A Reduced-Form Approach to Behavioral Public Finance

Sendhil Mullainathan,[1] Joshua Schwartzstein,[2] and William J. Congdon[3]

[1]Department of Economics, Harvard University, Cambridge, Massachusetts 02138, and Consumer Financial Protection Bureau, Washington, DC 20552; email: mullain@fas.harvard.edu

[2]Department of Economics, Dartmouth College, Hanover, New Hampshire 03755; email: josh.schwartzstein@dartmouth.edu

[3]Brookings Institution, Washington, DC 20036; email: wcongdon@brookings.edu

Annu. Rev. Econ. 2012. 4:511–40

The *Annual Review of Economics* is online at economics.annualreviews.org

This article's doi: 10.1146/annurev-economics-111809-125033

JEL codes: B40, D01, D03, D04, D60, D61, H00, H20, I10, I30, J65, Q50

1941-1383/12/0904-0511$20.00

Keywords

behavioral economics, taxation, social insurance, externalities

Abstract

Research in behavioral public finance has blossomed in recent years, producing diverse empirical and theoretical insights. This article develops a single framework with which to understand these advances. Rather than drawing out the consequences of specific psychological assumptions, the framework takes a reduced-form approach to behavioral modeling. It emphasizes the difference between decision and experienced utility that underlies most behavioral models. We use this framework to examine the behavioral implications for canonical public finance problems involving the provision of social insurance, commodity taxation, and correcting externalities. We show how deeper principles undergird much work in this area and that many insights are not specific to a single psychological assumption.

1. INTRODUCTION

Behavioral economics poses a challenge to traditional policy analysis. Failures of revealed preference should reshape how we think about basic policy problems. Realizing this potential systematically, however, has proven challenging. The parsimony of revealed preference allows for a powerful framework for policy design and analysis. It provides an integrated treatment for how different policy levers affect total welfare, accounting for how individuals will respond to policy changes. In many cases, empirically implementable formulas also follow. Commodity taxation generates deadweight loss, and the optimal tax rate depends on the demand elasticity. Health insurance generates moral hazard, and the optimal copay depends on the elasticity of utilization. The parsimony of the underlying model of behavior generates a uniform approach to the policy problem.

In contrast, behavioral economics appears scattered. Behavioral biases vary from paper to paper. People may be inattentive, not paying enough attention to certain prices or taxes. People may be overconfident in estimating the wage offers they will get while searching for a job. People may be present-biased in choosing how much to save. In many cases, the same behavior can be understood through a diverse set of psychologies. In savings, for example, researchers have discussed all three of these biases: present bias, inattention, and overconfidence. In a way, this diversity is both exciting and necessary. Given the complexity of decision making, it is no surprise that it is not captured easily in a single model.

At the same time, this diversity can also create a sense of unease. Will the results of a model change dramatically if we assume a different psychology? For example, O'Donoghue & Rabin (2006) present a model in which taxes on tempting goods would improve the welfare of hyperbolic discounters. Perhaps most interestingly, they show that when there are both hyperbolic and exponential discounters (the latter are not helped by taxing such goods), the optimal tax is always nonzero, irrespective of the proportions. Is this result specific to hyperbolic models? More broadly, it is hard to tell which theoretical results extend and which rely on the specific psychology considered.

The empirical work is also fragmented. Some papers document a behavioral anomaly in an important policy context. For example, papers have shown that people choose stock portfolios in privatized social security or prescription drug plans in Medicare Part D badly (Cronqvist & Thaler 2004, Kling et al. 2012). Some show the power of certain "nudges" (a behavioral intervention that ought not to work in traditional models). For example, Madrian & Shea (2001) demonstrate that a simple behavioral change—the default—can significantly affect 401(k) enrollments. Yet others calibrate very specific models. Such papers integrate a specific behavioral bias into a standard policy framework. For example, Fang & Silverman (2004) incorporate present bias into a model of optimal social welfare programs. Does our interpretation of a given nudge or observed error depend on the specific psychology we attribute to it? Consider the literature on defaults in 401(k)s (e.g., Beshears et al. 2008b). Besides the empirical challenges in drawing welfare conclusions from these papers (what happens to other savings; how do we know people were saving too little?), there is also an interpretational challenge. Why do defaults work? Procrastination, inattention, status quo bias and other reasons could all be listed. Distinguishing among these mechanisms is important. But can one say something without making these distinctions?[1]

[1]In addition to this fragmentation problem, there is also the welfare problem. If revealed preferences are inconsistent, how do we adjudicate between them to form a consistent social welfare function? Within this question are embedded

We argue that there are in fact more general lessons, that broader inferences are possible, that a deeper structure does exist. We show how a simple framework organizes many of the theoretical insights and helps interpret many of the empirical ones. This framework incorporates a diverse set of psychologies (although not all). What results are lessons that transcend specific applications and psychologies, analogous to how concepts such as deadweight loss are useful across traditional areas of public finance.[2]

At the heart of the framework is a simple observation. Much of public finance relies on one aspect of revealed preference. It presumes that people take an action when the action's private benefits exceed its private costs. This helps interpret the "demand" for an action. For example, when social costs are above private costs (perhaps there is an externality, or social insurance creates the price wedge), there will be excess demand. Any behavioral assumption—the particular psychology—can be modeled as it affects the propensity to take the action. A behavioral model specifies how choice is distorted: Rather than taking the action when private benefits exceed private costs, $b > p$, people now act when $b + \varepsilon > p$, where ε reflects the distortion caused by the psychology. This trivial modeling technique proves both broad (capturing many psychologies) and useful.

Using this framework, we illustrate how some principles, even when initially derived for a specific psychological bias, extend across a variety of psychological assumptions. We illustrate how some of the empirical estimates can be interpreted independent of a specific psychological mechanism. We emphasize the word "some" here because this framework is not a substitute for specific psychological modeling. It is a complement, allowing us to understand which of the predictions and empirical findings are more general, less tied to any one psychology. Behavioral public finance will inevitably be messy, but it is not as messy as it seems.

By focusing on the broader policy problem, the framework also points to new lessons. For example, by abstracting from traditional market failures, behavioral public finance has largely focused on situations in which eliminating biases improves social welfare. However, by embedding biases in situations with pre-existing distortions, we see that biases have consequences beyond creating internalities. Biases can make it easier or more difficult to correct traditional market failures; correcting a bias may benefit the individual but harm society. We use the framework to develop formulas for the welfare consequences of nudges and taxes that highlight the forces at play. We additionally illustrate ways in which these insights can be made empirically implementable. This article is an unconventional literature review: The centerpiece is our attempt to organize a broad set of papers into one particular (and we hope useful) perspective that highlights the broader lessons.

2. THE POLICY FRAMEWORK

Many insights in behavioral public finance span across policy problems. Similar forces are at work whether we are thinking about problems of social insurance or commodity

many other deep issues, such as the reasonableness of paternalistic policies, the proper role of government, and the definition of a market failure. As a result, this issue has produced a large theoretical literature (Bernheim & Rangel 2009, Green & Hojman 2009, and Rubinstein & Salant 2011 are among the most recent). Exactly because of the breadth and complexity of this problem, however, we ignore it: Doing it justice would require its own thorough review. Instead, we analyze a specific welfare function below, which presumes that the policy maker makes assumptions about the nature of hedonic and choice utilities prior to conducting welfare analyses. As we discuss further below, one interpretation of our approach is that we follow Bernheim & Rangel's (2009) lead in allowing the policy maker to be precise about which ancillary conditions should be considered when conducting revealed preference analysis.

[2]Some of these ideas are given an informal treatment in the book by Congdon et al. (2011).

taxation. Highlighting this requires the creation of a canonical model, one that embeds within it various policy problems. This section presents that model. Although it does not cover all the major public finance problems (e.g., adverse selection in insurance markets), it does cover many. Specifically embedded in it are moral hazard problems in social insurance, commodity taxation, and correcting externalities.

Suppose there is a continuum of consumers with measure one. Each must decide whether to take some action ($a \in \{0,1\}$), e.g., whether to buy a more fuel-efficient car, to go to the doctor, or to buy some product. Taking the action costs p, where this cost includes both direct costs (e.g., the market price of a product) and less direct costs (e.g., the time cost of going to the store). If the consumer chooses the action, it yields benefit b, which we write in income-equivalent units. This benefit differs across people, and it is distributed according to $b \sim F(b)$.[3] We also allow for actions to create externalities. We write ℓ as the disutility a given person experiences when someone else takes the action. If the total number of people who take the action is A (endogenized later), we write $A\ell$ as the total disutility a given person experiences from others taking the action.

The government can affect the price through t, $p = p(t)$, which is interpreted as a tax when positive and as a subsidy when negative. The government can also transfer T to individuals. For simplicity, we sometimes refer to t as taxes and T as transfers. All individuals face the same tax to taking the action and receive the same transfer payment. If the transfer is conditional (e.g., on being unemployed), we model this as affecting the tax to taking the action (e.g., the employment tax includes the forgone unemployment benefit).

Consumers have initial wealth Y.[4] Their utility is an increasing, concave function ($U' > 0$, $U'' \leq 0$) of final wealth (initial wealth plus the government transfer), any net benefit that accrues from taking the action, and any disutility derived from others taking the action. Utility therefore is

$$U(Y + T + (b - p(t)) - A\ell)$$

if the agent takes the action, and

$$U(Y + T - A\ell)$$

if she does not. We can write utility more simply as $U(Y + T + a(b - p(t)) - A\ell)$.

In this setup, when agents have no psychological biases (i.e., they are "standard agents"), the action is taken whenever the benefits exceed the costs, or

$$b > p(t).$$

Fixing the price at p, the total demand of standard agents is the number for which benefits exceed costs. Formally, let the total demand of standard agents be denoted by $A^S(p) \equiv 1 - F(p)$. [More generally, we denote the total demand by $A(p)$.] Sometimes we write demand as a function of taxes: $A(t) = A(p(t))$. We note a simplifying feature of the model: Because benefits to taking the action are in income-equivalent units, the agent's choice of whether to take the action depends only on its price, not the agent's income (i.e., there are no income effects).

[3]This variability across people can also be thought of as the ex ante uncertainty that any one person faces. For example, at the start of the year, someone may not know whether she will become unemployed and benefit from exerting effort to find a job.

[4]Wealth Y may be stochastic, but we assume (for simplicity) that it is deterministically related to b.

Table 1 Special cases of the framework

	Commodity taxation	Corrective taxation	Health insurance	Unemployment insurance
Income	Fixed wealth	Fixed wealth	Fixed wealth net of stochastic disease severity	Fixed wage if unemployed
Action	Buying commodity	Taking action that may create an externality	Getting treated	Exerting effort to become employed
Benefit	Value of commodity	Value of action	Value of treatment	Wage net of costs of search effort and of wage if unemployed
Tax	Sales tax on commodity	Corrective tax (e.g., Pigouvian) on action	Subsidy of getting treated, as implied by the copay	Tax on the employed
Externality (typically)	$\ell = 0$	$\ell \neq 0$	$\ell = 0$	$\ell = 0$
U'' (typically)	0	0	<0	<0

The government faces constraint $G(t,T) = 0$ in setting (t,T). This constraint implicitly defines transfers as a function of taxes: $T = T(t)$. Unless otherwise noted, to simplify matters we make the following assumption on $p(t)$ and $T(t)$.

> **Assumption 1**: Transfers are an increasing function of revenue raised by the government. Letting $R(t) = tA(t)$, $T(t) = g(R(t))$ for some increasing, differentiable function $g(\cdot)$. Taxes additively affect the price agents face to taking the action: $p(t) = p_0 + t$ for some constant p_0, where p_0 can be interpreted as the (constant) marginal cost to taking the action.

The government's problem is simple here. It maximizes expected utility:[5]

$$W(t) = E[U]. \tag{1}$$

This simple framework nests many more specific models. **Table 1** illustrates how the framework can be applied to the study of problems involving taxation and social insurance. As a brief illustration, in the case of commodity taxation, b can represent the value of a product produced by a competitive industry at constant marginal cost p_0, and $t > 0$ can stand for the sales tax. In the case of health insurance, b can represent the value of getting treated at marginal social cost p_0, and $t < 0$ the implicit subsidy implied by the copay.

3. ADDING BEHAVIORAL ELEMENTS

A large and growing body of evidence shows that behavioral biases may play a role in policy-relevant decisions. In the case of health care choices (Baicker et al. 2011), for instance, behavioral biases can distort care use. For example, diabetics fail to take glucose-lowering drugs even at zero cost and despite their large benefits and lack of side effects. For unemployment

[5]For problems where $a \in \{0, 1\}$ represents whether the agent buys a product, our formulation implicitly assumes that producers make zero profits.

insurance, false perceptions about how effort translates into job offers can distort the decision to search for (and accept) such offers (Spinnewijn 2009). With carbon externalities, people may buy energy-inefficient durables because they do not pay sufficient attention to future energy costs (Hausman 1979, Allcott et al. 2011). With standard commodities, people may fail to appreciate obscured nonsalient features of the price (Gabaix & Laibson 2006). Interestingly, taxes may be one of these nonsalient features (Chetty et al. 2009).

A remarkably diverse set of psychologies can be brought to bear on these problems. Even within any one problem, the exact psychology may not yet be clear. Do people fail to take medications because of hyperbolic discounting (the costs of adherence occur today, whereas the benefits are in the future) or because of limited attention (the symptoms may not be particularly salient)? In our simple framework, we capture these diverse psychologies using a simple modification to the $b > p$ choice rule. Specifically, we assume behavioral agents choose according to whether

$$b + \varepsilon > p, \tag{2}$$

where ε is an "error" that can depend on two factors: b and t. We make the simplifying assumption that $b + \varepsilon(b,t)$ is strictly increasing in b, and for any t, there is a unique $\tilde{b}$ in the support of $F(b)$ that satisfies

$$\tilde{b} + \varepsilon(\tilde{b}, t) = p(t).$$

Denote this level of $\tilde{b}$ by $b^B(t)$ and assume that $b^B(t)$ is differentiable with $db^B/dt > 0$. The level of demand of behavioral agents is the number for which decision benefits outweigh costs, or $A^B(t) = A^B(p(t)) = 1 - F(b^B(t))$.

This formulation implicitly captures a division between preference as revealed by choice and utility as it is experienced, or between decision utility and experienced utility (Kahneman et al. 1997). In our framework, $b - p$ affects the experienced utility of taking the action. Individuals instead choose as if $b + \varepsilon - p$ affects this utility.[6] We focus on behavioral models that imply a clear wedge between these two objects—in other words, models that generate a nonzero ε term. This clarifies a category of behavioral models that lies outside our focus: models of nonstandard preferences. For example, loss aversion may alter how individuals experience benefits: Benefits will vary depending on whether they are perceived as a loss or gain relative to some reference point (Kahneman & Tversky 1979). In job search, the prior wage may serve as a reference point against which future job offers are hedonically experienced. Similarly, other regarding preferences may affect how individuals value contributions to a public good. Fairness norms may lead people to contribute even if self-interest alone would dictate them not to. In these kinds of situations, it would be wrong to model the behavioral factor as a bias affecting ε; rather it should be modeled as a force that affects the mapping between outcomes (such as earnings) and benefits b.

In addition, although in Section 7 we briefly discuss the situation in which people have differing behavioral tendencies, for the most part we restrict attention to the situation in which biases are homogeneous and expressed with certainty—i.e., to the case where $\varepsilon(b,t)$ is nonstochastic (fixing b and t). This makes it difficult for the model to capture certain psychologies, such as forgetfulness, where ε is nonzero only with some probability. This

[6]In this setup, we deviate from Bernheim & Rangel (2009) by implicitly taking as primitive the notion of a true underlying preference from which actions may deviate. However, as we discuss further below, much of what we do can be understood in their framework as assuming that decisions under a specific ancillary condition reflect true preferences.

limitation is not intrinsic to the approach, however. We could extend much of the analysis to the case where ε is stochastic, but it complicates the formulas without adding much intuition. For the purpose of this review, we favor simplicity.[7]

At the same time, this simple formalization does capture a variety of behavioral phenomena. We give three examples of specific positive models of behavior consistent with our formulation.

Example 1 [present bias (e.g., Laibson 1997, O'Donoghue & Rabin 1999)]: Along the lines of O'Donoghue & Rabin (2006), suppose that agents decide whether to take some action that results in an immediate benefit but a delayed cost, such as consuming cigarettes or potato chips. Furthermore, suppose the delayed cost is the same across agents, but current benefits to consumption may differ. Formally, $b = v - h$, where $v \sim \tilde{F}(v)$ represents current benefits, and h, a constant, represents a delayed cost. The price agents face to taking the action may also be paid out over time, for example, if a purchase is financed. Denote the upfront price by p_1 and the rest, $p - p_1$, by p_2. Whereas standard agents are assumed not to discount future costs (for simplicity), behavioral agents discount future costs by factor $\beta \in (0,1)$, quantifying the degree to which they are present-biased. Instead of taking the action whenever $b = v - h > p = p_1 + p_2$, present-biased agents take the action whenever $v - \beta h > p_1 + \beta p_2$. In this case, $\varepsilon = (1 - \beta)(h + p_2)$.

Example 2 [inattention to nonsalient components of the price (e.g., Gabaix & Laibson 2006, Chetty et al. 2009, DellaVigna 2009)]: Similar to Chetty et al. (2009), suppose that agents decide to take some action for which benefits are salient, but part of the price vector is opaque. For simplicity (and in keeping with many applications of interest), we equate this portion of the price vector with taxes (e.g., sales taxes), but other aspects of prices may fail to be salient, for example, shipping costs and add-on prices (see DellaVigna 2009 for a review).[8] Inattentive agents discount nonsalient costs by factor $\theta \in (0,1)$, so instead of taking the action when $b > p_0 + t$, they take the action when $b > p_0 + \theta t$. In this case, $\varepsilon = (1 - \theta)t$.

Example 3 [false beliefs and overconfidence (e.g., Sandroni & Squintani 2007, Spinnewijn 2009)]: Finally, an extremely simple example consistent with our formulation is one in which agents misestimate or have false beliefs about the benefit to taking some action. To illustrate, people may be overconfident in their baseline ability to find a job and underweigh the value of searching harder or of taking up some social benefit, such as food stamps. Conversely, people may overweigh the benefits to taking an action, for example, when buying an herbal medicine with no known efficacy. Instead of taking the action when $b > p$, such agents take the action when $\hat{b} > p$, where $\hat{b}$ is the decision benefit to taking the action. In this case, $\varepsilon = \hat{b} - b$.

[7]The main difficulty seems to arise in the context of developing tools to identify biases—the topic of Section 8. Spinnewijn (2012) independently develops a related model in the context of studying insurance demand that focuses on the situation in which ε, the deviation between perceived and true insurance value in that context, is stochastic and averages to zero across the population (conditional on b).

[8]Some taxes may be fully salient if they are included in the posted price of the product (e.g., excise taxes).

4. RETHINKING THE DEMAND CURVE

The addition of behavioral elements forces us to rethink the demand curve. Recall the standard trade-off in commodity taxation. Suppose we raise taxes by Δ. The benefits of this increase stem from increasing revenue. The costs stem from distorting choices. Quantifying this inefficiency in the standard model is relatively straightforward. Prior to the increase, everyone with $b > p(t) = p_0 + t$ would buy the good. After the increase, only those with $b > p_0 + t + \Delta$ would do so. The marginally dissuaded people are those with $p_0 + t + \Delta > b > p_0 + t$. For small changes (small Δ), we know the marginal person roughly has benefit $b \approx p_0 + t$. The marginal inefficiency from the change in behavior therefore is simply the number of people who change their purchase (the demand response) multiplied by the difference between the marginal social value and marginal social cost of them buying the good, which (ignoring externalities) equals $p_0 + t - p_0 = t$. The key part of this argument is that we can infer the marginal private benefit b through the logic of revealed preference: Being marginal reveals indifference. In the case of commodity taxation, this insight allows us to use empirical quantities (demand) to estimate the deadweight loss.

The same logic is useful in many other examples. Take the case of health insurance for which there is a service that costs society p_0, but consumers face a copay $p_0 + t$, where $t < 0$. Suppose we are thinking of lowering the copay (i.e., lowering t). The benefit is the rise in insurance value. The cost is the moral hazard that follows. To quantify the moral hazard, we again would note that the marginal person roughly has a private benefit equal to the copay: $b \approx p_0 + t$. Because the cost of providing the service is p_0, the cost of moral hazard simply becomes t times the demand response. For the pure externality case, the Pigouvian tax formula follows directly from this logic. Suppose there is a good with marginal externality ℓ, and consumers face price $p_0 + t$. By the logic above (abstracting from the benefits of raising revenue), the marginal net benefit of increasing the tax is proportional to $\ell - t$, which suggests the optimal tax is $t = \ell$, the Pigouvian formula.

Many more examples make this same point. Revealed preference plays a crucial role in public finance. It is because we know that people choose according to whether $b > p$ that we can replace the marginal private benefit of taking the action with the price in analyzing everything from deadweight loss to the optimal corrective tax.

This point also makes clear how and why psychology can alter the basic logic. As we have modeled it, agents now choose according to whether

$$b + \varepsilon > p(t),$$

which affects our inference about who the marginal person is.

4.1. Who Is at the Margin Changes

Let us again consider the case of commodity taxation. Suppose we raise taxes by Δ. As above, the benefits stem from greater revenue and the "costs" from affecting choices. Prior to the increase, everyone with $b + \varepsilon > p_0 + t$ would buy the good, and after only those with $b + \varepsilon > p_0 + t + \Delta$ would. The marginally dissuaded people are those with $b \approx p_0 + t - \varepsilon$. This changes the logic of taxation, perhaps even drastically. Take the case where $\varepsilon > 0$, for what we might call "sin goods." These are goods that the marginal person buys despite the fact that their consumption provides less hedonic utility than the price. For example, in a model with hyperbolic discounting, $\varepsilon > 0$ arises when a good has long-term costs (e.g., cigarettes). In these cases, the cost of distorting behavior is no longer a cost but a benefit (at least for small taxes). For goods such as these, the government can raise revenues while solving an inefficiency (the internality).

More broadly, we can no longer say that the marginal person has private value equal to the price faced. The behavioral individual does not reveal indifference by being marginal. Perhaps b equals p or perhaps b is greater than p, but ε is less than zero, or perhaps b is less than p, but ε is greater than zero. The inverse demand function no longer coincides with the marginal private benefit curve. Any two people at the margin may value taking the action differently. It may even be impossible to use prices to get high-value individuals to take the action (e.g., when $\varepsilon \approx -\infty$) or to keep low-value individuals from taking the action (e.g., when $\varepsilon \approx \infty$). Behavioral agents may make inefficient choices even when they face what the standard model suggests is the correct price to taking the action.

It is worth noting that it is exactly this line of reasoning that led us to the reduced-form model. For many policy problems, the feature of the behavioral bias that matters most is how it affects who is marginal. This change is at the heart of our entire approach.

4.2. Nudges Matter

Before we draw out the consequences of this change, it is worth pointing out a second key implication of the behavioral approach to public policy: There exist new policy levers that influence demand. We call these nudges, after Thaler & Sunstein (2008). More precisely, in our framework nudges influence the choices of behavioral, but not standard, agents. Formally, there is a set N of available nudges, where a generic nudge is denoted by $n \in N$. Given nudge n, a behavioral agent takes the action when $b + \varepsilon_n > p$.[9]

A wide variety of nudges have been documented in the literature, for example, defaults and active choice in 401(k)s (Madrian & Shea 2001, Carroll et al. 2009). Another powerful nudge is social comparison. For example, giving people information on how their residential energy consumption compares with that of their neighbors has been shown to have significant effects on usage (Schultz et al. 2007, Allcott 2011). Other nudges belong to a class that operates by presenting relevant information in a salient way. For example, giving seniors simple, personalized information about the costs of prescription drug plans (as in Kling et al. 2012), or parents clear information about school quality (as in Hastings & Weinstein 2008), changes observed behavior. A related class of nudges works to great effect just by simplifying processes by which individuals achieve outcomes. For example, Bettinger et al. (2009) find impacts on college enrollment from an intervention that provided assistance with financial aid applications. Similarly, Choi et al. (2009) show that 401(k) enrollment increases just by simplifying enrollment forms. These are merely a few.

Nudges can influence the impact of more traditional price levers. In redistribution, outcomes depend not just on benefit levels but also on the details of their form, or how they nudge. For example, the timing of benefits, given their level, should matter little for outcomes when individuals are rational. But accumulated evidence suggests that this timing is in fact consequential. For example, consistent with present bias, individuals exhibit a drop in calorie intake over the month when they receive food stamp benefits (Wilde & Ranney 2000, Shapiro 2005). Similar timing effects are found in Supplemental Security Income (Dobkin & Puller 2007) and Social Security (Stephens 2003, Mastrobuoni &

[9]Furthermore, suppose that, for any n,t, there is a unique $\tilde{b}$ in the support of $F(b)$ that satisfies $\tilde{b} + \varepsilon_n(\tilde{b}, t) = p(t)$, with $b + \varepsilon_n(b,t)$ strictly increasing in b.

Weinberg 2009). Increasing the frequency at which benefits are paid, without changing their level, can improve program outcomes (Stephens & Unayama 2008).

In a similar example, evidence suggests that the impact of subsidization to retirement savings for low-income households depends on both the level and the form of the benefit, for which the subsidy appears to be more effective in the form of a match rather than a credit (Duflo et al. 2006, Saez 2009). A match might be more salient or easier for behavioral types to understand. In either case, the key point is that it is not just the net present value of benefits that matters, but also their structure.

5. RETHINKING GOVERNMENT INTERVENTION

When and how should the government intervene in the market? The behavioral framework forces us to rethink these questions.

5.1. Standard Model

First let us consider the standard model. The government has a basic lever it must set: the level of the tax. A simple way to think about when the government should intervene or change policy is to analyze the welfare impact of policy changes. Formally, let W^S denote welfare when there are only standard agents in the population, and totally differentiate W^S with respect to t:

$$\frac{dW^S}{dt} = -E^S[U'(C)]\left[\ell A^{S'}(t) - T'(t)\right] - E^S[U'(C)|a=1]A^S(t)\frac{dp}{dt}, \tag{3}$$

where $C = Y + T + a(b-p) - A\ell$, and $E^S\,[\cdot]$ denotes the expectation operator when there are only standard agents in the population. A key step in the derivation of Equation 3 is the envelope theorem, which allows us to ignore the welfare impact of the behavioral change on the marginal agent's utility as he is indifferent between taking and not taking the action.[10]

Imposing Assumption 1, convert dW^S/dt into a money metric by normalizing the increase in welfare by the welfare gain from increasing income by 1:[11]

$$\begin{aligned}\frac{dW^S}{dt}\Big/\frac{dW^S}{dY} &= A^{S'}(t)[MSB - MSC]^S(t) + TV^S(t)\cdot A^S(t)\\ &= A^{S'}(t)[t + ME(t)] + TV^S(t)\cdot A^S(t),\end{aligned} \tag{4}$$

where $TV^S\,(t) = \left(\frac{g'(R(t))E^S[U'(C)] - E^S[U'(C)|a=1]}{E^S[U'(C)]}\right)$ equals the value per dollar transferred from people who take the action to the population overall, $ME(t) = (g'(R(t)) - 1)t - \ell$ equals the

[10]More formally, W^S can be expressed as $E^S\left[U\left(Y+T+a(b-p)-A^S\ell\right)\right] = \int_{-\infty}^{p} U\left(Y+T-A^S\ell\right)dF(b) + \int_{p}^{\infty} U(Y+T+(b-p)-A^S\ell)dF(b) = \max_{\tilde{b}}\left\{\int_{-\infty}^{\tilde{b}} U\left(Y+T-A^S\ell\right)dF(b) + \int_{\tilde{b}}^{\infty} U\left(Y+T+(b-p)-A^S\ell\right)dF(b)\right\}$, where, for notational simplicity, we have suppressed how variables (e.g., p and T) depend on t. Differentiating $W^S(t)$ and employing the envelope condition on $\tilde{b}$, we yield Equation 3.

[11]Assumption 1 implies that $dp/dt = 1$ and $dT/dt = g'(R(t))\left[A^S(t) + tA^{S'}(t)\right]$. Plugging these equalities into Equation 3 and rearranging, we yield $E^S[U'(C)]A^{S'}(t)[g'(R(t))t - \ell] + \left(E^S[U'(C)]g'(R(t)) - E^S[U'(C)|a=1]\right)\cdot A^S(t)$. Normalizing by $dW^S/dY = E^S[U'(C)]$ then gives Equation 4.

marginal external effect of taking the action, and $[MSB - MSC]^S(t) = t + ME(t)$ equals the difference between the marginal social benefit and marginal social cost of taking the action.

There are two terms in Equation 4. The second term captures the redistribution motive for the government: to transfer money from low- to higher-value use. This motive can arise from the desire to transfer money to those with high marginal utility, or to provide insurance. Alternatively, it can arise from the desire to raise revenue to fund a public good with high marginal return.

The first term is the one we focus on, which captures the inefficiencies that come from price wedges. Absent pre-existing government involvement ($t = 0$), this term is nonzero if $\ell \neq 0$, i.e., if there is an externality.[12] With government involvement (e.g., when the government provides social insurance), this term is nonzero when $\ell - tg'(R(t)) \neq 0$. For example, in the case of health insurance, this term can be nonzero because insurance effectively subsidizes treatment ($t < 0$). More generally, in a standard model, there are distortions to activity when the prices people face do not reflect social cost, for example, because those prices omit externalities or because those prices include pre-existing taxes or subsidies. A change in policy can reduce these distortions by bringing prices closer to first best levels.

In short, in a standard framework, opportunities to improve on outcomes through government intervention arise because of a failure of redistribution or insurance, or because people do not face the social cost of taking an action. How does the analysis change in the behavioral model?

5.2. Behavioral Model

Let W^B denote welfare when there are only behavioral agents in the population. The analog to Equation 3 is

$$\frac{dW^B}{dt} = -E^B[U'(C)]\Big[(\ell + \tilde{\varepsilon}(t))A^{B'}(t) - T'(t)\Big] - E^B[U'(C)|a = 1]A^B(t)\frac{dp}{dt}, \qquad (5)$$

where $\tilde{\varepsilon}(t) = \frac{U(Y+T-A\ell)-U(Y+T-A\ell-\varepsilon(b^B(t),t))}{E^B[U'(C)]}$ is an increasing function of the error of the marginal behavioral type. (Note that $\varepsilon = \tilde{\varepsilon}$ for linear U, so we have the approximation $\varepsilon \approx \tilde{\varepsilon}$ if we take U to be approximately linear.) The main difference between Equations 5 and 3 is the addition of $\tilde{\varepsilon}(t)$, reflecting that the envelope theorem no longer holds in the behavioral model—the marginal agent is not truly indifferent between taking and not taking the action [the normed utility difference is $-\tilde{\varepsilon}(t)$].

Once again, to better interpret Equation 5, impose Assumption 1 and convert dW^B/dt into a money metric:

$$\begin{aligned}\frac{dW^B}{dt}\Big/\frac{dW^B}{dY} &= A^{B'}(t)[MSB - MSC]^B(t) + TV^B(t)\cdot A^B(t)\\ &= A^{B'}(t)[t + ME(t) + MI(t)] + TV^B(t)\cdot A^B(t),\end{aligned} \qquad (6)$$

where $TV^B(t) = \left(\frac{g'(R(t))E^B[U'(C)]-E^B[U'(C)|a=1]}{E^B[U'(C)]}\right)$ equals the value per dollar transferred from people who take the action to the population overall, $MI(t) = -\tilde{\varepsilon}(t)$ equals the marginal

[12] Absent government involvement, the private cost of taking an action can also differ from the social cost for reasons not captured in our simple framework, for example, due to monopoly pricing.

internal effect of taking the action, and $[MSB- MSC]^B\ (t) = t + ME(t) + MI(t)$ equals the difference between the marginal social benefit and marginal social cost of taking the action when there are behavioral types in the population.

By comparing $\frac{dW^S}{dt}$ with $\frac{dW^B}{dt}$, we can see how psychological factors alter the logic of how to assess inefficiencies. There are two terms in Equations 4 and 6. The second term in both captures the redistribution motive for government: Transfer money to higher-value use. This motive is often similar in the standard and behavioral cases.

The first term, which captures choice consequences of the policy change, differs between the two. Although in both cases the welfare effect of the change in behavior equals the number of people who are on the margin times the wedge between the marginal social benefit and marginal social cost of taking the action, this wedge differs across the behavioral and standard models. In the standard model, it equals $t + ME(t)$: Differences between the marginal social benefit and marginal social cost to taking the action arise from the existence of taxes and externalities. In the behavioral model, there is an additional term, the marginal internality, $MI(t)$, which arises from the fact that demand no longer reflects the marginal private benefit to taking the action. An analyst who ignores this term may misdiagnose problems, or mispredict when and how behavior will be distorted.

Consider the case of health insurance in which we know that prices are "too low." People face copays below cost when deciding whether to go to the doctor or take a pill. The standard implication is that too many people will go to the doctor or take a pill. Overutilization is the only concern. By focusing on the direction of the distortion implied by the price wedge, we fail to appreciate that underutilization, possibly resulting from procrastination or inattention, may be a big problem as well (McGlynn et al. 2003, Baicker et al. 2011).

One obvious implication of the existence of marginal internalities is that the government might intervene simply to correct them. Absent pre-existing government involvement and externalities $(t = 0, \ell = 0)$, the difference between the marginal social benefit and marginal social cost to taking the action becomes $MI(t)$. Pricing errors (like pricing externalities) can improve welfare. To illustrate, when $U'' = 0$, $\ell = 0$, and $g' = 1$, Equation 6 equals zero when $t = -MI(t)$: The optimal tax perfectly offsets the bias. This is, for example, the logic behind imposing sin taxes on commodities such as cigarettes (Gruber & Koszegi 2001).

5.3. Nudges

When people are behavioral, the government can also intervene with nudges. Suppose N is some interval in $\mathbb{R}$, and $\varepsilon_n(b,t)$ is differentiable in n. Then

$$\frac{dW^B}{dn} \Big/ \frac{dW^B}{dY} = \frac{dA_n^B}{dn}[MSB - MSC]_n^B(t) = \frac{dA_n^B}{dn}[t + ME(t) + MI_n(t)], \tag{7}$$

where A_n^B equals demand, $MI_n(t) = -\tilde{\varepsilon}_n(t) = -\frac{U(Y+T-A\ell)-U(Y+T-A\ell-\varepsilon_n(b_n^B(t),t))}{E^B[U'(C)]}$ equals the marginal internal effect, and $[MSB - MSC]_n^B(t) = t + ME(t) + MI_n(t)$ equals the difference between the marginal social benefit and marginal social cost of taking the action given the nudge. Comparing Equation 7 with Equation 6 illustrates a difference between nudges and prices: Nudges affect behavior without directly transferring money across individuals. The government may want to nudge whenever there is a wedge between the marginal social benefit and marginal social cost of taking the action, fixing the tax.

In particular, the government may want to nudge to correct internalities. Absent other distortions ($U'' = 0$, $\ell = 0$, and $g' = 1$), the difference between the marginal social benefit and marginal social cost to taking the action becomes $MI_n(t)$ when $t = 0$. In this case, the government will want to nudge to eliminate biases if possible: Equations 7 and 6 both equal zero when $MI_n(t) = 0$ and $t = 0$.

More generally, if the government can implement the first best when people do not make errors, then, if possible, it is optimal to nudge in a way that eliminates errors. One simple way to formalize this argument is as follows. Suppose the first best is implementable through feasible tax and transfer scheme (t', T') in a world where people do not make errors. If there exists a nudge $n' \in N$ that eliminates all errors at t' [i.e., $\varepsilon_{n'}(b, t') = 0$ for all b]—an idea that is implausible but instructive—then it is clear that the first best is implementable by setting $(n^*, t^*, T^*) = (n', t', T')$.

This result applies, for example, if the only market failure stems from internalities, or from internalities as well as externalities that can be corrected with Pigouvian taxes. For example, if we believe the only harm from failing to take some protective health behavior befalls individuals directly, then nudges that encourage such behavior will improve social welfare. Similarly, if we believe that the only welfare consequences to undersaving are the private hardships such individuals face in retirement, then nudges that lead individuals to save more for retirement, such as defaults, will be welfare improving.

It also will be socially optimal to nudge individuals to their private optimum in cases in which it is possible to perfectly correct for market failures. So, for example, when the externality that arises from energy consumption is perfectly mitigated through a corrective tax, nudges that correct for any remaining individual tendency to consume at suboptimal levels will improve social welfare.

Much of the behavioral policy literature discusses this case, such as the large literature on behavioral economics and retirement savings (e.g., Madrian & Shea 2001, Thaler & Benartzi 2004). In fact, when one thinks of behavioral policy, the focus is often on correcting these pure internalities, i.e., situations in which the government is at the first best, absent the internality.

However, even more interesting is the case in which internalities operate in a "second best" world.

6. EXTERNAL EFFECTS OF CORRECTING INTERNALITIES

In situations in which there is already room for the government to improve on outcomes, biases may make it more or less expensive for the government to correct traditional market failures. Conversely, correcting biases may exacerbate or attenuate these failures.

6.1. The Double Dividend

Let us consider social insurance. In the standard model, there is an inherent trade-off between the risk-smoothing benefits such programs provide to recipients and the moral hazard they create by distorting incentives. Examining Equation 4, the sign of $A^{S'}(t)[MSB - MSC]^S(t) = A^{S'}(t)t$ differs from the sign of $TV^S(t) \cdot A^S(t)$. From Equation 6, behavioral biases exacerbate this trade-off when $MI(t)$ shares the sign of t and attenuate it when $MI(t)$ is of the opposite sign.

Unemployment insurance provides a possible example of the first case. When individuals are making a private error in staying out of the workforce for too long, greater insurance not only has costs in the moral hazard it creates, but also may go against the private interests of workers. For example, when individuals are making an error by underweighing the future while unemployed, policy reforms that increase the spell length might leave individuals worse off in the long run (DellaVigna & Paserman 2005, Paserman 2008). Similarly, when individuals are making errors about the benefits to search, or the likelihood of their re-employment, policies that serve to discourage re-employment may hurt individuals (Spinnewijn 2009).

Conversely, biases may make it less expensive for the government to provide health insurance. Procrastination, limited attention, and other biases lead people to underuse certain care (Baicker et al. 2011). As a consequence, health insurance can provide more than just financial protection: It can increase the efficiency of health delivery by counteracting biases. Whereas the standard model suggests a clear tension between incentives and insurance, the behavioral model may attenuate or reverse this logic.

To take another example, let us consider externalities. Taxes addressing environmental externalities may, and sometimes must, address both those externalities and any decision-making errors on the part of consumers that interact with the externality. For example, policy that addresses auto emissions has to deal with both the environmental externality and any time inconsistency on the part of consumers (Heutel 2011). As a result, policies that correct for these externalities might also help individuals help themselves (Allcott et al. 2011). For example, if individuals are inattentive to energy prices, the optimal corrective tax rate may be higher than with rational agents as it has to correct for not just the externality, but also the internality. In doing so, these taxes also pay a double dividend by improving both private welfare on the part of behavioral individuals and social welfare.

6.2. Privately Optimal Nudges May Not Be Socially Optimal

Section 5.3 provides a condition under which the government should do what it can to debias individuals: Absent the bias, the government can implement the first best. Unfortunately, this result does not apply to most problems relevant to public policy as there is typically a traditional market failure that cannot be fully corrected with the available levers. For example, the cleverest policies may not be able to fully eliminate adverse selection and moral hazard in the provision of social insurance, or provide public goods without the imposition of taxes that distort other activities away from first best levels.

For this broad class of second best problems, it is not necessarily efficient to nudge people toward making privately optimal decisions. In the presence of other market failures, even nudges that improve private welfare may not be socially optimal. Improving outcomes for individuals might undermine broader social welfare. It may be socially optimal to allow errors to persist. In some cases, it might even be welfare improving to implement nudges that actively move individuals away from their private optimum if doing so has benefits to social welfare that overwhelm any private cost.[13]

[13]In a way, this observation can be seen as an extension of "The General Theory of Second Best" (Lipsey & Lancaster 1957), which basically says that holding some distortions in the economy fixed (e.g., stemming from distortionary commodity taxes), it is not necessarily desirable to correct other distortions. Suppose the policy maker is extremely powerful: She has access to nudges that can get behavioral types to take essentially any action she wants. In this case, for a broad class of second best problems, the policy maker will never want to nudge behavioral types into taking privately optimal decisions given the level of the traditional levers. The reason is simple. Fixing (t, T), marginally nudging behavioral types away from making privately optimal decisions has a second-order effect on their utility but a first-order effect by relaxing the government's constraint.

To examine this in a bit more detail, we start from a situation in which people are biased and the government has set taxes and transfers optimally given the bias. Consider the welfare impact of marginally nudging people away from being biased, i.e., by nudging people in a manner such that $dA_n^B/dn \cdot MI_n(t) > 0$ (e.g., by making taxes marginally more salient). From Equation 7,

$$\frac{dW^B}{dn} / \frac{dW^B}{dY} = \frac{dA_n^B}{dn} \cdot MI_n(t) + \frac{dA_n^B}{dn} \cdot [t + ME(t)]. \tag{8}$$

Equation 8 highlights the internal and external effects of marginally debiasing agents. The first term captures the internal effect and is unambiguously positive as debiased agents make (privately) better decisions. The second term captures the external effect, which is positive if and only if less biased agents are more (less) likely to take an action that is socially beneficial (harmful) at the margin. For example, suppose that we start at a situation in which there is some positive externality that the government partially but does not fully correct through subsidies (e.g., because subsidizing is costly). If agents are inattentive to the subsidy, making it more salient both helps them help themselves and has a positive external effect.

Conversely, suppose the government wants to finance a public good, but the public good should be undersupplied relative to the first best level if taxes are made fully salient because commodity taxation is distortionary (Atkinson & Stern 1974). If taxes are less salient, then there is a loss from some consumers buying too much of the taxed good relative to what is privately optimal, but there is a gain from being able to supply more of the public good, fixing the level of taxes. On net, it is unclear which effect dominates.

A set of examples in which the socially optimal nudge may not be privately optimal comes from redistribution, where achieving the social goals of such programs can be at odds with maximizing participants' private welfare. For example, individual labor supply responses to the Earned Income Tax Credit may result in part from error (Chetty & Saez 2009). Nudges, such as providing information to beneficiaries, might lead them to improve their personal welfare, which may manifest in individual labor supply becoming more responsive to the credit. But this might be at odds with the social goal of the program, at least in regions of the credit schedule where this would discourage work.

Or consider the potential conflicts between what is privately optimal for beneficiaries and social goals related to targeting benefits. Nudges such as automatic enrollment, the simplification of benefit schedules, and the streamlining of application and enrollment processes might work to increase take-up levels but could have dramatically differing effects in terms of how well they screen beneficiaries by, say, need. So, for example, nudges that help individuals overcome decision-making biases that lead them to fail to take up social benefits for which they are eligible will tend to improve the private welfare of those agents (Bertrand et al. 2006). But such nudges could attenuate or even undo efforts at targeting. Similarly, for the goal of targeting benefits to achieve particular outcomes, behavioral tendencies might lead to labeling effects, whereby, for example, benefits labeled as child benefits end up flowing to children merely because of the label (Kooreman 2000). A nudge that leads individuals to treat such benefits as fungible might improve private welfare for recipients but reduce targeting on the outcome of interest.

Another case is illustrated by a set of results in health insurance. Here a social goal might be to avoid adverse selection. When biases mitigate selection, then they might serve social purposes, even as they undermine private welfare. For example, in an employer-provided health insurance system with defaults, there is evidence that individuals are likely to stick with the status quo and that this may mitigate selection problems (Handel 2009). This might leave some individuals less well off, but it may be beneficial overall. Or take the example of Medicare Part D. Random assignment to plans for people who fail to sign up seems at first glance clearly inferior to an intelligent assignment scheme. It seems clear that smart choice architecture in this program could improve welfare at the individual level (Kling et al. 2012). However, random assignment in this case might lead to better pooling outcomes.

One final illustration of this principle comes from taxation. Evidence suggests that people may not respond optimally to nonsalient taxes either because of inattention or other sources of error (Chetty et al. 2009, Saez 2010). Policy makers may be able to use the salience of corrective taxes to manipulate their incidence in ways that make such taxes less regressive (Goldin & Homonoff 2010). In the case of tax compliance, biases or inaccurate perceptions about audit probabilities, for example, might lead to improved compliance (Dhami & al-Nowaihi 2007, 2010). Setting policy to perpetuate such errors may be socially beneficial.

7. TARGETING OF BEHAVIORAL AGENTS

How do we think about policy when people have differing behavioral tendencies? To analyze, suppose there are both standard and behavioral agents in the population, where behavioral agents are in proportion $\alpha \in (0,1)$, and define total welfare in the population to equal $W = (1-\alpha)W^S + \alpha W^B$. Imposing Assumption 1, totally differentiate W with respect to t and normalize

$$\begin{aligned} M_W(t) \equiv \frac{dW}{dt} \Big/ \frac{dW}{dY} &= (1-\alpha)A^{S'}(t)[tg'(R(t)) - \ell] + \alpha A^{B'}(t)[tg'(R(t)) - \ell - \tilde{\varepsilon}(t)] \\ &\quad + \left(\frac{g'(R(t))E[U'(C)] - E[U'(C)|a=1]}{E[U'(C)]} \right) \cdot A(t) \\ &= (1-\alpha)A^{S'}(t)[t + ME(t)] + \alpha A^{B'}(t)[t + ME(t) + MI(t)] \\ &\quad + TV(t) \cdot A(t), \end{aligned} \tag{9}$$

where we now let $MI(t) = -\tilde{\varepsilon}(t) = -\frac{U(Y+T-A\ell) - U(Y+T-A\ell-\varepsilon(b^B(t),t))}{E[U'(C)]}$.[14]

Equation 9 illustrates that, absent any other distortions, the presence of standard agents does not affect the conclusion that the optimal tax should be nonzero to help counteract

[14]We can easily extend the analysis to incorporate the heterogeneity of biases among behavioral agents, as considered by O'Donoghue & Rabin (2006). Suppose instead of there just being standard and behavioral individuals, there are N groups of individuals, $j = 1,2,\ldots,N$, with associated error maps $\varepsilon^j(b,t)$. Proportion α^j of individuals are of type j. In this case, we have $\frac{dW}{dt} / \frac{dW}{dY} = \sum_{j=1}^{N} \alpha^j A^{j'}(t)\left[t + ME(t) + MI^j(t)\right] + TV(t) \cdot A(t)$, where $MI^j(t) = -\tilde{\varepsilon}^j(t) = -\frac{U(Y+T-A\ell) - U\left(Y+T-A\ell-\varepsilon^j\left(b^B(t),t\right)\right)}{E[U'(C)]}$.

biases, so long as some behavioral agents are responsive to the tax (O'Donoghue & Rabin 2006). When $U'' = 0,\ \ell = 0$, and $g' = 1$, Equation 9 reduces to

$$M_{\mathbb{W}}(t) = A'(t)t - \alpha A^{B'}(t)\varepsilon(t). \tag{10}$$

Because the right-hand side of Equation 10 takes on the sign of $-A^{B'}(0)\varepsilon(0)$ at $t = 0$, the optimal tax is nonzero whenever behavioral agents are elastic and biased at $t = 0$. The intuition is that a marginal change in the tax away from zero does not have a first-order effect on standard agents' welfare because they are optimizing, but it does have a first-order effect on behavioral agents' welfare because the marginal behavioral agent is not truly indifferent between taking and not taking the action. O'Donoghue & Rabin (2006) make this point in the context of the argument that sin taxes should optimally be positive on goods that present-biased agents overconsume, such as cigarettes, even if a substantial fraction of the population does not suffer from self-control problems.[15]

Conversely, the presence of standard agents does affect the size of the optimal tax, as well as the relationship between the size of the tax and the size of behavioral agents' marginal internality. Continue to assume $U'' = 0,\ \ell = 0$, and $g' = 1$. Fixing $A'(t)$,

$$\frac{\partial M_{\mathbb{W}}(t)}{\partial \alpha} = -A^{B'}(t)\varepsilon(t), \tag{11}$$

which takes on the sign of $\varepsilon(t)$. Equation 11 indicates that a greater presence of behavioral agents in the population raises the marginal benefit to changing the tax to help correct their internalities. The intuition is that such a change helps improve behavioral agents' decisions, while distorting standard agents' choices. The greater the fraction of behavioral agents, the more this trade-off tilts toward considering the decisions of behavioral agents. In domains in which behavioral errors might be relatively common, such as in preparing adequately for retirement, the benefits to policies that address such errors—payroll taxes to fund Social Security, or restricted choice sets for investment levels or allocations—become larger relative to their welfare costs.

More generally, the number of behavioral agents who are on the margin matters for the determination of the size of the optimal tax, not just the extent to which behavioral agents are biased. From Equation 9, we see that what is important is $|\alpha A^{B'}(t)MI(t)|$, not $|MI(t)|$. For this reason, the optimal tax need not be monotonic in the "size" of the bias: More biased agents may be more inframarginal at tax rates in some relevant neighborhood.[16] Returning to Example 1, suppose that agents decide whether to consume some product that results in an immediate benefit but a delayed cost, such as cigarettes or potato chips. In this case, $\varepsilon(t) = (1 - \beta)h$ and $A^{B'}(t) = -\tilde{f}(p_0 + t + \beta h)$. In many applications, it is possible that, when t is small, $\tilde{f}(p_0 + t + \beta h) \approx 0$ for sufficiently small β. How many people would be close to indifferent to consuming chocolate at low prices if they were not sufficiently worried about caloric intake? In this case, $|\alpha A^{B'}(t)MI(t)| = \alpha\tilde{f}(p_0 + t + \beta h)(1 - \beta)h$ can fall in $1 - \beta$, reducing the benefits to taxing the good. To get grossly present-biased agents to

[15] Similarly, Sheshinski (2010) argues that restricting investment choices (e.g., in a privatized version of Social Security) might yield benefits to behavioral agents that outweigh the costs to perfect optimizers.

[16] O'Donoghue & Rabin (2006, proposition 2) provide certain conditions under which the size of the optimal sin tax does in fact increase in the "prevalence" of self-control problems in the population. Allcott et al. (2011) provide examples illustrating that the size of the optimal tax on energy-inefficient appliances may decrease in the degree to which behavioral agents are inattentive to future energy costs.

consume less chocolate, the tax may need to be so high that it severely reduces the demand of people who correctly trade off the costs and benefits of chocolate consumption.

To take another example, consider subsidies to energy-saving behaviors that pay off only over time, such as subsidies to weatherization programs. Such programs may simply target those who were already attentive to energy costs, or those who are sufficiently future oriented (Allcott et al. 2011). That is, such a program might fail to target those individuals who are behavioral with respect to energy consumption, and who could potentially most efficiently cut back on energy use. They are the first people we want such a policy to target, but the least likely to respond.

7.1. Matching the Price to the Bias

When there are multiple levers available to help correct behavioral agents' mistakes, Equations 9 and 10 indicate that, all else equal, care should be taken to use those that minimally affect standard agents' decisions.[17] We return to the case in which the only rationale for government intervention are internalities ($U'' = 0,\ \ell = 0,\ g' = 1$), and agents, some fraction of whom are present-biased, are deciding whether to consume a product with a delayed cost. Suppose the government has two options: It can levy a tax experienced by the agents today or tomorrow. For example, the government could levy a tax on potato chips or obesity. Whether the tax is experienced today or tomorrow does not affect standard agents' decisions, but behavioral agents discount tomorrow's tax by factor β: If the tax is experienced tomorrow, behavioral agents consume the good if $v - \beta h > p_0 + \beta t$. Starting from a situation of no tax, the marginal benefit of levying a small tax today is $1/\beta$ times greater than the marginal benefit to levying a small tax tomorrow: $M_W(0) = \alpha\tilde{f}(p_0 + \beta h)(1 - \beta)h$ if the tax is levied today and $M_W(0) = \beta\alpha\tilde{f}(p_0 + \beta h)(1 - \beta)h$ if the tax is levied tomorrow.

A related implication is that rationally equivalent prices can target differentially in the presence of behavioral agents. For example, two subsidies of equivalent net present value would be expected to have similar impacts with rational agents. But with present-biased agents, for example, front-loaded subsidies may have different effects than back-loaded subsidies of equivalent net present value. As a result, the mix of individuals who respond to otherwise equivalent price levers can be very different.

One domain in which the differential impacts of equivalent prices appear to manifest is health insurance policy. A core challenge in health insurance policy is mitigating moral hazard. In order to realize this, policy is concerned with the design of cost-sharing elements, such as copays and coinsurance, that work to align private incentives with social costs. In the standard model, copays do not need to distinguish within classes that have the same demand elasticity. Individuals are presumed to cut back on the demand for care in response to the copay in an efficient manner. In a behavioral model, however, this need not be the case. For example, individuals might have mistaken beliefs about the benefits

[17]Even better would be if the policy maker could "tag" people more likely to be behavioral agents and target interventions to them. For example, suppose that susceptibility to defaults with respect to retirement savings (Madrian & Shea 2001) indicates that someone is relatively present biased. If so, the default policy could be designed with present-biased agents in mind. Failing to adhere to drug (DiMatteo 2004) or treatment regimens, such as diabetics getting recommended blood tests (McGlynn et al. 2003), can be thought of as revealing someone as a behavioral type in this domain, and interventions can be targeted and designed accordingly. Tagging is likely to be imperfect, however, in which case it is still important to consider the relative degree to which different levers affect behavioral agents' and rational agents' decisions.

associated with different medical treatments. Or they may be present biased so that immediate costs associated with care weigh heavily.

With behavioral agents, then, individuals respond to cost sharing in potentially inefficient ways. Evidence that demand for medical care is very price sensitive at copays below cost does not necessarily indicate moral hazard, but rather may reflect error. In the standard model, this price sensitivity signals that a significant fraction of demand is for low-valued care, implying a greater benefit to increasing the copay. However, empirically, there is evidence that demand for "effective care" is often as elastic as demand for "ineffective care" (Baicker et al. 2011). For example, to understand the welfare impact of higher copays on cholesterol-lowering medications, it is informative that Goldman et al. (2006) estimate that a \$10 increase in copayments drives similar reductions in use among those with high risk (and thus high health benefits) as those with much lower risk. This is part of a literature that finds that individuals may cut back in inefficient ways (Lohr et al. 1986, Gross et al. 2006, Hsu et al. 2006, Chandra et al. 2010).

Consistent with these findings, evidence suggests that cost sharing that makes distinctions within classes of care that have the same elasticity can be more effective at targeting the right individuals to cut back on the right kind of care. Value-based insurance design, in which cost-sharing components of health insurance are set so as to reflect the cost-effectiveness of alternative medical treatments, can potentially lead to more efficient responses to cost sharing (Chernew et al. 2007, Choudhry et al. 2010). The key feature of value-based insurance design is not the differential overall level of cost sharing, but rather its structure.

Another application of the principle that the structure of prices can change how prices target individuals comes from incentives for educational attainment. The benefits to education are substantial, but often only accrue with a great delay. Behavioral biases such as present bias might therefore lead individuals to underinvest in education. In this environment, policies that operate on long-term incentives, such as subsidizing student loans, might not target the individuals of interest. Otherwise equivalent policies that structure subsidies differently in time, however, might. This is one way to view conditional cash transfers for outcomes in educational attainment (Angrist et al. 2010, Fryer 2010). By providing immediate incentives, such subsidies might more effectively target individuals.

With detailed knowledge of the underlying psychology, it may even be possible for the policy maker to use a combination of levers in a way that affects only behavioral agents' decisions. To take an extreme case, consider the present-bias example from above. In this example, levying tax $t_1 = h$ experienced today and tax $t_2 = -h$ experienced tomorrow implements the first best. To see this, standard agents would choose to consume the good if $v - h > p_0 + t_1 + t_2 \iff v - h > p_0 + h - h \iff v - h > p_0$, whereas behavioral agents would choose to consume the good if $v - \beta h > p_0 + t_1 + \beta t_2 \iff v - \beta h > p_0 + h - \beta h \iff v - h > p_0$. To take a related example, an effective policy to encourage energy-efficiency upgrades might be to cover part of the upfront cost of the upgrade and amortize that cost over several years on the homeowner's energy bills (Allcott et al. 2011).[18] It is this logic that gives an underpinning to libertarian paternalism, which emphasizes the virtue of nudges that, by definition, barely affect standard agents' decisions (Camerer et al. 2003, Thaler & Sunstein 2003). So, for example, defaults in 401(k) enrollment might affect retirement outcomes for behavioral agents but, given the relatively trivial and essentially

[18]Many utility companies around the country have instituted similar "on-bill financing" programs.

symmetric costs associated with opting either in or out of such plans, do not significantly affect rational savers' decisions.

7.2. Matching the Nudge to the Bias

The mere fact that a nudge works, in the sense that it changes how people behave, is not sufficient to conclude that the nudge is operating in ways that improve social welfare. Nudges, like prices, can be blunt instruments and can operate on mismatched margins. As a practical result, different nudges will target differentially.

A good example of this principle is in retirement security policy. Defaults work at getting more individuals to participate in savings plans (Madrian & Shea 2001). But it is not clear how the nudge—enrollment default—relates to the bias we want to correct—procrastination in asset accumulation. When defaults in retirement savings plans lead individuals who were previously not saving in such plans to participate, it is not clear that the propensity to be sensitive to the default maps neatly back onto whether that individual actually experiences a benefit from the move. Some individuals influenced by the default may truly have been better off not saving. Conversely, some of the individuals not influenced by the default may in fact be among those who would have experienced a benefit. A bit more formally, we do not know whether $\frac{dA_n^B}{dn} MI_n > 0$.

Moreover, different nudges are available to encourage retirement saving, and they may draw in different groups. Knowledge of the underlying psychology of behavioral agents could suggest levers that more effectively match the nudge to the bias. Defaults in some ways implicitly embody a model in which behavioral agents suffer from status-quo bias (Samuelson & Zeckhauser 1988). This bias could arise from present bias; i.e., behavioral agents overweigh the (minor) immediate hassle cost of switching from the default option relative to the future benefit of saving properly for retirement (Carroll et al. 2009). Less obvious, this bias could also arise from a form of limited attention, if the proclivity to pay attention depends on the cost of failing to attend (Bernheim et al. 2011). Both these interpretations suggest that active choice (Carroll et al. 2009), whereby people are "forced" to decide on whether to participate in 401(k) plans, would lead behavioral agents to make efficient decisions.[19] (Under different interpretations, however, active choice may not be as desirable, for example, if agents underweigh the benefits of participation when they are forced to choose.)

From a different perspective, consider nudges that address environmental externalities. Take, for example, the now substantial literature finding that social appeals can lead to reduced energy consumption, and a reduction in the associated externalities (Schultz et al. 2007, Allcott 2011). Interventions often take the form of including information on how the level of each household's energy use compares with that of their neighbors on residential energy statements. These interventions tend to reduce energy consumption. But the mechanism, social comparison, seems unlikely to be closely related to any underlying bias that might drive the overconsumption of energy. In this sense, the nudge, although effective, and possibly desirable from a social point of view, does not necessarily reduce internalities. It operates in some respects like a regulation, rather than a price: It is an effective instrument for moving the quantity of interest, but blunt.

[19] Of course, if agents face real costs to making decisions (e.g., time costs), then defaults may be preferable to active choice (Carroll et al. 2009).

8. RETHINKING MEASUREMENT

How can we actually measure the welfare impact of a tax change? When the government changes the tax, there are two effects that influence welfare: Resources are redistributed and choices change. We are particularly interested in the latter effect. To gauge the choice consequences in any application, we need to know first how many people are marginal and second who is marginal, i.e., what their type is (in this model, b, their private benefit). The first is, of course, the easiest quantity to measure empirically: the demand response to a tax change. The second appears to be significantly more challenging because it requires us to know something that is effectively private to the individual: their type.

8.1. Standard Model: Demand Response Forms a Sufficient Statistic

A key insight of the standard framework is that we actually know the private value of marginal agents by virtue of them being marginal: As shown above, demand coincides with marginal private benefit. This insight implies a powerful sufficient statistic result (Chetty 2009). Consider the first term of Equation 4, which captures the welfare impact of marginals' decisions:

$$A^{S'}(t)[MSB - MSC]^{S}(t) = A^{S'}(t)[t + ME(t)]. \tag{12}$$

This term equals the number of people who are at the margin multiplied by the difference between the marginal social benefit and marginal social cost of their taking the action: the marginal social distortion. Equation 12 omits the private benefit of the individual because we know that this benefit simply equals the price p. As a result, so long as the return on government investment (g') and the size of the externality (ℓ) are known, the sensitivity of demand is a sufficient statistic for measuring the welfare impact of the reduction in demand (Chetty 2009).

For this reason, much of empirical public finance is concerned with estimating and drawing policy conclusions from the sensitivity of demand. This is best illustrated through examples.

In the case of simple commodity taxation ($\ell = 0$, $g' = 1$), Equation 12 becomes $tA^{S'}(t)$ (Harberger 1964).[20] This equation reflects the standard textbook result that the deadweight loss from commodity taxation increases in the sensitivity of demand. From this theoretical basis, empirical public finance estimates the responsiveness of economic activity to various taxes and draws conclusions about their relative efficiency. For example, Hausman (2000) uses estimates of the price elasticity of cell phone services to calculate the social cost of taxes on those services.[21]

Knowledge of the sensitivity of demand is likewise useful in analyzing the welfare impact of expanding social insurance coverage. Consider the case of health insurance. With uncertainty about future health status and the associated costs of treatment, there are welfare gains from health insurance coverage. But there are also social costs, in the form of the moral hazard that insurance creates; when individuals do not face the full marginal

[20]In the case of corrective taxation, Equation 12 becomes $-A^{S'}(t)\ [\ell - t]$, which captures the standard result that, although the reduction in activity caused by taxation is often distortionary, small taxes ($t < \ell$) are welfare-improving when there are negative externalities.

[21]Note that the larger, parallel literature on income taxes, looking at the elasticity of taxable income with respect to the tax rate (as in Feldstein 1999 and Gruber & Saez 2002) proceeds from a similar logic.

costs of medical treatment decisions, they may consume medical care at inefficiently high levels (Arrow 1963, Pauly 1968, Zeckhauser 1970). Consider the welfare impact of reducing demand through increasing the copay p. If the cost of getting treated is $p_{0,}$ then the tax is the difference between the copay p and the cost p_0 (i.e., $t = p - p_0$). In this case, Equation 12 becomes $A^{S'}(p)[p - p_0]$ (assuming $\ell = 0$ and $g' = 1$). As long as the copay is below cost, the welfare impact of the reduction in demand is positive because it reduces moral hazard. The magnitude of this welfare impact is proportional to the sensitivity of demand.[22]

Estimating the responsiveness to copays is thus a central focus of much of the empirical literature, as it gives guidance on the optimal level of insurance coverage. The touchstone empirical work in health insurance is the RAND Health Insurance Experiment (Newhouse et al. 1981), which evaluated the responsiveness of utilization to the terms of coverage. This tradition carries through to more recent work, such as that by Finkelstein et al. (2011), who look at how individuals respond to public health insurance coverage.[23]

In short, across all these examples, the rational framework proves so useful for a simple reason. Revealed preference generally (and the envelope theorem specifically) means that the demand response alone goes a long way toward estimating the optimal policy. The analysis becomes more complicated when agents are biased.

8.2. Behavioral Model: Demand Response Is Not Enough

A consequence of the fact that demand no longer reflects the marginal private benefit in the behavioral model is that the demand response alone can be misleading when assessing the welfare effects of raising the tax: It is necessary to understand the marginal behavioral error. Consider the first term of Equation 6, which captures the welfare impact of behavioral agents' decisions:

$$A^{B'}(t)[MSB - MSC]^{B}(t) = A^{B'}(t)[t + ME(t) + MI(t)]. \tag{13}$$

In applications in which the marginal social distortion is a known function of the tax assuming standard agents [i.e., $ME(t)$ is known], it is often not a known function of the tax with behavioral agents because the marginal internality—$MI(t)$—needs to be estimated. The sensitivity of demand is then no longer a sufficient statistic for measuring the welfare impact of the reduction in demand. Not taking this fact into account can lead a policy maker to interpret data on the demand response in a way that supports incorrect conclusions about the benefits or costs of raising the tax.

Suppose for a health insurance policy we see a large demand response to a marginal increase in drug copays that start below cost. We would normally view this reduction as beneficial as the marginal person must value the drug at roughly the copay, which lies below the social cost. In other words, dissuading people to use those drugs is a good thing: They were of low value to them. Suppose, however, that some people have a tendency to

---

[22]Of course, there is also a cost of increasing the copay—diminished insurance value—which is captured by the second term of Equation 4, the term left out of Equation 12.

[23]Note too that unemployment insurance, which replaces a fraction of wages to the covered unemployed, presents an analytically analogous policy problem, as indicated in work such as Baily (1978) and Chetty (2008). The primary trade-off is balancing the consumption-smoothing benefits that unemployment insurance provides against the moral hazard it can create in job search. And here, too, a large empirical literature informs the design of unemployment insurance based in large part on estimating the demand response. For example, Meyer (1990) and Katz & Meyer (1990) look at how spell length responds to the generosity of benefits, in terms of either level or duration.

act as if they undervalue certain drugs, perhaps because they treat a chronic, asymptomatic, condition (e.g., hypertension). Then those who were marginal at the copay, those who reduce demand in response to a copay, may actually be people for whom the value of utilization is very high, above the social cost. Dissuading them from using the drugs is a bad thing (Baicker et al. 2011).

8.3. Supplementing Demand Responses to Assess Welfare

If knowledge of the demand response is not enough to assess the choice consequences of a policy change, then how do we do it? This is a central empirical challenge introduced by the behavioral approach to policy.

We return to the simplest case in which people have homogeneous behavioral biases. Assuming $U'' = 0$, the marginal internality can be expressed as

$$\begin{aligned} MI(t) &= b^B(t) - p(t) \\ &= b^B(t) - (p_0 + t), \end{aligned} \tag{14}$$

where the second line of Equation 14 follows from Assumption 1. Although we have taken p_0 to be exogenous, we sometimes highlight how certain quantities depend on its value by writing them as functions of p_0 [e.g., by writing $A^B = A^B(p_0, t)$ or $A^S = A^S(p_0, t)$]. Recall that $A^S(p, 0) = 1 - F(p)$ equals the number of standard types who take the action given a total price of taking the action of p, and let $P^S(A) = (A^S)^{-1}(p, 0)$ denote the corresponding "inverse demand function." We then have $p_0 + t = P^S(A^S(p_0, t))$ and $b^B(p_0, t) = P^S(A^B(p_0, t))$. With these identities, we can rewrite Equation 14 as

$$MI(p_0, t) = P^S(A^B(p_0, t)) - P^S(A^S(p_0, t)). \tag{15}$$

8.3.1. Using knowledge of when choices better reflect true preferences. Knowledge of what drives behavioral errors can be used to estimate ε. To illustrate, suppose all errors stem from misoptimization with respect to taxes: $A^B(p', 0) = A^S(p', 0)$ for all p'. Chetty et al. (2009) demonstrate that Equation 15 can then be approximated as a function of demand responses to prices and taxes. To see this, suppose A^B is approximately linear and separable in p_0 and t. Then Equation 15 can approximately be expressed as

$$MI(p_0, t) \approx -t + \frac{dP^S}{dA}\left(A^B(p_0, t) - A^S(p_0, 0)\right) = -t + \frac{dP^B}{dA}\frac{dA^B}{dt} \cdot (t) = -t(1 - \theta), \tag{16}$$

where $\theta = \frac{dA^B}{dt} / \frac{dA^B}{dp}$. Alternatively, Equation 15 demonstrates that perfect knowledge of the demand function absent taxes, $A^B(p, 0)$, together with the level of demand at (p_0,t) is sufficient to calculate $MI(p_0, t)$ when tax salience is the only bias. In this case, it would be possible to invert $A^B(p, 0)$ to recover the only unknown, $P^S(A^B(p_0, t))$.

This illustrates a more basic point: If a positive model of behavior indicates that choices made under some condition (e.g., absent taxes) reveal true preferences, then knowledge of the demand function under that condition (together with the observed level of demand) is enough to recover the marginal behavioral error.[24] As Chetty et al. (2009) point out, this

[24]Alternatively, if a positive model of behavior indicates that moving to some condition results in choices that are closer to optimal, then knowledge of the demand function under that condition can be used to bound the marginal behavioral error (Baicker et al. 2011).

observation can be viewed as an application of Bernheim & Rangel's (2009) framework for choice-based behavioral welfare economics to applied policy problems. Roughly speaking, if choices made in one situation conflict with those made in another, it may be possible to use a positive model of behavior to select those situations that are "welfare relevant." If welfare-relevant choices further satisfy standard axioms, then familiar techniques can be used to recover preferences. To take another example, when the behavioral error results solely from the simple form of present bias laid out above, standard and behavioral agents' choices agree when a tax of $t_2 = -h$ is experienced tomorrow for taking the action today.[25] Under such a tax, all net benefits and costs to taking the action are in the present, so any underweighing of future benefits does not bias choice. Fixing $t_2 = -h$, demand as a function of $p_0 + t_1$ can then be used to back out ε.

Similarly, with externalities, there is a divergence between how demand for energy-using durables such as appliances and cars responds to changes in the direct purchase price and how it responds to changes in usage costs influenced by the price of energy (Hausman & Joskow 1982, Allcott & Wozny 2010). We might take the direct response to purchase prices as roughly tracing the "true demand."

8.3.2. Using nudges. There is an interesting feature of the sales tax example, in which the behavioral agent underweighs nonsalient sales taxes by factor $\theta \in (0, 1)$, so $MI(t) = -\varepsilon(t) = -(1-\theta)t$. Fixing $p = p_0 + t$, the size of the tax t affects the choices of behavioral agents (as ε is affected) but not standard ones. Thus, fixing p, the size of the sales tax fits the definition of a nudge. Likewise, fixing p, the size of the tax tomorrow fits the definition of a nudge when the behavioral error stems from present bias. In this case, the behavioral error can be expressed as $\varepsilon = (1-\beta)(h + t_2)$, which is influenced by the size of t_2, fixing $p = p_0 + t_1 + t_2$.

Nudges can be used to help estimate who is at the margin. Let the total demand of behavioral agents given nudge n and total price p be denoted by $A_n^B(p)$, and the corresponding inverse demand function be denoted by $P_n^B(A)$. Suppose a positive model of behavior indicates that there is some nudge, $n = *$, under which standard and behavioral agents' decisions are perfectly aligned. In the case of tax salience, the nudge would be $n = t = 0$. In the case of present bias, the nudge would be $n = t_2 = -h$. In the case of retirement savings decisions, the nudge might be active choice (Carroll et al. 2009). In the case of asset allocation, the nudge might be one that re-expresses the portfolio choice in terms of the implied distribution of some relevant outcome (Benartzi & Thaler 2002). The marginal internality given some nudge n and total price p can be calculated by examining behavior under nudge $n = *$. In particular, Equation 14 can be expressed as

$$MI_n(p) = P_*^B(A_n^B(p)) - p. \tag{17}$$

Equation 17 highlights that knowledge of the demand function under nudge $n = *$ is enough to recover the marginal behavioral error. An alternative way of looking at this is that the error can be identified by comparing demand responses to nudges with demand responses to prices, analogous to comparing demand responses to taxes and prices in the tax

[25]To see this, note that standard agents take the action if $v - h > p_0 + t_1 + t_2$, whereas behavioral agents take the action if $v - h + \underbrace{(1-\beta)(h + t_2)}_{=\varepsilon} > p_0 + t_1 + t_2$. Clearly, $\varepsilon = 0$ when $t_2 = -h$.

salience example. We can re-express $P_*^B(A_n^B(p))$ as the value $\tilde{p}$ that satisfies $A_*^B(\tilde{p}) = A_n^B(p)$. Equivalently, $\tilde{p}$ solves

$$\underbrace{A_*^B(\tilde{p}) - A_*^B(p)}_{\text{demand response to price change}} = \underbrace{A_n^B(p) - A_*^B(p)}_{\text{demand response to nudge}}. \tag{18}$$

Thus, to find $P_*^B(A_n^B(p))$ and recover $MI_n(p)$ through Equation 17, it suffices to find the price at which the demand response to moving to that price (fixing the nudge at $n = *$) is equal in magnitude but opposite in sign to the demand response to moving to nudge $n = *$ (fixing the price at p).

This insight could potentially be applied, for example, in work such as that by Bertrand et al. (2010), who observe the response to loan offers with independently varied nudges and interest rate terms. Somewhat similarly, studies such as those by Ashraf et al. (2006), who observe the impact of a nudge in relation to an interest rate, and Duflo et al. (2006), who vary their presentation nudge along with the subsidy rate, could potentially inform the marginal internality.

8.3.3. Using a structural model of utility and behavior. Even if there does not exist a situation or nudge under which choices reflect true preferences, one can make progress in estimating or bounding the marginal internality with the aid of structural assumptions about utility and behavior (Beshears et al. 2008a). To take one example, Gruber & Koszegi (2001) write a model of cigarette-purchase decisions in which the marginal internality can be expressed as a function of the future health costs of smoking (in utility terms), as well as the degree of present bias β. To calibrate the optimal tax per pack of cigarettes, they make assumptions linking (estimable) health consequences of smoking to utility costs and make assumptions about the degree of present bias.[26] Similarly, Baicker et al. (2011) model health treatment choices, in which the marginal internality is the wedge between the price (copay) of treatment and the marginal health value of treatment (in dollars). They bound the marginal internality for certain treatment choices (e.g., the decision of whether to fill a prescription for antihypertensives) by bounding the health costs of not getting treated (e.g., of failing to fill a prescription for antihypertensives).

8.3.4. Using survey data. Another data source is survey data. In the case of savings rates, one commonly cited piece of evidence for undersaving is that people report wanting to save more in the coming year than they do (Choi et al. 2002, 2006). The criticisms of such data are well described elsewhere (see, e.g., Bertrand & Mullainathan 2001). Yet these data can play an important role in behavioral public finance. We see two ways they can be incorporated into our framework. The first is to presume (in some cases at least) that survey responses reflect what people truly want. In our framework, this would be tantamount to presuming that responding to survey questions is a decision frame that reveals true hedonic preference (or more realistically approximates true preference). Those who use survey data to build the case for mistaken choices take this approach. A related approach is to use self-reported hedonic measures (such as self-reported well-being) to gauge the welfare impacts of policy. For example, Gruber & Mullainathan (2005) use self-reported happiness measures to argue that cigarette taxes can improve welfare.

[26]Other studies structurally estimate the degree of present bias (e.g., Laibson et al. 2006).

A second approach uses survey data to rule out certain explanations of behavior. For example, we may be unsure about whether a patient is nonadherent to a high-return medication because of painful side effects or for a more behavioral reason (e.g., because he suffers from self-control problems). Suppose that in self-reported data patients never report side effects (of any kind) as a reason for nonadherence. This may give us some confidence in ruling out this alternative as an important driver of nonadherence. And if this were the primary alternative to a behavioral explanation, this would increase our confidence that the choice is in error. To more formally incorporate the role of these and other types of nonchoice data, we could model the policy maker as a Bayesian who is initially uncertain about the positive model driving choice. Armed with a prior belief over models and a likelihood function of data (choice, survey, and otherwise) given underlying models, she can use the data to update her beliefs. Under such a framework, survey data can help shift the policy maker's belief about the nature of error, even if it is not fully diagnostic.

9. CONCLUSION

Our framework provides a reduced-form approach to understanding policy problems in traditional public finance. Yet behavioral economics has been applied more broadly to policy. An important avenue of future work would be to understand whether the reduced-form approach yields important insights elsewhere. Three areas stand out. First, we have not considered the interaction of markets and consumer biases. Gabaix & Laibson (2006), for example, show that competition does not necessarily remove firms' incentives to obfuscate product features to inattentive or myopic customers. One wonders whether results such as theirs might generalize to a broader set of biases. Second, in consumer finance, there has been an intense focus on how mistakes by consumers can translate into poor choice of credit, savings, or other financial products (Heideus & Koszegi 2010). In this area, the focus has typically been on hyperbolic discounting and limited attention. A reduced-form approach might allow a clearer understanding of how different financial contracts interact with biases more broadly, or of the benefits of regulating exploitive contract features. Finally, a very applied literature in development uses behavioral insights to design policy solutions to problems. Understanding how to frame these policy exercises within a broader model could also give empirical guidance.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review. The views expressed are those of the authors and do not necessarily represent those of the Director of the Consumer Financial Protection Bureau nor those of the staff.

ACKNOWLEDGMENTS

We thank Jeff Kling for conversations that inspired some of these ideas, as well as David Laibson, Richard Thaler, and participants in the BEAM 2009 conference for helpful comments.

LITERATURE CITED

Allcott H. 2011. Social norms and energy conservation. *J. Public Econ.* 95:1082–95

Allcott H, Mullainathan S, Taubinsky D. 2011. *Externalities, internalities, and the targeting of energy policy.* Work. Pap., Harvard Univ.

Allcott H, Wozny N. 2010. *Gasoline prices, fuel economy, and the energy paradox*. Work. Pap., Mass. Inst. Technol.

Angrist J, Oreopoulos P, Williams T. 2010. *When opportunity knocks, who answers? New evidence on college achievement awards*. NBER Work. Pap. 16643

Arrow KJ. 1963. Uncertainty and the welfare economics of medical care. *Am. Econ. Rev.* 53:941–73

Ashraf N, Karlan D, Yin W. 2006. Tying Odysseus to the mast: evidence from a commitment savings product in the Philippines. *Q. J. Econ.* 121:635–72

Atkinson AB, Stern NH. 1974. Pigou, taxation and public goods. *Rev. Econ. Stud*. 41(1):119–28

Baicker K, Mullainathan S, Schwartzstein J. 2011. *Choice hazard in health insurance*. Work. Pap., Dartmouth Coll.

Baily MN. 1978. Some aspects of optimal unemployment insurance. *J. Polit. Econ.* 10:379–402

Benartzi S, Thaler RH. 2002. How much is investor autonomy worth? *J. Finance* 57:1593–616

Bernheim BD, Fradkin A, Popov I. 2011. *The welfare economics of default options: a theoretical and empirical analysis of 401(k) plans*. Work. Pap., Stanford Univ.

Bernheim BD, Rangel A. 2009. Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics. *Q. J. Econ.* 124:51–104

Bertrand M, Karlan D, Mullainathan S, Shafir E, Zinman J. 2010. What's advertising content worth? Evidence from a consumer credit marketing field experiment. *Q. J. Econ.* 125:263–306

Bertrand M, Mullainathan S. 2001. Do people mean what they say? Implications for subjective survey data. *Am. Econ. Rev.* 91:67–72

Bertrand M, Mullainathan S, Shafir E. 2006. Behavioral economics and marketing in aid of decision-making among the poor. *J. Public Policy Mark*. 25(1):8–23

Beshears J, Choi J, Laibson D, Madrian B. 2008a. How are preferences revealed? *J. Public Econ.* 92:1787–94

Beshears J, Choi J, Laibson D, Madrian B. 2008b. The importance of default options for retirement saving outcomes: evidence from the United States. In *Lessons from Pension Reform in the Americas*, ed. SJ Kay, T Sinha, pp. 59–87. New York: Oxford Univ. Press

Bettinger EP, Long BT, Oreopoulos P, Sanbonmatsu L. 2009. *The role of simplification and information in college decisions: results from the H&R Block FAFSA experiment*. NBER Work. Pap. 15361

Camerer C, Issacharoff S, Loewenstein G, O'Donoghue T, Rabin M. 2003. Regulation for conservatives: behavioral economics and the case for 'asymmetric paternalism.' *Univ. Pa. Law Rev.* 151:1211–54

Carroll GD, Choi JJ, Laibson D, Madrian BC, Metrick A. 2009. Optimal defaults and active decisions. *Q. J. Econ.* 124:1639–74

Chandra A, Gruber J, McKnight R. 2010. Patient cost-sharing and hospitalization offsets in the elderly. *Am. Econ. Rev.* 100:193–213

Chernew ME, Rosen AB, Fendrick AM. 2007. Value-based insurance design. *Health Aff*. 26:W195–203

Chetty R. 2008. Moral hazard versus liquidity and optimal unemployment insurance. *J. Polit. Econ.* 116:173–234

Chetty R. 2009. Sufficient statistics for welfare analysis: a bridge between structural and reduced-form methods. *Annu. Rev. Econ.* 1:451–88

Chetty R, Looney A, Kroft K. 2009. Salience and taxation: theory and evidence. *Am. Econ. Rev.* 99:1145–77

Chetty R, Saez E. 2009. *Teaching the tax code: earnings responses to an experiment with EITC recipients*. NBER Work. Pap. 14836

Choi J, Laibson D, Madrian B. 2009. Reducing the complexity costs of 401(k) participation through quick enrollment. In *Developments in the Economics of Aging*, ed. DA Wise, pp. 57–85. Chicago: Univ. Chicago Press

Choi J, Laibson D, Madrian B, Metrick A. 2002. Defined contribution pensions: plan rules, participant decisions, and the path of least resistance. In *Tax Policy and the Economy*, Vol. 16, ed. J Poterba, pp. 67–114. Cambridge, MA: MIT Press

Choi J, Laibson D, Madrian B, Metrick A. 2006. Saving for retirement on the path of least resistance. In *Behavioral Public Finance: Toward a New Agenda*, ed. E McCaffrey, J Slemrod, pp. 304–51. New York: Russell Sage Found.

Choudhry N, Rosenthal MB, Milstein A. 2010. Assessing the evidence for value-based insurance design. *Health Aff.* 29:1988–94

Congdon W, Kling J, Mullainathan S. 2011. *Policy and Choice: Public Finance Through the Lens of Behavioral Economics*. Washington, DC: Brookings Inst.

Cronqvist H, Thaler RH. 2004. Design choices in privatized social-security systems: learning from the Swedish experience. *Am. Econ. Rev.* 94:424–28

DellaVigna S. 2009. Psychology and economics: evidence from the field. *J. Econ. Lit.* 47(2):315–72

DellaVigna S, Paserman MD. 2005. Job search and impatience. *J. Labor Econ.* 23:527–88

Dhami S, al-Nowaihi A. 2007. Why do people pay taxes? Prospect theory versus expected utility theory. *J. Econ. Behav. Organ.* 64:171–92

Dhami S, al-Nowaihi A. 2010. Optimal income taxation in the presence of tax evasion: expected utility versus prospect theory. *J. Econ. Behav. Organ.* 75:313–37

DiMatteo MR. 2004. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. *Med. Care* 42(3):200–9

Dobkin C, Puller SL. 2007. The effects of government transfers on monthly cycles in drug abuse, hospitalization, and mortality. *J. Public Econ.* 91:2137–57

Duflo E, Gale W, Liebman J, Orszag P, Saez E. 2006. Saving incentives for low- and middle-income families: evidence from a field experiment with H&R Block. *Q. J. Econ.* 121:1311–46

Fang H, Silverman D. 2004. On the compassion of time-limited welfare programs. *J. Public Econ.* 88:1445–70

Feldstein MS. 1999. Tax avoidance and the deadweight loss of the income tax. *Rev. Econ. Stat.* 81:674–80

Finkelstein A, Taubman S, Wright B, Bernstein M, Gruber J, et al. 2011. *The Oregon Health Insurance Experiment: evidence from the first year.* NBER Work. Pap. 17190

Fryer RG. 2010. *Financial incentives and student achievement: evidence from randomized trials*. NBER Work. Pap. 15898

Gabaix X, Laibson D. 2006. Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Q. J. Econ.* 121:505–40

Goldin J, Homonoff T. 2010. *Smoke gets in your eyes: cigarette tax salience and regressivity.* Work. Pap. 561, Ind. Relat. Sect., Princeton Univ.

Goldman DP, Joyce GF, Karaca-Mandic P. 2006. Varying pharmacy benefits with clinical status: the case of cholesterol-lowering therapy. *Am. J. Manag. Care* 12(1):21–28

Green JR, Hojman DA. 2009. *Choice, rationality and welfare measurement*. Work. Pap., Harvard Univ.

Gross CP, Andersen MS, Krumholz HM, McAvay GJ, Proctor D, Tinetti ME. 2006. Relation between Medicare screening reimbursement and stage at diagnosis for older patients with colon cancer. *JAMA* 296:2815–22

Gruber J, Koszegi B. 2001. Is addiction 'rational'? Theory and evidence. *Q. J. Econ.* 116:1261–303

Gruber J, Mullainathan S. 2005. Do cigarette taxes make smokers happier? *B.E. J. Econ. Anal. Policy* 5(1):4

Gruber J, Saez E. 2002. The elasticity of taxable income: evidence and implications. *J. Public Econ.* 84:1–32

Handel B. 2009. *Adverse selection and switching costs in health insurance markets: when nudging hurts*. Work. Pap., Univ. Calif., Berkeley

Harberger AC. 1964. The measurement of waste. *Am. Econ. Rev.* 54(3):58–76
Hastings JS, Weinstein JM. 2008. Information, school choice, and academic achievement: evidence from two experiments. *Q. J. Econ.* 123:1373–414
Hausman JA. 1979. Individual discount rates and the purchase and utilization of energy-using durables. *Bell J. Econ.* 10(1):33–54
Hausman JA. 2000. Efficiency effects on the U.S. economy from wireless taxation. *Natl. Tax J.* 53:733–42
Hausman JA, Joskow PL. 1982. Evaluating the costs and benefits of appliance efficiency standards. *Am. Econ. Rev.* 72(2):220–25
Heideus P, Koszegi B. 2010. Exploiting naivete about self control in the credit market. *Am. Econ. Rev.* 100:2279–303
Heutel G. 2011. *Optimal policy instruments for externality-producing durable goods under time inconsistency.* NBER Work. Pap. 17083
Hsu J, Price M, Huang J, Brand R, Fung V, et al. 2006. Unintended consequences of caps on Medicare drug benefits. *N. Engl. J. Med.* 354:2349–59
Kahneman D, Tversky A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–91
Kahneman D, Wakker P, Sarin R. 1997. Back to Bentham? Explorations of experienced utility. *Q. J. Econ.* 112:375–406
Katz LF, Meyer BD. 1990. The impact of the potential duration of unemployment benefits on the duration of unemployment. *J. Public Econ.* 41:45–72
Kling JR, Mullainathan S, Shafir E, Vermeulen L, Wrobel MV. 2012. Comparison friction: experimental evidence from Medicare drug plans. *Q. J. Econ.* 127:199–235
Kooreman P. 2000. The labeling effect of a child benefit system. *Am. Econ. Rev.* 90:571–83
Laibson D. 1997. Golden eggs and hyperbolic discounting. *Q. J. Econ.* 112:443–77
Laibson D, Repetto A, Tobacman J. 2006. *Estimating discount functions with consumption choices over the lifecycle*. Work. Pap., Harvard Univ.
Lipsey RG, Lancaster K. 1957. The general theory of second best. *Rev. Econ. Stud.* 24(1):11–32
Lohr K, Brook RH, Kamberg CJ, Goldberg GA, Leibowitz A, et al. 1986. Use of medical care in the Rand Health Insurance Experiment: diagnosis and service-specific analyses in a randomized controlled trial. *Med. Care* 24(9):S1–87
Madrian BC, Shea DF. 2001. The power of suggestion: inertia in 401(k) participation and savings behavior. *Q. J. Econ.* 116:1149–87
Mastrobuoni G, Weinberg M. 2009. Heterogeneity in intra-monthly consumption patterns, self-control, and savings at retirement. *Am. Econ. J. Econ. Policy* 1(2):163–89
McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, et al. 2003. The quality of health care delivered to adults in the United States. *N. Engl. J. Med.* 348:2635–45
Meyer BD. 1990. Unemployment insurance and unemployment spells. *Econometrica* 58:757–82
Newhouse JP, Manning WG, Morris CN, Orr LL, Duan N, et al. 1981. Some interim results from a controlled trial of cost sharing in health insurance. *N. Engl. J. Med.* 305:1501–7
O'Donoghue T, Rabin M. 1999. Doing it now or later. *Am. Econ. Rev.* 89:103–24
O'Donoghue T, Rabin M. 2006. Optimal sin taxes. *J. Public Econ.* 90:1825–49
Paserman MD. 2008. Job search and hyperbolic discounting: structural estimation and policy evaluation. *Econ. J.* 118:1418–52
Pauly MV. 1968. The economics of moral hazard. *Am. Econ. Rev.* 58:531–37
Rubinstein A, Salant Y. 2011. *Eliciting welfare preferences from behavioral datasets*. Work. Pap., New York Univ.
Saez E. 2009. Details matter: the impact of presentation and information on the take-up of financial incentives for retirement saving. *Am. Econ. J. Econ. Policy* 1(1):204–28
Saez E. 2010. Do tax filers bunch at kink points? *Am. Econ. J. Econ. Policy* 2(3):180–212
Samuelson W, Zeckhauser R. 1988. Status quo bias in decision making. *J. Risk Uncertain.* 1(1):7–59

Sandroni A, Squintani F. 2007. Overconfidence, insurance, and paternalism. *Am. Econ. Rev.* 97:1994–2004

Schultz PW, Nolan JM, Cialdini RB, Goldstein NJ, Griskevicius V. 2007. The constructive, destructive, and reconstructive power of social norms. *Psychol. Sci.* 18:429–34

Shapiro JM. 2005. Is there a daily discount rate? Evidence from the food stamp nutrition cycle. *J. Public Econ.* 89:303–25

Sheshinski E. 2010. *Optimal policy to influence individual choice probabilities*. Work. Pap., Hebrew Univ.

Spinnewijn J. 2009. *Unemployed but optimistic: optimal insurance design with biased beliefs.* Work. Pap., London Sch. Econ.

Spinnewijn J. 2012. *Heterogeneity, demand for insurance and adverse selection*. Work. Pap., London Sch. Econ.

Stephens M. 2003. '3rd of tha month': Do Social Security recipients smooth consumption between checks? *Am. Econ. Rev.* 93(1):406–22

Stephens M, Unayama T. 2008. *Can governments help households smooth consumption? Evidence from Japanese public pension benefits*. Work. Pap., Carnegie Mellon Univ.

Thaler RH, Benartzi S. 2004. Save More Tomorrow: using behavioral economics to increase employee savings. *J. Polit. Econ.* 112:164–87

Thaler RH, Sunstein CR. 2003. Libertarian paternalism. *Am. Econ. Rev.* 93(2):175–79

Thaler RH, Sunstein CR. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale Univ. Press

Wilde PE, Ranney CK. 2000. The monthly food stamp cycle: shopping frequency and food intake decisions in an endogenous switching regression framework. *Am. J. Agric. Econ.* 82(1):200–13

Zeckhauser R. 1970. Medical insurance: a case study of the trade-off between risk spreading and appropriate incentives. *J. Econ. Theory* 2(1):10–26

Annual Review of Economics

Volume 4, 2012

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Economics* articles may be found at http://econ.annualreviews.org