online with recorded audio and graphs, while the in-store treatments were presented by a live person without graphs. While we assume for policy analysis that the information treatments were pure nudges and would thus have identical effects for a given consumer in a given choice environment, this could perhaps be violated in either of the experiments. For example, in-store information effects could be smaller if the in-store treatment were somehow more difficult to understand, or if consumers chose not to process that information in the absence of a quiz. However, our RAs for the in-store experiment report that most consumers did seem to engage with and understand the information.

The in-store experiment does not allow us to directly estimate the effects of information on WTP. We showed empirically in Section IIID that there is no clear relationship between the EPM and the CATE on WTP, and we showed theoretically in Section II that the average marginal bias can be large even when the EPM is zero. Furthermore, online Appendix D.B calibrates an example with demand parameters

[32] Remarkably, the in-store data suggest that incandescent buyers do not buy more bulbs per trip than CFL buyers. Because incandescents have much shorter lives than CFLs, people who prefer incandescents will thus need to purchase bulbs more often and will appear with higher probability in the in-store sample than in the TESS sample. The TESS CATEs are smaller for consumers that have lower baseline WTP (and thus more strongly prefer incandescents), and this could partially explain the differences between the two experiments.

similar to the in-store experiment and shows that the CATE on WTP could easily be as large as \$2 when the effect on market shares is zero, even with fairly restrictive assumptions on the distribution of bias across consumers. We thus use the in-store results primarily to answer our positive research question about the effects of information on demand. Theoretically, we cannot reject the null hypothesis that the CATEs from the in-store experiment are equivalent to the CATEs in the TESS experiment. Notwithstanding, the fact that the EPM and the information effects on market shares are generally smaller than in the TESS experiment is certainly consistent with the idea that the CATE on WTP would also be smaller. In this case, the in-store experiment results would strengthen the qualitative conclusion that information and inattention do not justify the incandescent lightbulb ban.

VI. Conclusion

While imperfect information and inattention are commonly used to justify energy efficiency policies, the arguments are often qualitative, without formal welfare analysis and relevant empirical tests. In this paper, we derived welfare effects of subsidies and standards in terms of two sufficient statistics, the baseline demand curve and the average marginal bias, and carried out a randomized experiment specifically designed to identify the two statistics.

Our main results suggest that moderate CFL subsidies may be optimal, but that imperfect information and inattention alone cannot justify a ban on incandescents. The approach requires the assumption that the effects of information on WTP equal the average marginal bias. While we carefully designed our treatments to make this as plausible as possible, we still view this assumption as only an approximation, and we explored plausible alternative assumptions in the welfare analysis. We showed how the analysis can be easily extended to incorporate externalities and other distortions not addressed by information provision, and we gave an example of this in studying what may be an excess mass of consumers with valuations near zero. In this alternative scenario, our model suggests that a ban does increase welfare.

To begin to address the question of whether the TESS results generalize to other populations and choice environments, we implemented a complementary in-store experiment. While the standard 2-by-2 design in this experiment does not allow a good approximation to the average marginal bias, the smaller effects on market shares might reinforce the qualitative conclusion that informational and attentional biases do not justify a ban.

The paper makes several important contributions. First, while it was plausible to believe that information provision could substantially affect the lightbulb market, both of our experiments show that large shares of consumers still prefer incandescents even after being powerfully informed. Second, while incandescent lightbulb bans have become important features of energy policy in many countries, our results suggest that more careful thought is needed about why these might increase welfare. Third, our basic approach is generally useful for studying behaviorally motivated policies outside of the lightbulb market. We show that approximations like the EPM that require homogeneity can give biased empirical estimates, meaning that precisely identifying the necessary statistics for behavioral welfare analysis may require more complex empirical designs than had previously been anticipated.

REFERENCES

Abaluck, Jason, and Jonathan Gruber. 2011. "Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program." *American Economic Review* 101 (4): 1180–1210.

ABC News. 2011. "Rand Paul's Toilet Tirade." http://abcnews.go.com/blogs/politics/2011/03/randpauls-toilet-tirade/.

Allcott, Hunt. 2013. "The Welfare Effects of Misperceived Product Costs: Data and Calibrations from the Automobile Market." *American Economic Journal: Economic Policy* 5 (3): 30–66.

Allcott, Hunt. Forthcoming. "Paternalism and Energy Efficiency: An Overview." *Annual Review of Economics*.

Allcott, Hunt, Sendhil Mullainathan, and Dmitry Taubinsky. 2014. "Energy Policy with Externalities and Internalities." *Journal of Public Economics* 112 (April): 72–88.

Allcott, Hunt, and Dmitry Taubinsky. 2015. "Evaluating Behaviorally Motivated Policy: Experimental Evidence from the Lightbulb Market: Dataset." *American Economic Review.* http://dx.doi.org/10.1257/aer.20131564.

Allcott, Hunt, and Nathan Wozny. 2014. "Gasoline Prices, Fuel Economy, and the Energy Paradox." *Review of Economics and Statistics* 96 (5): 779–95.

Anderson, C. Dennis, and John D. Claxton. 1982. "Barriers to Consumer Choice of Energy Efficient Products." *Journal of Consumer Research* 9 (2): 163–70.

Andreoni, James, and Charles Sprenger. 2012. "Estimating Time Preferences from Convex Budgets." *American Economic Review* 102 (7): 3333–56.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Attari, Shahzeen Z., Michael L. DeKay, Cliff I. Davidson, and Wändi Bruine de Bruin. 2010. "Public Perceptions of Energy Consumption and Savings." *Proceedings of the National Academy of Sciences* 107 (37): 16054–59.

Australian Department of the Environment, Water, Heritage, and the Arts (DEWHA). 2008. "Regulatory Impact Statement: Proposal to Phase-Out Inefficient Incandescent Light Bulbs." http://www.energyrating.gov.au/wp-content/uploads/Energy_Rating_Documents/Library/Lighting/Incandescent_Lamps/200808-ris-phaseout.pdf.

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein. 2015. "Behavioral Hazard in Health Insurance." National Bureau of Economic Research Working Paper 18468.

Bernheim, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *Quarterly Journal of Economics* 124 (1): 51–104.

Blumstein, Carl, Betsy Krieg, Lee Schipper, and Carl York. 1980. "Overcoming Social and Institutional Barriers to Energy Conservation." *Energy* 5 (4): 355–71.

Bollinger, Bryan, Phillip Leslie, and Alan Sorensen. 2011. "Calorie Posting in Chain Restaurants." *American Economic Journal: Economic Policy* 3 (1): 91–128.

Borenstein, Severin, and Lucas Davis. 2012. "The Equity and Efficiency of Two-Part Tariffs in US Natural Gas Markets." *Journal of Law and Economics* 55 (1): 75–128.

Bureau of Labor Statistics (BLS). 2013. "Consumer Expenditure Survey: Expenditure Tables." http://www.bls.gov/cex/csxstnd.htm.

Busse, Meghan R., Christopher R. Knittel, and Florian Zettelmeyer. 2013. "Are Consumers Myopic? Evidence from New and Used Car Purchases." *American Economic Review* 103 (1): 220–56.

Caplin, Andrew, and Mark Dean. 2014. "Revealed Preference, Rational Inattention, and Costly Information Acquisition." National Bureau of Economic Research Working Paper 19876.

Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri. 2011. "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *Quarterly Journal of Economics* 126 (2): 749–804.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99 (4): 1145–77.

Davis, Lucas W. 2012. "Evaluating the Slow Adoption of Energy Efficient Investments: Are Renters Less Likely to Have Energy Efficient Appliances?" In *The Design and Implementation of US Climate Policy*, edited by Don Fullerton and Catherine Wolfram, 301–18. Chicago: University of Chicago Press.

Davis, Lucas W., and Erich Muehlegger. 2010. "Do Americans Consume too Little Natural Gas? An Empirical Test of Marginal Cost Pricing." *RAND Journal of Economics* 41 (4): 791–810.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47 (2): 315–72.

Deutsch, Matthias. 2010a. "Life Cycle Cost Disclosure, Consumer Behavior, and Business Implications: Evidence from an Online Field Experiment." *Journal of Industrial Ecology* 14 (1): 103–20.

Deutsch, Matthias. 2010b. "The Effect of Life-Cycle Cost Disclosure on Consumer Behavior: Evidence from a Field Experiment with Cooling Appliances." *Energy Efficiency* 3 (4): 303–15.

Diamond, Peter A. 1973. "Consumption Externalities and Imperfect Corrective Pricing." *Bell Journal of Economics and Management Science* 4 (2): 526–38.

Dubin, Jeffrey A., and Daniel L. McFadden. 1984. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption." *Econometrica* 52 (2): 345–62.

Energy Star. 2013. "ENERGY STAR Unit Shipment and Market Penetration Report Calendar Year 2012 Summary." http://www.energystar.gov/ia/partners/downloads/unit_shipment_data/2012_USD_Summary_Report.pdf.

Falk, Armin, and Florian Zimmermann. 2013. "A Taste for Consistency and Survey Response Behavior." *CESifo Economic Studies* 59 (1): 181–93.

Fong, Christina M., and Erzo F. P. Luttmer. 2009. "What Determines Giving to Hurricane Katrina Victims? Experimental Evidence on Racial Group Loyalty." *American Economic Journal: Applied Economics* 1 (2): 64–87.

Formisano, Bob. 2008. "2007 Energy Bill: Are They Phasing Out or Making Incandescent Bulbs Illegal?" http://homerepair.about.com/od/electricalrepair/ss/2007 energybill.htm.

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

Gabaix, Xavier, and David Laibson. 2006. "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets." *Quarterly Journal of Economics* 121 (2): 505–40.

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg. 2006. "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model." *American Economic Review* 96 (4): 1043–68.

Gayer, Ted. 2011. "A Better Approach to Environmental Regulation: Getting the Costs and Benefits Right." Hamilton Project Policy Brief 2011-06.

Gillingham, Kenneth, Matthew Harding, and David Rapson. 2012. "Split Incentives in Residential Energy Consumption." *Energy Journal* 33 (2): 37–62.

Gillingham, Kenneth T., and Karen L. Palmer. 2013. "Bridging the Energy Efficiency Gap: Policy Insights from Economic Theory and Empirical Evidence." Resources for the Future Discussion Paper 13-02.

Goldberg, Pinelopi Koujianou. 1998. "The Effects of the Corporate Average Fuel Economy Standards in the US." *Journal of Industrial Economics* 46 (1): 1–33.

Gruber, Jonathan, and Botond Kőszegi. 2004. "Tax Incidence when Individuals Are Time-Inconsistent: The Case of Cigarette Excise Taxes." *Journal of Public Economics* 88 (9–10): 1959–87.

Harberger, Arnold C. 1964a. "The Measurement of Waste." *American Economic Review* 54 (3): 58–76.

Harberger, Arnold C. 1964b. "Taxation, Resource Allocation, and Welfare." In *The Role of Direct and Indirect Taxes in the Federal Revenue System*, edited by John F. Due, 25–70. Princeton, NJ: Princeton University Press.

Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature* 42 (4): 1009–55.

Hassett, Kevin A., and Gilbert E. Metcalf. 1995. "Energy Tax Credits and Residential Conservation Investment: Evidence from Panel Data." *Journal of Public Economics* 57 (2): 201–17.

Hausman, Jerry A. 1979. "Individual Discount Rates and the Purchase and Utilization of Energy-Using Durables." *Bell Journal of Economics* 10 (1): 33–54.

Heidhues, Paul, Botond Kőszegi, and Takeshi Murooka. 2014. "Inferior Products and Profitable Deception." Unpublished.

Heiss, Florian, Daniel McFadden, and Joachim Winter. 2007. "Mind the Gap! Consumer Perceptions and Choices of Medicare Part D Prescription Drug Plans." National Bureau of Economic Research Working Paper 13627.

Herberich, David H., John A. List, and Michael K. Price. 2011. "How Many Economists Does It Take to Change a Light Bulb? A Natural Field Experiment on Technology Adoption." Unpublished.

Heutel, Garth. Forthcoming. "Optimal Policy Instruments for Externality-Producing Durable Goods under Time Inconsistency." *Journal of Environmental Economics and Management*.

Hossain, Tanjim, and John Morgan. 2006. "...Plus Shipping and Handling: Revenue (Non)Equivalence in Field Experiments on eBay." *Advances in Economic Analysis & Policy* 5 (2).

Houde, Sébastien. 2012. "How Consumers Respond to Product Certification and the Value of Energy Information." National Bureau of Economic Research Working Paper 20019.

Jaffe, Adam B., and Robert N. Stavins. 1994. "The Energy Paradox and the Diffusion of Conservation Technology." *Resource and Energy Economics* 16 (2): 91–122.

Jessoe, Katrina, and David Rapson. 2014. "Knowledge Is (Less) Power: Experimental Evidence from Residential Energy Use." *American Economic Review* 104 (4): 1417–38.

Kallbekken, Steffen, Håkon Sælen, and Erlend A. T. Hermansen. 2013. "Bridging the Energy Efficiency Gap: A Field Experiment on Lifetime Energy Costs and Household Appliances." *Journal of Consumer Policy* 36 (1): 1–16.

Kőszegi, Botond, and Adam Szeidl. 2013. "A Model of Focusing in Economic Choice." *Quarterly Journal of Economics* 128 (1): 53–104.

Mullainathan, Sendhil, Joshua Schwartzstein, and William J. Congdon. 2012. "A Reduced-Form Approach to Behavioral Public Finance." *Annual Review of Economics* 4: 511–40.

National Highway Traffic Safety Administration (NHTSA). 2010. "Final Regulatory Impact Analysis: Corporate Average Fuel Economy for MY 2012–MY 2016 Passenger Cars and Light Trucks." Office of Regulatory Analysis and Evaluation, National Center for Statistics and Analysis (March).

Natural Resources Defense Council (NRDC). 2011. "Better Light Bulbs Equal Consumer Savings in Every State." www.nrdc.org/policy.

Newell, Richard G., and Juha Siikamäki. 2013. "Nudging Energy Efficiency Behavior: The Role of Information Labels." Resources for the Future Discussion Paper 13-17.

O'Donoghue, Ted, and Matthew Rabin. 2006. "Optimal Sin Taxes." *Journal of Public Economics* 90 (10–11): 1825–49.

Parry, Ian W. H., David Evans, and Wallace E. Oates. 2010. "Are Energy Efficiency Standards Justified?" Resources for the Future Discussion Paper 10-59.

Parry, Ian, Carolyn Fischer, and Winston Harrington. 2007. "Do Market Failures Justify Tightening Corporate Average Fuel Economy (CAFE) Standards?" *Energy Journal* 28 (4): 1–30.

Rabin, Matthew, and Georg Weizsacker. 2009. "Narrow Bracketing and Dominated Choices." *American Economic Review* 99 (4): 1508–43.

Sallee, James M. 2014. "Rational Inattention and Energy Efficiency." *Journal of Law and Economics* 57 (3): 781–820.

Sallee, James M., Sarah E. West, and Wei Fan. 2015. "Do Consumers Recognize the Value of Fuel Economy? Evidence from Used Car Prices and Gasoline Price Fluctuations." http://home.uchicago.edu/~sallee/swf_140127.pdf (accessed May 29, 2015).

Sanstad, Alan H., and Richard B. Howarth. 1994. "'Normal' Markets, Market Imperfections, and Energy Efficiency." *Energy Policy* 22 (10): 811–18.

Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3): 665–90.

Snyder, Mark. 1974. "Self-Monitoring of Expressive Behavior." *Journal of Personality and Social Psychology* 30 (4): 526–37.

Spinnewijn, Johannes. 2014. "Heterogeneity, Demand for Insurance and Adverse Selection." http://personal.lse.ac.uk/spinnewi/perceptions_welfare.pdf (accessed May 29, 2015).

Sylvania. 2012. "5th Annual Sylvania Socket Survey." https://www.sylvania.com/en-us/tools-and-resources/surveys/Pages/socket-survey.aspx.

US Department of Energy (DOE). 2009. "Technical Support Document: Impacts on the Nation of the Energy Independence and Security Act of 2007." http://www1.eere.energy.gov/buildings/appliance_standards/pdfs/en_masse_tsd_march_2009.pdf.

US Department of Energy (DOE). 2010. "Energy Star CFL Market Profile: Data Trends and Market Insights." http://www.energystar.gov/ia/products/downloads/CFL_Market_Profile_2010.pdf.

US Department of Energy (DOE). 2012. "2010 U.S. Lighting Market Characterization." http://apps1.eere.energy.gov/buildings/publications/pdfs/ssl/2010-lmc-final-jan-2012.pdf.

US Department of Energy (DOE). 2014. "Electric Power Monthly." http://www.eia.gov/electricity/monthly/epm_table_grapher.cfm?t=epmt_5_3.

US Environmental Protection Agency (EPA). 2011. "Energy Independence and Security Act of 2007 (EISA): Frequently Asked Questions." http://www.energystar.gov/ia/products/lighting/cfls/downloads/EISA_Backgrounder_FINAL_4-11_EPA.pdf.

US Environmental Protection Agency (EPA) and US Department of Transportation (DOT). 2010. "Light-Duty Vehicle Greenhouse Gas Emission Standards and Corporate Average Fuel Economy Standards; Final Rule." *Federal Register* 75 (88): 25324–25728.

Ward, David O., Christopher D. Clark, Kimberly L. Jensen, Steven T. Yen, and Clifford S. Russell. 2011. "Factors Influencing Willingness-to-Pay for the Energy STAR Label." *Energy Policy* 39 (3): 1450–58.