Regulating Internalities

Regulating Internalities

Hunt Allcott and Cass R. Sunstein
February 2015

Abstract

This paper offers a framework for regulating internalities. Using a simple economic model, we provide four principles for designing and evaluating behaviorally-motivated policy. We then outline rules for determining which contexts reliably reflect true preferences and discuss empirical strategies for measuring internalities. As a case study, we focus on energy efficiency policy, including Corporate Average Fuel Economy (CAFE) standards and appliance and lighting energy efficiency standards.

Hunt Allcott is an Assistant Professor at the New York University Department of Economics, 19 West 4th St., New York, NY 10012. Email: hunt.allcott@nyu.edu. Cass R. Sunstein is the Robert Walmsley University Professor at Harvard University.

<u>Introduction</u>

It is well-established that governments can increase welfare by providing public goods and addressing market failures such as externalities, asymmetric information, and market power. In practice, governments also regulate to address "internalities" – that is, costs we impose on ourselves by taking actions that are not in our own best interest. An internality might arise when agents act to obtain short-term benefits, or to avoid short-term costs, even though that action produces long-term net costs for those very agents. When this happens, there may be a "behavioral market failure," adding to the standard catalogue of justifications for regulatory action.

The U.S. has consumer protection laws, seatbelt laws, retirement savings incentives, cigarette and alcohol taxes, bans on addictive drugs, prescription requirements for many medicines, food and drug laws, product safety and quality standards, and a series of other "hard" mandates. Requirements of this kind have multiple justifications, but they are sometimes defended on the paternalistic ground that they improve people's welfare by taxing, subsidizing, or restricting their own choices (Conly 2013). Governments also use "soft" or "libertarian" paternalism, such as default rules, disclosure mandates, graphic health warnings, and nutrition labels, which are designed to "nudge" people toward better decisions without changing prices or restricting choice (Thaler and Sunstein 2008). The goal of such approaches is to make people better off by their own lights (ibid.), but they nonetheless have a paternalistic dimension insofar as they steer people in particular directions.

To take an example of an approach that is probably difficult to defend without resort to paternalism, and that plausibly involves internalities, it is unlawful in the United States and many other industrialized countries for employers to allow employees to subject themselves to certain occupational risks. It is possible, and central to our argument, that many workers would choose to run risks (say, involving carcinogenic substances) at Time 1 that would cause them serious or fatal health problems at Time 2. Although efficiency and distributive justifications are sometimes offered, many workplace health and safety regulations can be understood as paternalistic in the sense that they restrict employees' choices in order to protect the long-term interests of employees themselves. An obvious alternative to occupational safety and health mandates would be to rely on competition (alongside common law safeguards against tortious behavior), with the idea that firms would provide the information necessary for workers to make knowledgeable decisions. Grossman and Hart (1980) and others show theoretically how information costs can prevent this from happening. A second alternative would be to require all employers to disclose workplace risks and to let employees choose for themselves; more aggressive nudges, such as graphic warnings and default rules, could also be imagined.

A policymaker's decision about whether or not to ban workplace risks or instead to adopt a choice-preserving approach requires an assessment of the benefits and costs of centralized decisionmaking. (We bracket the philosophical issues; see Conly (2013) and Rebonato (2011).) Delegating decisions to a regulatory agency eliminates the need for employees to assess health and safety

risks from various sources and also the need to require workplaces to label various products and activities. On the other hand, the regulatory agency might have its own biases and make mistakes. Furthermore, employees have heterogeneous situations and tastes, so even under the socially-efficient regulation, some workers will be net losers, in the sense that they will lose more in salary and other benefits than they will receive from safer workplaces.

This paper describes an approach to regulating internalities that sets and evaluates regulations on the basis of individuals' own welfare, and to the extent feasible using the individual's own preferences, instead of imposing the regulator's preferences on individuals. (We put to one side some philosophical issues about the proper conception of welfare, see Adler (2011).) The key to this approach is to focus regulation on situations when an individual's choices are inconsistent across contexts, but principled analysis can be used to determine which context more reliably reflects "true preferences," or the preferences that people would show under ideal decisionmaking conditions. As examples, Kling *et al.* (2012) show that people are more likely to choose a lower-cost health insurance plan when given simplified comparison information; Carroll *et al.* (2009) show that more people enroll in 401k savings plans when making active choices instead of passive "opt-in" choices; and Hossein and Morgan (2006) show that consumers are less likely to buy a product when more of the cost is included in the base price instead of "shrouded" as part of shipping and handling charges.

In all three examples, people's choices differ between two contexts (informed vs. uninformed, active vs. passive choice, clearly-presented vs. shrouded costs), and the first of the two contexts more plausibly reflects true preferences. In such situations, there might be scope for policy intervention, and welfare analysis would be conducted using only true preferences. The workplace safety example is more controversial, but the motivation for regulation should be similar. Ideally, regulators are making the decision that most workers would make for themselves if they had the time to become fully informed and if they considered the full set of benefits and costs over their lifetimes (acknowledging the existence of a range of discount rates).

In our view, the interesting questions related to paternalistic regulation are not whether consumers ever make inconsistent decisions (our major focus here) or are always perfectly informed. The above papers and many others reviewed by DellaVigna (2009) and others demonstrate that there are *some* situations in which we make predictably inconsistent choices, and asymmetric or imperfect information is a well-established market failure. The question of whether governments should *consider* paternalistic regulation is also largely settled within the utilitarian perspective (for another perspective, see Waldron 2014): if and when there are internalities, it is mathematically true that policy intervention can increase welfare.[1] We think that many of the most interesting questions in this domain relate to the extensive margin, instrument choice, and the intensive margin. In other words, in what settings should the

---

[1] See, among many others, O'Donoghue and Rabin (2006), Mullainathan, Schwartzstein, and Congdon (2012), or our simple model below.

government intervene at all? If intervening, what policy instrument should be used - a tax, a ban, a "nudge," or some other policy? How stringent should the policy be? And crucially, how do we use data to infer "true preferences" in a principled way?

This paper offers a framework for regulating internalities, drawing on an extensive body of work by many researchers. Using a simple economic model, we provide four principles for designing and evaluating behaviorally-motivated policy. We then provide rules for determining which contexts reliably reflect true preferences and discuss strategies for measuring internalities.

Throughout the paper, we use energy efficiency policy as a case study, focusing on Corporate Average Fuel Economy (CAFE) standards and appliance and lighting energy efficiency standards. (It is important to see that the same present-future tradeoff holds for many attributes of durable goods or capital investments; you may pay more for a larger car upfront, but enjoy it over many years of ownership.) On the basis of our four principles, we discuss legislative changes that could improve energy efficiency policy. Because recent empirical work raises the possibility that internalities in the vehicle market may not be as large as previously hypothesized, we explore approaches for future CAFE standard regulatory impact analyses that might involve less scope for paternalism.

Background: The Behavioral Motivation for Energy Efficiency Policy

There are three main categories of market failures that might motivate energy efficiency policies: externalities (such as emissions of particulate matter and greenhouse gases); "inter-consumer" distortions such as asymmetric information between landlords and tenants; and (ex-post) consumer mistakes due to imperfect information and internalities. We could imagine an approach that is solely focused on externalities and other standard market failures, but three factors make the latter category *theoretically* plausible. First, energy efficiency often involves a present-future tradeoff, where people may have to reduce other consumption now to pay for better home insulation or a hybrid car that saves money in the future. Second, energy efficiency is only one of many product attributes, so people might be inattentive to energy costs when choosing between products. Third, some numerical calculation is typically required, and many consumers find this complicated (Larrick and Soll 2008, Turrentine and Kurani 2007).

The CAFE regulatory impact analyses (RIAs) use engineering models of different fuel economy technologies, such as hybrid powertrains and redesigned engines. In essence, these models predict firms' marginal cost curves for increasing fuel economy (holding constant other attributes such as horsepower). The RIAs then compare these costs to the present discounted value (PDV) of private benefits (largely fuel savings, but also increased driving and less time spent refueling) and net effects on externalities (including lower carbon and local air pollutant emissions, more congestion, and less dependence on foreign oil). Other factors such as effects on safety must also be considered.

Consider the NHTSA (2012) RIA for the light duty CAFE standards for 2011-2025. Projected social benefits far outweigh projected social costs ($629 billion vs. $153 billion). The net externality reductions are valued at $61 billion, while the net *private* benefits (i.e. net social benefits excluding externalities) are $475 billion.[2] Of course, positive net private benefits could arise only in the presence of some market distortion.[3] Drawing directly on behavioral research, the RIA points to "phenomena observed in the field of behavioral economics, including loss aversion, inadequate consumer attention to long-term savings, or a lack of salience of relevant benefits (such as fuel savings, or time savings associated with refueling) to consumers at the time they make purchasing decisions." It adds, "Both theoretical and empirical research suggests that many consumers are unwilling to make energy-efficient investments even when those investments appear to pay off in the relatively short-term. This research is in line with related findings that consumers may undervalue benefits or costs that are less salient, or that they will realize only in the future."

The 2009 technical support document for appliance and lighting energy efficiency standards promulgated under the Energy Independence and Security Act of 2007 (EISA) similarly finds large net private cost savings: $27 to $64 billion in net savings over 30 years (DOE 2009). As with CAFE standards, net private savings are much more important than externalities: summing across all durable good energy efficiency standards promulgated under the EISA, the net private cost savings outweigh the value of carbon externality reductions by 34 to 194 percent. While DOE (2009) is not as explicit in arguing that consumers misoptimize, the RIA for a parallel lighting efficiency standard in Australia is explicit (DEHWA 2008, page vii):

> [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 ...

Energy efficiency standards are thus an important case study of behaviorally-motivated regulation. There are many such standards, but the usual pattern is that they involve large gross costs, and

---

[2] These figures are from Table 13, for NHTSA's "preferred alternative," assuming a three percent discount rate and using the 2010 baseline fleet.

[3] It is sometimes argued that the fact that *gross* private benefits (without netting out increases in vehicle costs) are much larger than externality benefits is evidence that the CAFE RIA assumes internalities. This is not correct. Any CAFE cost-benefit analysis (even assuming zero non-externality market failure) will find that private benefits are a large share of gross benefits, because the policy primarily saves gasoline, and private costs are a large share of the assessed social costs of gasoline. What is true is that some non-externality market failure is required for *net* private benefits (excluding externality benefits) to be positive. Net private benefits of any binding CAFE standard will be negative in the absence of a non-externality market failure.

internality reduction represents a large share of net benefits. As Fisher, Harrington, and Parry (2006) and others show in the context of CAFE, the total costs can exceed the total benefits unless internality reduction is taken into account.

A Model of Optimal Policy under Inconsistent Choice

We present a simplified version of the "reduced form model of behavioral public finance" from Chetty (2015), Mullainathan, Schwartzstein, and Congdon (2012), and Allcott and Taubinsky (2015). Individuals decide whether or not to take an action, such as whether to open a retirement account, quit smoking, get preventive healthcare, or buy a higher fuel economy car. The action has equilibrium price p, and the perfectly-competitive supply curve is S(p).

Consumers incur true utility v when taking the action. In the standard economic model, consumers take the action if v>p. The "behavioral" model generalizes the standard model by allowing consumers to take the action if d>p, where d=v-b is "decision utility" and b is some bias which could be zero (as in the standard model), positive, or negative. Because the model nests the standard model as a special case, it does not assume that consumers are biased; this is an empirical question. In the language of Kahneman *et al.* (1997), v represents "experienced utility" and d represents "decision utility."

The "reduced form" bias parameter b could reflect different specific inconsistencies in different contexts. Under present bias (Laibson 1997), individuals might decide today to quit smoking tomorrow, but when tomorrow arrives they might not quit smoking. Under inattention (Gabaix 2014, Gabaix and Laibson 2006), individuals might buy an energy efficient lightbulb if energy costs and prices are presented together, but they might ignore energy costs and buy the cheaper energy inefficient bulb if energy costs are not salient. Under biased beliefs or otherwise imperfect information, individuals might not switch insurance plans, whereas they might switch health plans when informed about another lower-cost plan.

Figure 1 illustrates the model. $D_M(p)$ is market demand – it reflects the heterogeneity in decision utility d across consumers. Imagine that all consumers have homogenous bias $b_1 > 0$ against the action, so the demand curve if consumers were unbiased would be $D_U(p)$, and it is $D_U(p)$ that reflects experienced utility. Bias b_1 is quantified in dollars as the vertical difference between D_M and D_U. The market equilibrium is at point a, whereas in the efficient equilibrium, all consumers to the left of points f and k would take the action.

While consumer bias is not a traditional market failure, the source of inefficiency is clear: consumers between points a and k on the market demand curve don't take the action even though their true utility exceeds the cost. There is thus a distortion with welfare loss equal to triangle aef. The role of government in this model is also clear: if the policymaker subsidizes the action by amount $s^*=b_1$, the equilibrium moves from point a to point k, and the social optimum is achieved. This government

Figure 1: Graphical Model of Internalities
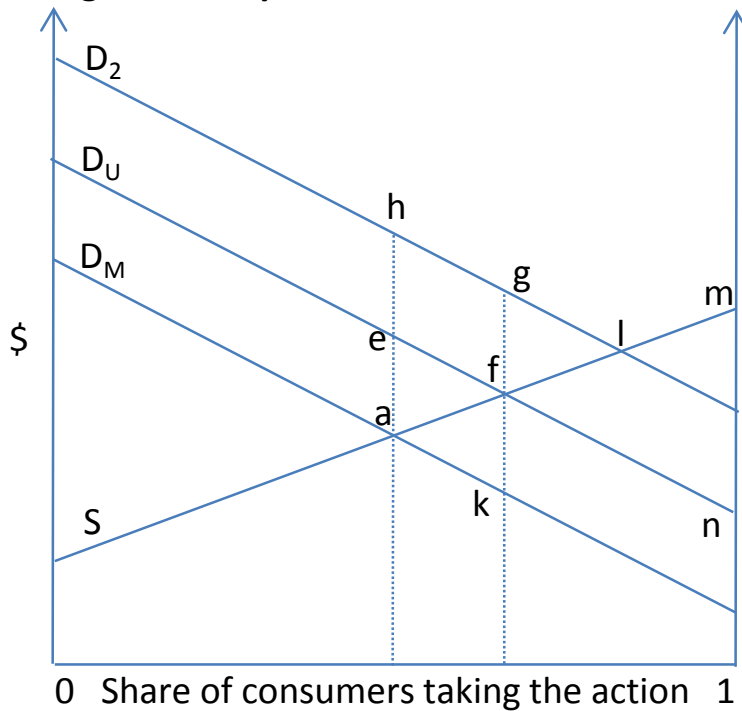
intervention has eliminated the distortion, increasing welfare by amount aef. Notice here the direct analogy between internalities and externalities: if we reinterpret b_1 as an externality, this model simply restates the standard Pigouvian model of externality taxation. It is thus useful to think of internalities as "externalities that individuals impose on themselves."

Principles of Behaviorally-Informed Regulation

The analogy between internalities and externalities is useful because a rich set of theoretical results on optimal externality regulation translate to internalities. Here we discuss four principles that follow from our simple model and use them to comment on major energy efficiency policies.

Principle 1: Taxes and subsidies are more allocatively efficient than mandates or bans

A well-established literature compares price vs. quantity regulation to address externalities; see Hepburn (2006) for a review. In our model of internalities, imagine that the policymaker has two instruments available: a subsidy of amount s and a mandate that all consumers take the action. Which instrument should the policymaker use? Per above, the optimal subsidy generates welfare gain aef. Similarly, the mandate also increases welfare for consumers between a and k on the market demand curve. However, for consumers to the right of k (those with true utility $v<p$), the mandate decreases welfare by forcing them to take an action that they value at less than cost. Thus, the net welfare effects of a ban are aef minus fmn. Except for the special case where *all* consumers have $v>p$ and an infinite subsidy is thus optimal, a mandate is thus strictly worse than the optimal subsidy. Intuitively, since an infinite subsidy achieves the same outcome as a ban, a non-infinite subsidy is preferred to a ban if a non-infinite subsidy is preferred to an infinite subsidy. In this drawing, a mandate is also worse than no policy at all, because aef is smaller than fmn.

As we discuss below, the optimal internality subsidy (or tax) with heterogeneous bias is a "compromise" that is too large for some and too small for others. But this feature does not affect the argument that in this model, a price instrument is more efficient than a mandate or ban. Note that many externality taxes also have this same problem: for example, Knittel and Sandler (2013) highlight that uniform gasoline taxes are also an imperfect "compromise" due to the significant heterogeneity across vehicles in local air pollutant emissions.

Of course, there are many additional considerations; here we highlight four. First, if the price policy is implemented as a tax, we have the well-established "double dividend" from corrective taxation (Bovenberg and Goulder 1996, Parry 1995): aside from correcting distorted choices, the tax also earns revenues, which can be used to reduce distortionary income taxes. Subsidies have the opposite property: they must be funded somehow, likely through distortionary taxation. Second, bans may be easier to implement than taxes – in the workplace safety example, it may be easier to disallow particular activities than to set up an infrastructure for assessing corrective taxes from all firms. In some cases, however, taxes may be easier to implement, because bans may require a complex regulatory framework. Third, the tax salience results of Chetty, Looney, and Kroft (2009) and Finkelstein (2009) highlight that corrective taxes are more effective if included in the posted price, not added on later at the cash register or implemented through individual taxes paid on April 15th.

A preference for price instruments instead of mandates or bans has direct implications for energy efficiency policy. Federal minimum standards ban energy inefficient appliance models. Given that these standards are mandated by existing law, we are not criticizing the fact that they are being implemented. Moreover, the benefits of bans might significantly exceed the costs. However, a different law that would tax energy inefficient appliances at the level of the average marginal distortion (including internalities, uninternalized externalities, and any other market failures) would be more allocatively efficient. Intuitively, this is because there are some people who use air conditioners infrequently or face low electricity prices and thus would optimally own an energy inefficient air conditioner that is currently prohibited. Allcott and Taubinsky (2015) quantify this point in the lightbulb market. For the consumers in their randomized experiments, a moderate tax on energy inefficient lightbulbs might increase welfare by reducing internalities, but some consumers still strongly prefer traditional incandescent lightbulbs even when given clear information about their energy costs, and banning them from buying incandescents reduces welfare. (Note, however, that there is a question whether their ex ante preference predicts their ex post welfare (Kahneman and Thaler 2006).)

By our Principle 1, minimum energy efficiency standards are thus not preferred. On the other hand, many other energy efficiency policies are not bans: federal and state governments often subsidize energy efficiency, and CAFE standards, based on fleet-wide averages, are a quantity policy that imposes a shadow price on low fuel economy vehicles but does not ban them (and in this sense operate a lot like a price incentive).

Principle 2: Target the distortion

Now allow bias to be heterogeneous. The optimal subsidy level is the average marginal bias – i.e. the average bias of consumers that are on the margin at the subsidized price. This is analogous to Diamond's (1973) result that the optimal externality tax equals the average marginal externality. In our simple model, imagine that half of individuals have zero bias $b=0$ and the other half have bias $b_2=2b_1$. Because the average marginal bias is unchanged from the above homogeneous case, the optimal subsidy is still $s^*=(0+b_2)/2=b_1$. Graphically, the subsidy generates welfare gains of afgh for the biased types but welfare losses of afk for the unbiased types. The optimal subsidy is thus a "compromise" such that a marginal increase equates gains to the biased types (the distance fg) with losses to the unbiased types (the distance fk). This illustrates that holding constant the average marginal bias, heterogeneity in bias (and even the presence of many unbiased consumers) does not weaken the case for corrective policy.

However, with heterogeneous bias, this "optimal" subsidy is optimal only if the uniform subsidy is the policymaker's only instrument. In theory, the policymaker can do better: in the first best, all consumers with bias b_2 to the left of point "l" would take the action, whereas all unbiased consumers to the left of point "a" would take the action. Relative to the uniform subsidy, the first best would generate additional welfare gains of fgl for the biased consumers and eliminate the welfare loss of afk for the unbiased consumers. In theory, this could be approximated through behavioral "tagging," which Allcott, Knittel, and Taubinsky (2015) and Mullainathan, Schwartzstein, and Congdon (2012) discuss as an extension to Akerlof's (1978) analysis of targeted social programs. For example, larger subsidies could be made available to consumers with characteristics associated with larger bias, or energy efficiency programs could be preferentially marketed to such consumers.

This heterogeneity motivates nudges as a policy instrument. If the bias b derives from imperfect information or inattention, providing salient information can help misinformed and inattentive consumers without affecting the already informed and attentive types. If the bias b derives from present bias, offering commitment contracts can similarly help biased consumers make the privately-optimal decision without distorting choices by time consistent types. In this model, the benefit of nudges is thus not the (important) philosophical idea that they interfere less with choice. Instead, this model highlights that nudges can be preferred from an efficiency perspective to the extent that they target unbiased consumers. Of course, if nudges are costly or do not fully remove bias, it is possible that a mandate or a ban could have higher net benefits, and so the preference for nudges is only presumptive.

How well-targeted are energy efficiency policies? Allcott, Knittel, and Taubinsky (2015) argue that many energy efficiency subsidies could be poorly targeted for two reasons. First, many of the subsidies available in the U.S. are offered by local utilities and are not fully publicized, and the types of people who are aware that subsidies are available are more likely to be well-informed about and attentive to energy costs. Thus, consumers marginal to these subsidies could be mechanically less biased than the average consumer. Second, some energy efficient goods such as hybrid cars and energy efficient

lightbulbs are niche goods with small market shares and preferentially appeal to environmentalists. Environmentalists, however, are more likely to be well-informed and attentive to energy costs. Thus, relatively unbiased consumers are more likely to be marginal to any moderate subsidy. This logic also raises questions about the targeting of CAFE standards: to the extent that hybrids are largely bought by environmentalists and automakers comply with the standard by selling more hybrids, it seems likely that the CAFE policy does not affect consumers who are more biased by the factors that the RIAs use to motivate the policy.

On the other hand, some energy efficiency programs are already well-targeted. Opower and similar companies mail conservation tips to residential energy consumers, and recipient groups are often restricted to heavier users who might be unaware of their heavy usage and who might benefit more from information. Furthermore, many utilities offer "on-bill financing" through which a household can invest in insulation or other forms of energy conservation. On-bill financing programs directly address market distortions from credit constraints, and they also can help potentially present biased consumers by deferring investment costs to the same future time periods as the benefits from reduced energy costs.

While nudges are by definition better targeted than uniform subsidies and standards, using nudges alone is not optimal unless the nudge removes *all* systematic bias. For example, while energy use information labels such as appliance "yellow tags" and fuel economy labels are the natural first step to addressing imperfect information and inattention, many consumers do not see and internalize this information: 40 percent of Americans report that they "did not think about fuel costs at all" during their most recent auto purchase (Allcott 2011). Thus, the fact that information is available does not mean that the policymaker should not consider CAFE standards or other policies to address imperfect information or other distortions; see Bubb and Pildes (2014). There is a great deal of room for more empirical work here.

Principle 3: Minimize losses from regulators' imperfect information

Environmental economists often argue for flexible policies such as carbon taxes or cap-and-trade programs instead of command-and-control regulations such as specific technology mandates. In theory, a regulator with perfect information could exactly determine the socially optimal outcome and mandate firms to carry it out. A key reason to prefer flexible policies is that regulators typically do not have the information required to do this. Weitzman's (1974) model of prices vs. quantities is similarly relevant because the regulator may not know the marginal cost of compliance.

A similar logic applies to behaviorally-motivated regulation: we should prefer policies that exploit what regulators know and minimize the losses from what they do not know.

- If the policymaker is concerned about a particular behavioral bias, the policymaker should use a nudge well-matched to that bias. For example, information disclosure is the natural first approach to imperfect information and inattention.

- If the policymaker does not know the source of bias but does know the average marginal bias, the policymaker should set a tax or subsidy equal to the average marginal bias. For example, analyses by Allcott and Wozny (2014), Busse, Knittel, and Zettelmeyer (2013), and Sallee, West, and Fan (2015) are designed to identify whether auto buyers have biased valuations of improved fuel economy, but they cannot identify whether any bias derives from present bias over cash flows, biased beliefs, inattention, or other factors. If such analyses were the most rigorous evidence available, then the least informationally-demanding policy would be to calibrate a tax on low-fuel economy vehicles based on the estimated bias. By contrast, translating an average marginal bias to a quantity policy such as the CAFE standard requires also knowing the cost curve for improved fuel economy. (In our model, if the policymaker knows b, then it is easy to set s*=b, but setting an optimal quantity also requires knowledge of S(p).) This discussion highlights a useful feature of the "reduced form approach to behavioral public finance": some optimal policies depend only on the average marginal bias, and do not require knowledge of what specific behavioral model generates the bias.

- In other cases, the policymaker suspects some bias but has no reliable estimate of the bias, but can try to construct S(p) and $D_U(p)$ and mandate the social optimum. This is the approach that the CAFE RIAs take. Unfortunately, this is extremely informationally demanding. The RIAs required extensive efforts to model vehicle technologies, which required agencies to learn a remarkable amount of information about automakers. As one measure of this, the 2012-2016 light duty CAFE standards involved 2142 pages of regulatory impact statements, technical support documents, and responses to comments, and the final rule itself is 406 pages. Much of this is devoted to regulatory agencies' models of automakers' technology adoption decisions. Even from the most competent and well-motivated officials, there is inevitably a risk of error here.

Principle 4: Adopt the policy that maximizes net benefits

A simple but practically important lesson: For energy-efficiency requirements, there is a range of possible levels of stringency; assuming that regulators are choosing among those levels (and not considering nudges or economic incentives), they should adopt the approach with the highest net benefits (including internalities). Of course, any subsidy that is not correctly calibrated to equal the average marginal bias reduces welfare relative to the optimally-calibrated subsidy. This principle emphasizes the importance of measuring the average marginal bias, which is the issue we take up below.

Measuring the Average Marginal Bias

Principled refinements of the welfare-relevant domain

   Welfare analysis in the standard model assumes that the market demand curve (D_M in our model) reflects individuals' true utility (v). The more general behavioral model relaxes this assumption. The challenge is to determine in a principled way which choices reflect true preferences and which reflect mistakes. Within economics, theoretical analyses such as Bernheim and Rangel (2009) and Koszegi and Rabin (2008) have proposed particular principled refinements, and Gruber and Koszegi (2001) and many others have empirically implemented refinements in "behavioral welfare analyses."

   On the basis of this literature, we propose four principles for applied work:

1. *Use well-informed choices.* By definition, such choices should be preferred to those that are based on ignorance of relevant factors.
2. *Use considered choices.* Here, "considered" means choices where the individual evaluates all relevant facets of a product or activity.
3. *Use active choices.* Such choices reflect the agent's own values and tastes, whereas passive choices (such as failing to opt in or opt out of a default setting) may not. Notwithstanding this point, it should be acknowledged that active choices might be inferior to passive ones if they are ill-informed or ill-considered.
4. *If individuals are present-biased, use long-run instead of present-biased (impulsive) choices.* By long-run choices, we mean choices that are pre-planned for future periods, discounting future utility at a consistent discount rate. (Of course, a time-consistent preference for short-term consumption does not reflect present bias.)

   In the language of Bernheim and Rangel (2009), choices that satisfy these four criteria would constitute the "welfare-relevant domain," while choice contexts which fail any of these criteria can plausibly be defined as "suspect" on welfare grounds. If suspect choices differ from the choices in the welfare-relevant domain, they would not be used for welfare analysis. In this sense, paternalistic regulation can be limited to situations in which individuals' choices are demonstrably inconsistent, and when it can be argued in a principled way that choices in one setting do not promote people's welfare or reflect true preferences.

   These refinements mean that there are cases when this framework does *not* justify paternalistic regulation. For example, there are cases when people make considered and active choices after gathering all relevant information, but their interpretation of the evidence differs from the regulator. The framework does not support paternalism in such cases.

Measuring bias

We follow DellaVigna (2009) and Mullainathan, Schwartzstein, and Congdon (2012) in defining three empirical approaches to measuring bias:

1. *Comparing demand responses.* Consumers should be equally responsive to variations in purchase prices and to variations in other product costs. For example, quantity demanded should decrease by the same amount if a good's purchase price increases by $1 or if other costs such as shipping and handling, sales taxes, or present discounted energy costs increase by $1. Abaluck and Gruber (2012), Allcott and Wozny (2014), Busse, Knittel, and Zettelmeyer (2013), Chetty, Looney, and Kroft (2009), Hossein and Morgan (2006), Sallee, West, and Fan (2015), and others use this idea to measure whether consumers fully value particular potentially-shrouded costs. The weakness of this approach is that it is often not possible to find econometrically exogenous variation in prices and the potentially-shrouded cost.

2. *Measuring effects of nudges*. If a nudge is something that only removes bias, then the change in demand caused by a nudge is a measure of bias. Allcott and Taubinsky (2015), Chetty, Looney, and Kroft (2009), and others use this approach, providing information to a randomly-selected treatment group to measure the bias caused by imperfect information or inattention. The challenge is finding an intervention that only removes bias but has no other effects. For example, one might be worried that an "information" intervention does not just remove bias, but also subtly persuades consumers to take a different action.

3. *Comparing elicited beliefs to factual data.* Allcott (2013), Attari *et al.* (2010), and Larrick and Soll (2008) show that elicited beliefs differ from factual data in the context of energy efficiency. Many others have done this in other contexts; see Bollinger, Leslie, and Sorensen (2011) for an example related to calorie consumption.

One potential alternative to using revealed preference is to use engineering or accounting models to measure true utility. For example, Gruber and Koszegi (2001) approximate the true disutility of smoking by multiplying expected life years lost to smoking by the estimated value of a statistical life. In theory, an approach of this kind could be extremely helpful and might well deserve priority, because it would measure utility (or welfare) itself, rather than using choice, which is merely a prediction (and which might go wrong, see Kahneman and Thaler (2006)). At least in the context of energy efficiency, however, some engineering models do not accurately predict true utility in some contexts. Inaccuracies can arise in both estimated benefits and estimated costs. Davis, Fuchs, and Gertler (2013), Fowlie, Greenstone, and Wolfram (2014), Joskow and Marron (1992), and others have shown that engineering models overstate actual energy savings in particular settings. Anderson and Newell (2004), Fowlie, Greenstone, and Wolfram (2015), and Klemick, Kopitz, Sargent, and Wolverton (2014) argue that costs not included in engineering models reduce takeup of energy efficiency relative to models' predictions.

Applying Recent Estimates of Internalities to Future CAFE Regulatory Impact Analyses

The recent papers by Allcott and Wozny (2014), Busse, Knittel, and Zettelmeyer (2013), and Sallee, West, and Fan (2015) use the "comparing demand responses" approach to measure whether consumers fully value gasoline costs relative to vehicle purchase prices. The exact empirical strategies differ, but the basic approach is to compare how the relative prices of low- vs. high-fuel economy used vehicles change as gas prices change. Intuitively, as gas prices increase, the price of a Prius should increase relative to the price of a Hummer, because the higher fuel economy rating of the Prius is more valuable. An exact relationship between gas prices and used vehicle prices can be predicted under assumptions about future gas price expectations, vehicle lifetimes and miles traveled, and other factors.

If vehicle prices are not highly responsive to changes in gas prices, it would appear that consumers undervalue the true future benefits of fuel economy, perhaps due to present bias over cash flows, inattention, biased beliefs, or other factors. Conversely, if vehicle prices are fully responsive to gas prices, it is harder to argue that consumers are present-biased or poorly informed about fuel economy ratings or that fuel economy is a "shrouded" attribute. (We say "harder" rather than "impossible": the three papers show that results can be sensitive to assumptions about discount rates and future fuel prices, Allcott and Taubinsky (2015) show that the comparing demand responses approach is only an approximation to consumer bias that requires homogeneity and linearity assumptions, and one could also formulate inattention models under which consumers appear responsive to changes in gas prices but still undervalue fuel economy.)

The main estimates in Allcott and Wozny (2014) suggest that consumers value 76 percent of discounted future gas prices, i.e. a moderate bias relative to full (100 percent) valuation. Busse, Knittel, and Zettelmeyer (2013) and Sallee, West, and Fan (2015) find results more consistent with full valuation, i.e. zero average bias. All three papers agree that used vehicle prices are highly responsive to gasoline prices and that the average bias is at most moderate. This finding is important given the simulation model of Fischer, Harrington, and Parry (2006), which argues that "the efficiency rationale for raising fuel economy standards appears to be weak unless carbon and oil dependency externalities are far greater than mainstream economic estimates, or consumers perceive only about a third of the fuel savings benefits from improved fuel economy." Allcott (2013), Allcott, Mullainathan, and Taubinsky (2014), Heutel (2011), and Parry, Evans, and Oates (2010) use different simulation models and find similar qualitative results.

The empirical literature is continuing to develop. However, there is a discrepancy between the engineering model results, which suggest very large internalities, and these foregoing empirical estimates, which do not. The engineering model may overstate net benefits, or the empirical estimates may not correctly measure internalities, or there may be some additional market failure other than internalities that the empirical estimates do not capture. To be clear, this latter possibility is not implausible: for example,

there are well-established market failures related to information spillovers from new technologies. Future CAFE RIAs might productively engage with the developing literature and with this possibility.

Conclusion

A large body of existing research suggests that imperfect information and internalities are real and significant in some markets, so it is appropriate to consider them in regulatory impact analyses. In these cases, our simple model (and many more formal models in the literature) show how policy intervention can increase welfare, whether by restricting choices, changing prices, or by "nudging" individuals toward privately-optimal decisions. In this paper, we review an approach to paternalistic regulation that relies on consumers' own preferences as revealed through well-informed, considered, active, and pre-planned choices, acknowledging that uninformed, unconsidered, passive, and impulsive choices may not accurately reflect true preferences.

This approach suggests potential policy improvements in energy efficiency and other domains: taxing instead of banning choices that generate internalities, targeting policies directly toward market distortions, using regulations that rely on less regulator knowledge, and correctly calibrating the stringency of a tax, subsidy, or standard. We have plausible theoretical accounts of why consumers might not make optimal decisions, but we need to know much more empirically, above all to establish the magnitudes of internalities in different settings. Recent evidence raises important issues about these magnitudes in the context of fuel economy. Given the many billions of dollars of social welfare at stake, it is clear that further research is needed to understand these issues.

Counterpoint

Here we consider six potential counter-arguments.

Argument 1: Internalities don't exist.

Such an argument is difficult to square with a large and growing body of research showing that consumers can be present-biased, overoptimistic, inattentive, and misinformed in some settings; see DellaVigna (2009) for a review. Of course, there are many other settings where we do not make systematic mistakes, so policymakers need contextually-relevant empirical evidence. And it is important to note that the magnitude of the internalities, when they exist, may be either small or large.

Argument 2: As a general principle, governments should never interfere in individual decisions unless there is harm to others.

This idea raises deep philosophical questions (see Sunstein 2013; Conly 2012; Rebonato 2011), which we cannot engage here. Many of the policies mentioned in our first paragraph (seat belt laws, product safety regulations, etc.) enjoy broad support even though a paternalistic justification is hard to avoid, which suggests that most people do not think that the government should *never* act paternalistically (Conly 2012; Sunstein 2013). In markets where there are internalities, numerous papers have shown theoretically that taxes or other forms of government intervention can increase welfare (Allcott, Mullainathan, and Taubinsky 2014, Heutel 2011, Mullainathan, Schwartzstein, and Congdon 2012, O'Donoghue and Rabin 2006, etc.), and our simple model above gives the basic intuition for this result. Of course, the fact that policy *can* increase welfare does not mean that *any* policy will increase welfare – each specific proposal requires cost-benefit analysis.

Argument 3: Policymakers should not impose their preferences on constituents.

We agree with this argument. The central feature of the approach we advocate is that policymakers are not imposing their own preferences on constituents. Instead, policymakers should set and evaluate policy to match individuals' own preferences from the subset of contexts when they are fully informed, making active choices, and fully considering present and future costs and benefits.

Argument 4: Policymakers should maximize behavior change instead of maximizing welfare.

While economists hardly ever make this argument, it frequently arises in environmental and health policy debates. For example, one reading of some provisions of the Clean Air Act is that the EPA must impose economically and technically feasible requirements instead of maximizing net benefits. In the health context, one might argue that some foods be taxed or banned because they are unhealthy, without regard for how sugar, salt, and fat can make food taste better. We think that welfare maximization is the right criterion (for one explanation, see Sunstein 2014; on the notion of welfare, see Adler 2011).

Energy conservation and healthy eating can produce large gains, but their socially optimal quantities are not infinite, so we must use cost benefit analysis to determine the social optimum.

Argument 5: Policymakers should not regulate internalities without solid evidence that consumers are biased.

Taken in one sense, this statement is important and unobjectionable; policies should always be supported by solid evidence. Taken in another sense, however, the statement is more provocative; it suggests a strong presumption against regulation based on internalities. We believe that policymakers should use the best available evidence, and should not adopt any such presumption.

If a policymaker is risk neutral and uncertain about the average marginal bias, uncertainty about internalities enters the optimal subsidy formula just like heterogeneity in bias. For example, imagine that the policymaker believes that there is a 50 percent chance that there is no bias and a 50 percent chance that the average marginal bias is $2b_1$. The optimal subsidy is still $(2b_1+0)/2=b_1$, just as in the case with heterogeneous bias when the policymaker knows with certainty that the average is b_1. Thus, being uncertain about bias is not a reason to avoid regulation. Of course, being uncertain about whether there is a bias at all (e.g. suggesting that the average marginal bias is 0 when it was previously thought to be b_1) does imply that the optimal corrective policy is less stringent, or even that the optimal policy is to do nothing.

This is just a special case of a broader issue: policymakers almost always act with imperfect information, whether regulating internalities, providing public goods, or acting in any other traditional function. When maximizing expected welfare under uncertainty, uncertainty does not generically imply that the optimal policy is to do nothing.

Argument 6: Regulators make mistakes, so they shouldn't regulate in response to consumer mistakes

Glaeser (2006) provides three models under which bounded rationality reduces the quality of government decisionmaking more than it reduces the quality of private decisionmaking, implying that bounded rationality strengthens the case for limited government. His three models show that (1) regulators have less incentive to make the right decisions for consumers than consumers do for themselves, (2) if errors come from the influence of firms or interest groups, it is easier to "buy" a small number of regulators than a large number of consumers, and (3) individuals have more incentives when making private decisions than they do when voting, so elected politicians are less likely to have the right views on how to interfere.

We would make two notes about Glaeser's argument. First, while the models focus on paternalistic policies, the results would also generalize to regulation in traditional spheres such as regulating externalities or providing public goods. Thus, while bounded rationality strengthens the case

for limited regulation in these models, this is true for all kinds of regulation, not just paternalistic regulation.

Second, and more important, these are not the only three models, and it's easy to formulate alternative models that *strengthen* the case for intervention. Glaeser (2006, page 144) proposes one such model that could reverse his main results, under which there are returns to scale in information acquisition and decisionmaking. In this case, it is more likely to be optimal to centralize decisionmaking. This model is broadly relevant to our consumer protection and energy efficiency examples, where it can be difficult to gather and weigh information about chemical composition, future energy use, and obscure but relevant safety risks. Turrentine and Kurani (2007) document how calculating the present discounted value of energy costs can be difficult, even for accountants, bankers, and financial analysts.

Bibliography

Abaluck, Jason, and Jonathan Gruber (2011). "Choice Inconsistencies Among the Elderly: Evidence from Plan Choice in the Medicare Part D Program." American Economic Review, Vol. 101, No. 4 (June), pages 1180-1210.

Adler, Matthew (2011). Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis. Cambridge: Cambridge University Press.

Akerlof, George (1978). "The Economics of "Tagging" As Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning." American Economic Review, Vol. 68, No. 1 (March), pages 8-19.

Allcott, Hunt (2011). "Consumers' Perceptions and Misperceptions of Energy Costs." American Economic Review, Papers and Proceedings, Vol. 101, No. 3 (May), pages 98-104.

Allcott, Hunt (2013). "The Welfare Effects of Misperceived Product Costs: Data and Calibrations from the Automobile Market." American Economic Journal: Economic Policy, Vol. 5, No. 3 (August), pages 30-66.

Allcott, Hunt, Christopher Knittel, and Dmitry Taubinsky (2015). "Tagging and Targeting of Energy Efficiency Subsidies." American Economic Review, Papers and Proceedings, forthcoming.

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

Allcott, Hunt, and Dmitry Taubinsky (2015). "Evaluating Behaviorally-Motivated Policy: Experimental Evidence from the Lightbulb Market." American Economic Review, forthcoming.

Allcott, Hunt, and Nathan Wozny (2014). "Gasoline Prices, Fuel Economy, and the Energy Paradox." Review of Economics and Statistics, Vol 96., No. 10 (December), pages 779-795.

Anderson, Soren T., and Richard G. Newell (2004). "Information Programs for Technology Adoption: The Case of Energy-Efficiency Audits." Resource and Energy Economics, Vol. 26, No. 1, pages 27–50.

Bernheim, B. Douglas, and Antonio Rangel (2004). "Addiction and Cue-Triggered Decision Processes." American Economic Review, Vol. 90, No. 5 (December), pages 1558-1590.

Bovenberg, Lans, and Lawrence Goulder (1996). "Optimal Environmental Taxation in the Presence of other Taxes: General Equilibrium Analyses." American Economic Review, Vol. 86, No. 4, pages 985–1000.

Bubb, Ryan, and Richard H. Pildes (2014). "How Behavioral Economics Trims Its Sails and Why." Harvard Law Review, Vol. 127, pages 1593-1678.

Busse, Meghan, Christopher Knittel, and Florian Zettelmeyer (2013). "Are Consumers Myopic? Evidence from New and Used Car Purchases." American Economic Review, Vol. 103, No. 1 (February), pages 220-256.

Carroll, Gabriel, James Choi, David Laibson, Brigitte Madrian, and Andrew Metrick (2009). "Optimal Defaults and Active Decisions." Quarterly Journal of Economics, Vol. 124, pages 1639-1674.

Chetty, Raj (2015). "Behavioral Economics and Public Policy: A Pragmatic Perspective." Working Paper, Harvard University (January).

Chetty, Raj, Adam Looney, and Kory Kroft (2009). "Salience and Taxation: Theory and Evidence." American Economic Review, Vol. 99, No. 4 (September), pages 1145-1177.

Conly, Sarah (2012). Against Autonomy: Justifying Coercive Paternalism. Cambridge, England: Cambridge University Press.

Davis, Lucas, Alan Fuchs, and Paul Gertler (2014). "Cash for Coolers: Evaluating a Large-Scale Appliance Replacement Program in Mexico. American Economic Journal: Economic Policy, Vol. 6, No. 4, 207-238.

DellaVigna, Stefano (2009). "Psychology and Economics: Evidence from the Field." Journal of Economic Literature, Vol. 47, No. 2 (June), pages 315-372.

Diamond, Peter (1973). "Consumption Externalities and Imperfect Corrective Pricing." Bell Journal of Economics and Management Science, Vol. 4, No. 2 (Autumn), pages 526-538.

DOE (U.S. Department of Energy) (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

Dubin, Jeffrey, Allen Miedema, and Ram Chandran (1986). "Price Effects of Energy-Efficient Technologies: A Study of Residential Demand for Heating and Cooling." RAND Journal of Economics, Vol. 17, No. 3 (Autumn), pages 310-325.

Federal Register (1993). Executive Order 12866: Regulatory Planning and Review. Available at http://www.archives.gov/federal-register/executive-orders/pdf/12866.pdf

Federal Register (2010). "Light-Duty Vehicle Greenhouse Gas Emission Standards and Corporate Average Fuel Economy Standards; Final Rule." Available from http://www.gpo.gov/fdsys/pkg/FR-2010-05-07/pdf/2010-8159.pdf

Finkelstein, Amy (2009). "E-ZTAX: Tax Salience and Tax Rates." Quarterly Journal of Economics, Vol. 124, No. 3 (August), pages 969-1010.

Fischer, Carolyn, Winston Harrington, and Ian Parry (2007). "Do Market Failures Justify Tightening Corporate Average Fuel Economy (CAFE) Standards?" The Energy Journal, Vol. 28, No. 4, pages 1-30.

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram (2014). "Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program." Working Paper (September).

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram (2015). "Are the Non-Monetary Costs of Energy Efficiency Investments Large? Understanding Low Take-up of a Free Energy Efficiency Program." E2e Working Paper 016 (January).

Glaeser, Edward (2006). "Paternalism and Psychology." University of Chicago Law Review, Vol. 73, No. 1 (Winter), pages 133-156.

Grossman, Sanford and Oliver Hart (1980), "Disclosure Laws and Takeover Bids," The Journal of Finance, Vol. 35, No. 2, pages 323-334.

Heutel, Garth (2011). "Optimal Policy Instruments for Externality-Producing Durable Goods under Time Inconsistency." NBER Working Paper 17083.

Hossain, Tanjim, and John Morgan (2006). "...Plus Shipping and Handling: Revenue (Non)Equivalence in Field Experiments on eBay." Advances in Economic Analysis and Policy, Vol. 6.

Joskow, Paul, and Donald Marron. 1992. "What Does a Negawatt Really Cost? Evidence from Utility Conservation Programs." Energy Journal, Vol. 13, No. 4, pages 41–74.

Kahneman, Daniel, and Richard Thaler (2006). "Anomalies: Utility Maximization and Experienced Utility." Journal of Economic Perspectives, Vol. 20, No. 1 (Winter), pages 221-234.

Kahneman, Daniel, Peter Wakker, and Rakesh Sarin (1997). "Back to Bentham? Explorations of Experienced Utility." Quarterly Journal of Economics, Vol., No. (May), pages 375-405.

Klemick, Heather, Elizabeth Kopits, Keith Sargent, and Ann Wolverton (2014). "Heavy-Duty Trucking and the Energy Efficiency Paradox." National Center for Environmental Economics Working Paper #14-02 (January).

Kling, Jeffrey, Sendhil Mullainathan, Eldar Shafir, Lee Vermeulen, and Marian Wrobel (2012). "Comparison Friction: Experimental Evidence from Medicare Drug Plans." Quarterly Journal of Economics. Vol. 127, No. 1, pages 199-235.

Knittel, Christopher, and Ryan Sandler (2013). "The Welfare Impact of Indirect Pigouvian Taxation: Evidence from Transportation." Working Paper, MIT (February).

Larrick, Richard, and Jack Soll. 2008. "The MPG Illusion." Science, Vol. 320, No. 5883 (June 20th), pages 1593–94.

Mullainathan, Sendhil, Joshua Schwartzstein, and William Congdon (2012). "A Reduced-Form Approach to Behavioral Public Finance." Annual Review of Economics, Vol. 4, pages 17.1-17.30.

NHTSA (National Highway Traffic Safety Administration) (2012). "Final Regulatory Impact Analysis: Corporate Average Fuel Economy for MY 2017-MY 2025 Passenger Cars and Light Trucks." Office of Regulatory Analysis and Evaluation, National Center for Statistics and Analysis (March).

O'Donoghue, Edward, and Matthew Rabin (2006). "Optimal Sin Taxes." Journal of Public Economics, Vol. 90, pages 1825-1849.

Parry, Ian (1995). "Pollution Taxes and Revenue Recycling." Journal of Environmental Economics and Management, Vol. 29, pages S64-S77.

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

Rebonato, Riccardo (2011). Taking Liberties: A Critique of Libertarian Paternalism. London, England: Palgrave McMillan.

Sallee, James, Sarah West, and Wei Fan (2015). "Do Consumers Recognize the Value of Fuel Economy? Evidence from Used Car Prices and Gasoline Price Fluctuations." Working Paper, University of Chicago (January).

Sunstein, Cass R. (2013). Why Nudge: The Politics of Libertarian Paternalism. New Haven, CT: Yale University Press.

Sunstein, Cass R. (2014). Valuing Life. Chicago, IL: University of Chicago Press.

Thaler, Richard and Cass R. Sunstein (2008). Nudge: Improving Decisions About Health, Wealth, and Happiness. New Haven, CT: Yale University Press.

Turrentine, Thomas, and Kenneth Kurani (2007). "Car Buyers and Fuel Economy?" Energy Policy, Vol. 35, pages 1213-1223.

Waldron, Jeremy (2014). "It's All for Your Own Good." New York Review of Books. October 9, 2014. http://www.nybooks.com/articles/archives/2014/oct/09/cass-sunstein-its-all-your-own-good/

Weitzman, Martin (1974). "Prices vs. Quantities." *Review of Economic Studies*, Vol. 41, No. 4 (October), pages 477-491.