ELSEVIER

Journal of Public Economics 90 (2006) 1825–1849

JOURNAL OF PUBLIC ECONOMICS

www.elsevier.com/locate/econbase

Optimal sin taxes

Ted O'Donoghue [a,*,1], Matthew Rabin [b,1]

[a] *Department of Economics, Cornell University, 414 Uris Hall, Ithaca, NY 14853-7601, United States*
[b] *Department of Economics, 549 Evans Hall #3880, University of California, Berkeley, Berkeley, CA 94720-3880, United States*

Received 22 June 2005; received in revised form 6 March 2006; accepted 14 March 2006
Available online 3 May 2006

Abstract

We investigate "sin taxes" on unhealthy items, such as fatty foods, that people may (by their own reckoning) consume too much of. We employ a standard optimal-taxation framework, but replace the standard assumption that all consumers have 100% self control with an assumption that some consumers may have some degree of self-control problems. We show that imposing taxes on unhealthy items and returning the proceeds to consumers can generally improve total social surplus. Because such taxes counteract over-consumption by consumers with self-control problems while at the same time they naturally redistribute income to consumers with no self-control problems (who consume less), such taxes can even create Pareto improvements. Finally, we demonstrate with some simple numerical examples that even if the population exhibits relatively few self-control problems, optimal taxes can still be large.

Keywords: Hyperbolic discounting; Immediate gratification; Paternalism; Pigouvian taxation; Time inconsistency

"Never eat more than you can lift." — Miss Piggy

1. Introduction

We investigate the welfare effects of "sin taxes" on unhealthy items, such as fatty foods, that people may (by their own reckoning) consume too much of. The standard economic approach to taxation a priori assumes that there is no such "over-consumption", and hence the only reasons to

* Corresponding author.
E-mail addresses: edo1@cornell.edu (T. O'Donoghue), rabin@econ.berkeley.edu (M. Rabin).
URL's: http://www.people.cornell.edu/pages/edo1/ (T. O'Donoghue), http://emlab.berkeley.edu/users/rabin/ (M. Rabin).
[1] CB handles: "Puck Boy" and "Game Boy".

0047-2727/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpubeco.2006.03.001

tax commodities are to raise revenue, to correct externalities, or to redistribute wealth. If, however, people do exhibit such over-consumption – due to self-control problems or some other error – then the standard calculus of optimal taxation does not necessarily hold.[2]

We employ a standard framework to study optimal commodity taxation, but we replace the standard assumption that all consumers have 100% self control with an assumption that some consumers may have some degree of self-control problems — formalized as a time-inconsistent preference for immediate gratification. Using a social-welfare function that puts equal weight on all individuals, we show that imposing taxes on unhealthy items and returning the proceeds to consumers can generally improve social surplus. Moreover, we find that such taxes can even create Pareto improvements — in other words, such taxes need not involve helping people with self-control problems to the detriment of fully rational people. Finally, we demonstrate with some simple numerical examples that even if the prevalence of self-control problems in the population is relatively small, optimal sin taxes can still be large.[3]

In Section 2, we develop a framework to analyze optimal commodity taxation when agents might have self-control problems. We focus on a simple quasi-linear economy where, in addition to a composite good, there is a "sin good" – which we refer to as potato chips – that is enjoyable to consume but creates negative health consequences in the future.[4] Two basic results are immediate. First, self-control problems lead to over-consumption of potato chips. Intuitively, current consumption imposes a negative externality on one's future self — dubbed a "negative internality" by Herrnstein et al. (1993). Whereas people without self-control problems fully internalize this externality, people with self-control problems only partially internalize it. In particular, whereas optimal behavior gives full weight to any future health cost, the current self gives it only partial weight. This negative externality intuition immediately leads to our second basic result: With homogeneous consumers, a simple Pigouvian tax-and-transfer scheme (Pigou, 1920) can induce people to consume optimally.

Our primary interest, however, is the case in which there is population heterogeneity in both people's tastes and in their degree of self-control problems. Even in this case, the first-best outcome can be implemented with individual-specific taxes and transfers. But since such schemes are unrealistic, we investigate constrained-optimal taxation when the government is limited to using linear taxes and lump-sum transfers that are the same for all consumers.[5]

In Section 3, we analyze optimal taxation using a conventional social-welfare function that puts equal weight on all people. If there are no self-control problems in the population, then it

[2] Over the years, some researchers (e.g., Musgrave, 1959; Besley, 1988) have studied ''merit goods,'' for which the government has a different notion of individuals' optimal consumption than the individuals themselves have.

[3] Some of the ideas presented here appear in more preliminary form in O'Donoghue and Rabin (2003). Three other recent papers also study the welfare effects of sin taxes. Gruber and Koszegi (2001) also suggest that a Pigouvian tax could be used to counteract over-consumption due to self-control problems, and they conduct some simulations to assess the relevant magnitudes for optimal cigarette taxes. They do not address optimal taxation in a heterogeneous population, however, which is our primary focus. Gruber and Koszegi (2004) also study cigarette taxation in the presence of self-control problems, and they calibrate tax incidence for different income groups. But they do not derive optimal taxes, and once again they do not focus on population heterogeneity in self-control problems. In a somewhat different approach, Gruber and Mullainathan (2005) use survey data from Canada to provide empirical evidence that higher local cigarette taxes lead to increased happiness. For another example of tax analysis when taxpayers are boundedly rational, see Sheshinski (2002).

[4] Our analysis applies fully to chips both with and without ridges, and we believe our qualitative results extend to potato sticks, tator tots, and french fries.

[5] Our analysis is similar in spirit to analyses (following Diamond, 1973) of uniform taxation to correct non-uniform externalities.

would be optimal not to tax potato chips, because such taxes would merely distort the otherwise optimal potato-chip consumption. If instead some people have self-control problems, then it is indeed optimal to tax potato chips. Although taxes create consumption distortions for fully self-controlled people, such distortions are second-order relative to the benefits from reducing over-consumption by people with self-control problems.

In Section 4, we characterize Pareto-efficient taxation. Specifically, we investigate a constrained Pareto efficiency in which the government is restricted to uniform taxes and transfers. In this context, potato-chip taxes and the associated transfers have two effects on individuals' welfare: they affect the efficiency of potato-chip consumption, but they also redistribute income from those with high potato-chip consumption to those with low potato-chip consumption. Two results follow. First, if there is any taste heterogeneity, then even when there are no self-control problems in the population, small taxes are not Pareto-inefficient. Intuitively, for fully self-controlled individuals, small taxes create only second-order consumption distortions, while at the same time there are first-order income redistributions from people with a strong taste for potato chips to people with a weak taste for potato chips. Second, when there are self-control problems in the population, potato-chip taxes may, in fact, create Pareto improvements. People with self-control problems benefit because the taxes counteract their over-consumption. At the same time, because people with self-control problems consume more potato chips than people without self-control problems, income is naturally redistributed from people with self-control problems to people without self-control problems. Indeed, we show that, under quite reasonable conditions, it is possible to tax potato chips in a way that on average helps both people with self-control problems and those without. Under the less reasonable assumption that the variation in tastes for potato chips is sufficiently small, such taxes could help every single person.

Our formal results in Sections 3 and 4 are based on marginal arguments: we show that, when there are self-control problems in the population, infinitesimal departures from the policy that would be optimal given 100% rationality always increases average welfare and can sometimes make everyone better off. Yet we suspect that these results are not of marginal interest to economists. To highlight the potential importance of these results, in Section 5 we demonstrate with some simple numerical examples that, even if the prevalence of self-control problems in the population is relatively small, optimal taxes can still be significant.

Our formal analysis focuses on over-consumption due to a time-inconsistent preference for immediate gratification. In Section 6, we explore to what extent our conclusions would hold in other behavioral models of sin good consumption. Finally, we conclude in Section 7 by discussing the broader implications of our analysis, and also its limitations.

2. Model and basic results

Consider a simple consumption model of the form introduced by Pigou (1920). There are two goods, potato chips and a composite good. Both goods are produced with constant returns to scale, and we normalize units so that they have identical marginal costs. Moreover, we assume that both markets are competitive, and we normalize the price of the composite good to be one, which implies the marginal cost of each good is also one.[6] We assume that time is discrete, and that consumption (and production) occur in all periods.

[6] Given these assumptions, there are no distortions from mispriced goods — absent taxes, goods are priced at their constant marginal cost.

The essential feature of "sin goods" such as potato chips is that current consumption generates immediate enjoyment but future health costs or other negative consequences. To incorporate this feature, we assume that the person's instantaneous utility in period-t takes the quasi-linear form $u_t \equiv v(x_t;\rho) - c(x_{t-1};\gamma) + z_t$, where x_t and z_t denote, respectively, an individual's period-t consumption of potato chips and the composite good. The function $v(x_t;\ \rho)$ represents the immediate benefits from current potato-chip consumption, while $c(x_{t-1};\ \gamma)$ represents the negative health consequences from past potato-chip consumption.[7] We assume $v_x > 0$ and $v_{xx} < 0$, so that there are decreasing marginal benefits to consumption. We also assume $c_x > 0$, but we allow that there might be increasing, constant, or decreasing marginal health costs — that is, we allow $c_{xx} < 0$, $c_{xx} = 0$, or $c_{xx} > 0$.[8] The parameters ρ and γ capture population heterogeneity in tastes. In particular, we assume that $v_{x\rho} > 0$, so that a higher ρ corresponds to a higher marginal benefit from consumption; and we assume that $c_{x\gamma} > 0$, so that a higher γ corresponds to a higher marginal health cost from consumption.

We assume throughout that people cannot borrow or save. This assumption helps to clarify the basic logic of our results. In particular, it permits us to isolate the intratemporal distortion in potato-chip consumption without any confounding effects from intertemporal distortions in savings behavior. In a more general model, there might be interaction effects wherein distortions in saving behavior influence potato-chip consumption, or distortions in potato-chip consumption influence savings behavior. We suspect, however, that as long as potato-chip expenditures are small relative to people's income, such effects should be small.

How a person trades off the immediate benefits of potato-chip consumption against the future health costs depends on her intertemporal preferences. We assume that people might exhibit a tendency to pursue immediate gratification in a way that they themselves disapprove of in the long run.[9] Beginning with Laibson (1997), recent research on such present-biased preferences uses a simple and convenient functional form: a person's intertemporal preferences at time t are given by

$$U^t(u_t, \ldots,\ u_T) \equiv u_t + \beta \sum_{\tau=t+1}^{T} \delta^{\tau-t} u_\tau,$$

where u_τ is her instantaneous utility in period τ. This two-parameter model is a simple modification of the standard one-parameter, exponential-discounting model. The parameter δ represents standard time-consistent impatience; for $\beta = 1$ these preferences reduce to exponential discounting. The parameter β represents a time-inconsistent preference for immediate gratification, where $\beta < 1$ implies an extra bias for now over the future. To simplify our analysis, we assume throughout that there is no time-consistent discounting, or $\delta = 1$.[10]

[7] As will become clear, the fact that the negative consequences go only one period forward is not essential. Indeed, $c(x_t;\ \gamma)$ can be interpreted as the (exponentially) discounted sum of the future health costs due to period-t consumption.

[8] The combination of concave benefits and concave costs can of course create problems. Hence, whenever $c_{xx} < 0$ we impose the additional assumption that $v_{xx} - c_{xx} < 0$, which guarantees that behavior is well-behaved.

[9] For evidence that most humans do, indeed, have such a tendency, see Ainslie (1991, 1992), Ainslie and Haslam (1992a,b), Loewenstein and Prelec (1992), Thaler (1991), and Thaler and Loewenstein (1992). For a recent overview, see Frederick et al. (2002). This tendency is often referred to as "hyperbolic discounting".

[10] This model was originally developed by Phelps and Pollak (1968) in the context of intergenerational altruism. It has been used in recent years by numerous authors, including Laibson (1997,1998), Laibson et al. (1998), Angeletos et al. (2001), O'Donoghue and Rabin (1999a, 2001), Fischer (1999), Carrillo and Mariotti (2000), and Benabou and Tirole (2002).

Because we assume that the benefits and costs from period-t consumption are additively separable from the benefits and costs from consumption in any other period, the person effectively faces a series of independent decisions. In particular, in every period the person will choose her current consumption (x, z) to maximize $u^*(x, z) \equiv v(x; \rho) - \beta c(x; \gamma) + z$, subject to the budget constraint that we discuss below.[11]

We and other researchers often refer to $\beta < 1$ as representing a "self-control problem" because it reflects a short-term desire or propensity that the person disapproves of at every other moment in her life. Our welfare analysis therefore treats this preference for immediate gratification as an error, although the main points apply for essentially any welfare criterion. Specifically, we shall measure a person's welfare as a function of her choice by her long-run utility $u^{**}(x, z) \equiv v(x; \rho) - c(x; \gamma) + z$.[12]

The crucial feature that drives our results is that a person's behavior may not maximize her own welfare. This feature is quite common in the behavioral-economics literature, which often examines "errors" in utility maximization. Indeed, to highlight this feature, Kahneman (1994) makes an explicit distinction between a person's "decision utility", which is the utility function that explains a person's choices, and a person's "experienced utility", which is the utility function that reflects her welfare. In our model, $u^*(x, z)$ is the person's decision utility, whereas $u^{**}(x, z)$ is the person's experienced utility, and these differ when $\beta < 1$. Although our formal analysis focuses on this one specific source of decision utility deviating from experienced utility, we discuss in Section 6 the extent to which our conclusions hold in other behavioral models.

Consider ideal vs. actual behavior for an individual with per-period income I, where I is "large" relative to potato-chip consumption. The first-best optimal allocation for this individual, which we denote by (x^{**}, z^{**}), maximizes long-run utility $u^{**}(x, z)$ subject to the resource constraint $x + z \leq I$. Hence, x^{**} satisfies $v_x(x^{**}; \rho) - c_x(x^{**}; \gamma) - 1 = 0$ and $z^{**} = I - x^{**}$.

The person's actual behavior depends on taxes. Without taxes, the market price of potato chips will equal their marginal cost (which we have normalized to be 1), or $p_x = 1$. If the government imposes a per-unit tax t on potato chips, then the market price will be $p_x = 1 + t$. If in addition the person receives a lump-sum transfer ℓ from the government, her (per-period) budget constraint will be $(1+t)x + z \leq I + \ell$. The person will choose her consumption allocation to maximize $u^*(x, z)$ subject to this budget constraint. Hence, her consumption of potato chips, which we denote by $x^*(t)$, satisfies $v_x(x^*(t); \rho) - \beta c_x(x^*(t); \gamma) - (1+t) = 0$, and her consumption of the composite good is $z^*(t, \ell) = I + \ell - (1+t)x^*(t)$.[13]

From these conditions, it is straightforward to derive our first basic result: In the absence of taxes – when $t = 0$ – self-control problems lead to over-consumption of potato chips. In other words, for all ρ and γ, whereas actual potato-chip consumption $x^*(0)$ is identical to first-best

[11] The behavior of people with time-inconsistent preferences often depends on whether they are aware vs. unaware of their future self-control problems — on whether they are "sophisticated" vs. "naive". For our analysis in this paper, however, this distinction is irrelevant because there is no intertemporal link between decisions — that is, one's optimal behavior now is independent of her beliefs about her future behavior.

[12] In other words, we are using the long-run perspective for an agent's welfare function. While researchers sometimes worry about the appropriate welfare function for time-inconsistent agents, here there should be no controversy — u^{**} is appropriate under essentially any perspective. Perhaps most importantly, for any tax policy that takes effect *in the future*, under the β, δ model, the agent agrees that u^{**} is the appropriate welfare function. Alternatively, note that, if the person consumes a bundle (x', z') in all periods – as she does in our model – then her instantaneous utility will be exactly $u^{**}(x', z')$ in *all* periods except period 1. Hence, measuring welfare using $u^{**}(x', z')$ means we are also equating welfare to the person's per-period utility flow.

[13] Given our assumption that I is "large", potato-chip consumption is independent of ℓ. Also, note that, for notational simplicity, we suppress the arguments ρ, γ, and β in the expressions x^{**}, z^{**}, $x^*(t)$, and $z^*(t, \ell)$.

potato-chip consumption x^{**} for people with $\beta=1$, actual potato-chip consumption $x^*(0)$ is larger than first-best potato-chip consumption x^{**} for people with $\beta<1$. As we discuss in Section 1, this result can be interpreted in standard externality terms. Current consumption of potato chips imposes a negative externality on future selves. Whereas optimality requires giving full weight to the future cost, the current self only gives it weight $\beta<1$, and so ignores proportion $1-\beta$ of the future cost.

This negative externality intuition generates our second basic result: In a population of homogeneous consumers with self-control problems, a simple Pigouvian tax-and-transfer scheme can be used to correct this over-consumption. In particular, consider the case in which there are many identical consumers and the government imposes a per-unit tax t on potato chips and returns the proceeds to consumers via a uniform (per-capita) lump-sum ℓ. Since the lump-sum is independent of each consumer's own behavior, it is easy to see that a tax rate $t^{**}=(1-\beta)c_x(x^{**})$ will implement the first-best outcome — that is, will induce all consumers to choose $x^*(t^{**})=x^{**}$.[14]

While these basic results are straightforward, the heart of our analysis will focus on the case in which there is population heterogeneity in tastes, as captured by heterogeneity in ρ and γ, and – more importantly – also in the degree of self-control problems, as captured by heterogeneity in β. In this case, implementing the first-best outcome would require individual-specific taxes and lump-sum transfers. Such schemes are presumably unrealistic, because of informational constraints, implementation costs, arbitrage opportunities, and so forth. We will therefore assume that the government is limited to using linear taxes and lump-sum transfers that are the same for all consumers.

Formally, we assume that the population distribution of parameters is given by a cumulative distribution $F(\rho, \gamma, \beta)$. In addition, we assume that the distribution of tastes is independent from the distribution of self-control problems — that is, $F(\rho, \gamma, \beta)$ can be written as $G(\rho, \gamma)H(\beta)$. Given that an individual's demand for potato chips is $x^*(t)$, the aggregate demand for potato chips (in per-capita terms) is $X^*(t)=E_F[x^*(t)]$. Hence, a tax rate t will raise (per-capita) revenue $tX^*(t)$, and if the government returns all tax proceeds to consumers, the (per-capita) lump-sum transfer will be $\ell(t)=tX^*(t)$. The next two sections analyze optimal taxation given this heterogeneity.

3. Optimal taxes

In this section, we analyze optimal taxation given a specific social-welfare function, as in Diamond and Mirrlees (1971a,b) and the subsequent optimal-taxation literature. Specifically, we use a social-welfare function that puts "equal weight" on all people, although the basic ideas will clearly hold for other weights as well.

Recall that an individual's welfare function is $u^{**}(x, z)\equiv v(x;\rho)-c(x;\gamma)+z$. Given a tax t and a lump-sum $\ell(t)$, the person will choose consumption bundle $(x^*(t), z^*(t, \ell(t)))$, and so the person's welfare as a function of t will be $u^{**}(x^*(t), z^*(t, \ell(t)))$. To put equal weight on all people, the social-welfare function will be the expectation of individual welfare — that is,

$$\begin{aligned}\Omega(t)&\equiv E_F[u^{**}(x^*(t), z^*(t,\ell(t)))]\\&= E_F[v(x^*(t);\rho)-c(x^*(t);\gamma)+I+\ell(t)-(1+t)x^*(t)].\end{aligned}$$

[14] "Proof": Given tax $t=(1-\beta)c_x(x^{**};\gamma)$, $x^*(t)$ satisfies $v_x(x^*(t);\rho)-\beta c_x(x^*(t);\gamma)-(1+(1-\beta)c_x(x^{**};\gamma))=0$, which can be rewritten as $v_x(x^*(t);\rho)-c_x(x^{**};\gamma)-1+\beta[c_x(x^{**};\gamma))-c_x(x^*(t);\gamma))]=0$, which is satisfied if and only if $x^*(t)=x^{**}$.

(Note that we have substituted $z^*(t, \ell) = I + \ell(t) - (1+t)x^*(t)$.) Finally, because the tax payments and lump-sum transfers sum up to zero – because $\ell(t) = E_F[tx^*(t)]$ – we can simplify this equation to

$$\Omega(t) = E_F[v(x^*(t);\rho) - c(x^*(t);\gamma) - x^*(t) + I].$$

For each individual, $v(x^*(t);\ \rho) - c(x^*(t);\ \gamma) - x^*(t) \equiv \hat{u}(t)$ reflects the efficiency of potato-chip consumption. For any $(\rho,\ \gamma,\ \beta)$, $\hat{u}(t)$ is maximized when actual consumption $x^*(t)$ is equal to first-best consumption x^{**}, and otherwise there are consumption distortions. Hence, if policymakers put equal weight on all people, they will be concerned exclusively with minimizing the average distortion in potato-chip consumption. Proposition 1 characterizes optimal taxation given this social-welfare function. (All proofs are in the Appendix).

Proposition 1. *Suppose policymakers maximize $\Omega(t)$. For any distribution of tastes $G(\rho, \gamma)$:*

(1) If everyone in the population has $\beta = 1$, then the optimal tax $t^* = 0\%$; and
(2) If everyone in the population has $\beta \leq 1$ and some people have $\beta < 1$, then the optimal tax $t^* > 0\%$.

Part 1 establishes that if there are no self-control problems in the population then we should not tax potato chips. In the absence of taxes, people without self-control problems consume optimally. Hence, potato-chip taxes would merely distort consumption away from the first best. Part 2 establishes that, if instead some people have self-control problems, then it is always optimal to tax potato chips. Potato-chip taxes still create consumption distortions for people without self-control problems; however, because such people consume optimally when $t = 0\%$, these distortions are second-order. At the same time, because people with self-control problems over-consume when $t = 0\%$, potato-chip taxes create first-order reductions in consumption distortions for people with self-control problems.

Proposition 1 focuses on the case in which no one has a preference for *future* gratification — that is, no one has $\beta > 1$. More generally, because people with $\beta > 1$ under-consume when $t = 0\%$, the existence of such people would militate against taxing potato chips. (Indeed, if no one has $\beta < 1$ and some people have $\beta > 1$, the optimal tax would be $t^* < 0\%$ — that is, it would be optimal to subsidize potato-chip consumption). Even so, as long as the predominant error is a preference for immediate gratification rather than a preference for future gratification – as the evidence suggests – it will be optimal to tax sin goods.[15]

While Proposition 1 addresses whether it is optimal to tax potato chips, it does not address how the optimal tax depends on the prevalence of self-control problems in the population. A natural conjecture is that the more prevalent self-control problems are, the larger the optimal potato-chip tax. While this conjecture is correct for some specifications, it does not hold in general; Proposition 2 addresses when it is likely to hold.

[15] For instance, one can show that, if $v(x;\ \rho) = \rho \ln x$ and $c(x;\ \gamma) = \gamma \ln x$, it is optimal to tax potato chips whenever the average β in the population is less than 1.

Proposition 2. *Suppose policymakers maximize $\Omega(t)$. For a fixed distribution of tastes $G(\rho, \gamma)$, let t_0^* and t_1^* be the optimal taxes given distributions of self-control problems $H^0(\beta)$ and $H^1(\beta)$, respectively.*

(1) Suppose that $H^1(\beta) \geq H^0(\beta)$ for all β. If for all (ρ, γ) and $t \leq t_0^*$, $d\hat{u}/dt = (d\hat{u}/dx)(dx^*/dt)$ is larger for smaller β, then $t_1^* > t_0^*$; and

(2) Suppose that $H^1(\beta) \geq H^0(\beta)$ for all β and that $H^1(\beta|\beta \geq \beta_0) = H^0(\beta)$ for all $\beta \geq \beta_0$, where $\beta_0 \equiv \sup\{\beta|H^0(\beta) = 0\}$. If, for all (ρ, γ), $x^*(t_0^*) > x^{**}$ for $\beta = \beta_0$, then $t_1^* > t_0^*$.

The simple intuition behind the conjecture is that people with larger self-control problems have an increased propensity to over-consume and hence are more likely to be helped by taxes. This simple intuition is not entirely correct, however, because it does not account for people's sensitivity to tax changes. More precisely, the fact that people with larger self-control problems have an increased propensity to over-consume implies that $|d\hat{u}/dx|$ is larger for smaller β. Part 1 points out, however, that in order to conclude that any increase in the prevalence of self-control problems – in the sense of first-order stochastic dominance – implies an increase in the optimal tax, it must be that $d\hat{u}/dt$ is larger for smaller β, where $d\hat{u}/dt = (d\hat{u}/dx)(dx^*/dt)$. While this condition holds for some specifications – including our example in Section 5 – it does not hold in general.[16]

Part 2 establishes that the conjecture holds more broadly if we impose more structure on the way in which we increase the prevalence of self-control problems. Specifically, suppose we add to the population only people with larger self-control problems than currently exist — i.e., initially, everyone has $\beta \geq \beta_0$, and we add people with $\beta < \beta_0$. If these new people all over-consume under the initial optimal tax t_0^* (they have $x^*(t_0^*) > x^{**}$), then the new optimal tax t_1^* will be larger. Intuitively, amongst the initial population the marginal effect of a tax change is zero – because t_0^* is optimal – while at the same time increasing the tax above t_0^* helps every new member of the population.[17]

Thus far, our analysis has assumed that all revenue from potato-chip taxes is fully returned to consumers. An alternative – perhaps more realistic – assumption is that this revenue is used to reduce distortionary taxes elsewhere in the economy. If, for instance, the government raises revenue by taxing other commodities (in the spirit of Ramsey, 1927), such taxes create consumption distortions, and so taxing potato chips and using the proceeds to reduce taxes on other goods reduces consumption distortions.[18]

To illustrate, let $D(R)$ denote the reduction in distortions elsewhere when we raise (per-capita) revenue R from potato-chip taxes, in which case our social-welfare function becomes

$$\begin{aligned}\hat{\Omega}(t) &\equiv E_F[v(x^*(t);\rho) - c(x^*(t);\gamma) + I - (1+t)x^*(t)] + D(tX^*(t)) \\ &= \Omega(t) + [D(tX^*(t)) - tX^*(t)].\end{aligned}$$

[16] For our example in Section 5 – which assumes $v(x;\rho) = \rho x^{1-r}/(1-r)$ and $c(x;\gamma) = \gamma x$ – this condition holds as long as $t_0^* < (1+\gamma^{\min})r$, where $\gamma^{\min}$ is the minimum γ in the population.

[17] The condition that all new people over-consume given tax t_0^* will hold as long as the distribution of tastes is tight enough — indeed, it necessarily holds if everyone has the same (ρ, γ).

[18] In O'Donoghue and Rabin (2003), we in fact analyze a Ramsey framework. We show for a specific functional form that the existence of self-control problems implies that we should raise taxes on potato chips and reduce taxes on other goods relative to the Ramsey taxes that would be optimal if everyone were fully self-controlled.

Our analysis above assumes $D'(R)=1$, which reflects no distortions elsewhere in the sense that reducing (per-capita) tax collections elsewhere by \$1 is equivalent to giving everyone \$1. When there are distortions elsewhere, however, $D'(R)>1$, which says that reducing taxes elsewhere by \$1 is better than giving everyone \$1. In this case, potato-chip taxes have the added benefit of reducing distortions elsewhere — indeed, because of this second effect, it becomes optimal to tax potato chips even when everyone has $\beta=1$.

A natural extension of Proposition 1 is that, for whatever the optimal potato-chip tax might be under an assumption that everyone is fully self-controlled, if we recognize that there are self-control problems in the population, it becomes optimal to impose an even larger tax potato chips.[19] After all, such taxes still have the additional benefit of reducing over-consumption. It turns out, however, that this conclusion does not hold in general for much the same reason that our result in Proposition 2 does not hold in general — because in addition to affecting the degree of over-consumption, self-control problems also affect one's sensitivity to taxes. Even so, for most examples that we have worked out, this conclusion does hold.

4. Pareto-efficient taxes

In this section, we characterize Pareto-efficient taxes. As should be clear from our basic results in Section 2, if the government can use individual-specific taxes and lump-sum transfers, then it can implement first-best Pareto efficiency. Because such schemes are unrealistic, however, we investigate a constrained Pareto efficiency in which the government is restricted to uniform taxes and lump-sum transfers. In other words, because a potato-chip tax t implies that the (per-capita) lump-sum transfer will be $\ell(t)=tX^*(t)$, the choice set for policymakers is $\{(t, \ell(t)) \mid t\in[-1, \infty)\}$, and we analyze the set of constrained Pareto-efficient taxes within this set.[20]

To build intuition, consider how taxes affect an individual's long-run utility. For any t, the long-run utility for type (ρ, γ, β) is

$$\begin{aligned}\hat{u}^{**}(t) &\equiv v(x^*(t);\rho)-c(x^*(t);\gamma)+I+\ell(t)-(1+t)x^*(t)\\ &= [v(x^*(t);\rho)-c(x^*(t);\gamma)-x^*(t)]+[I+\ell(t)-tx^*(t)].\end{aligned}$$

The latter equation reveals that taxes have two effects on the individual's long-run utility. First, taxes affect the efficiency of potato-chip consumption, as reflected by $v(x^*(t);\rho)-c(x^*(t);\gamma)-x^*(t)$. Second, taxes and the associated lump-sum transfers redistribute income, as reflected by $I+\ell(t)-tx^*(t)$. Because $\ell(t)=tX^*(t)$, anyone whose consumption $x^*(t)$ is smaller than average consumption $X^*(t)$ on net receives income, while anyone whose consumption $x^*(t)$ is larger than average consumption $X^*(t)$ on net loses income. Whether an individual is better off under tax t vs. t' depends on the combination of these two effects.

The set of Pareto-efficient taxes depends on the support of preferences. Recall that $G(\rho, \gamma)$ and $H(\beta)$ represent, respectively, the population distributions of tastes and of self-control problems.

[19] While the simple framework used in the text highlights some basic ideas, it is inappropriate for a more formal analysis because it ignores the fact that $D(R)$ likely depends on the distribution of self-control problems in the population.

[20] Unless there are limits on free disposal, any tax $t<-1$ would lead people to demand infinite amounts merely to collect the net subsidy — hence why we bound taxes below at -1. This bound is not relevant for our analysis.

We let Γ and $\mathbf{B}$ denote the supports of G and H, respectively. Our analysis focuses on two types of Pareto comparisons:

Definition 1. Given population distributions $G(\rho, \gamma)$ and $H(\beta)$:

(1) A tax t is *Pareto-superior* to a tax t' if $\hat{u}^{**}(t) \geq \hat{u}^{**}(t')$ for all $(\rho, \gamma) \in \Gamma$ and $\beta \in \mathbf{B}$ and $\hat{u}^{**}(t) > \hat{u}^{**}(t')$ for some $(\rho, \gamma) \in \Gamma$ and $\beta \in \mathbf{B}$; and a tax t is *Pareto-efficient* if there does not exist any t' that is Pareto-superior to t.
(2) A tax t is *quasi-Pareto-superior* to a tax t' if $E_G[\hat{u}^{**}(t)] \geq E_G[\hat{u}^{**}(t')]$ for all $\beta \in \mathbf{B}$ and $E_G[\hat{u}^{**}(t)] > E_G[\hat{u}^{**}(t')]$ for some $\beta \in \mathbf{B}$; and a tax t is *quasi-Pareto-efficient* if there does not exist any t' that is quasi-Pareto-superior to t.

Hence, in addition to standard Pareto comparisons, we also consider a kind of group Pareto comparison, which we call a quasi-Pareto comparison, where we group people by their self-control problems. Specifically, we say that one tax is quasi-Pareto-superior to another if and only if every subpopulation with a fixed β is on average better off under that tax. This second criterion permits us to address a common worry that policies designed to combat "errors" involve a trade-off between helping people making errors while hurting people who are fully rational. This criterion allows us to assess whether, on average, fully rational people (people with $\beta = 1$) are hurt when we implement taxes.

Proposition 3 characterizes Pareto-efficient taxes when there are no self-control problems in the population.

Proposition 3. *Suppose that everyone has $\beta = 1$.*

(1) If there is no heterogeneity in (ρ, γ), then $t = 0\%$ is the unique Pareto-efficient tax; and
(2) If there is heterogeneity in (ρ, γ), then there exists $t' > 0\% > t''$ such that all taxes $t \in (t'', t')$ are Pareto-efficient.

Part 1 says that if there is no heterogeneity in tastes – as reflected by ρ and γ – then the unique Pareto-efficient policy is not to tax potato chips. With a homogeneous population, Pareto efficiency merely requires maximizing the sum of surplus, which, as we saw in Section 3, occurs for $t^* = 0\%$. Part 2, however, says that for the more interesting case of population heterogeneity in tastes, it is not Pareto-inefficient to tax potato chips as long as this tax is not too large. Although a potato-chip tax creates consumption distortions for everyone, since everyone is fully self-controlled, these consumption distortions are second-order near $t = 0\%$. At the same time, the potato-chip tax and the associated lump-sum transfer has first-order income-redistribution effects. In particular, although everyone whose potato-chip consumption is larger than average ($x^*(0) > X^*(0)$) is hurt, everyone whose potato-chip consumption is smaller than average ($x^*(0) < X^*(0)$) is helped. Hence, a small-enough potato-chip tax is a movement along the Pareto frontier. (Analogously, a small enough potato-chip subsidy is also a movement along the Pareto frontier).[21]

Proposition 4 characterizes Pareto-efficient taxes when some people have self-control problems, and in particular establishes that in this case a potato-chip tax can yield Pareto improvements relative to a zero tax — that is, $t = 0\%$ may be Pareto-inefficient:

[21] Trivially, when everyone has $\beta = 1$, quasi-Pareto efficiency becomes equivalent maximizing $\Omega(t)$, and so $t = 0\%$ is the unique quasi-Pareto-efficient policy even when there is taste heterogeneity.

Proposition 4. *Suppose that everyone has $\beta \leq 1$ and some people have $\beta < 1$. Suppose further that for all $(\rho, \gamma) \in \Gamma$ and $\beta \in \boldsymbol{B}$,*

$$V_{xxx} - \beta c_{xxx} \geq \frac{2c_{xx}}{c_x}(v_{xx} - \beta c_{xx}) \text{ for all } x. \quad \text{(A)}$$

(1) If there is no heterogeneity in (ρ, γ), then there exists $t' > 0\%$ such that all taxes $t \in (0\%, t')$ are Pareto-superior to $t = 0\%$; and

(2) If there is heterogeneity in (ρ, γ), then there exists $t' > 0\%$ such that all taxes $t \in (0\%, t')$ are quasi-Pareto-superior to $t = 0\%$; and if $\max_{(\rho,\gamma)\in\Gamma,\beta=1} x^*(0) < X^*(0)$, then there exists $t' > 0\%$ such that all taxes $t \in (0\%, t')$ are Pareto-superior to $t = 0\%$.

Part 1 considers the case in which there is no heterogeneity in tastes, and so the only heterogeneity is in the degree of self-control problems. In this case, potato-chip consumption is larger for people with larger self-control problems, and thus income is naturally redistributed from people with large self-control problems to people with small or no self-control problems. Because small taxes create only second-order consumption distortions for fully rational people, it immediately follows that small taxes make fully rational people better off (relative to $t = 0\%$). It does not immediately follow that small taxes make everyone better off, because for people with self-control problems it must be that the benefits from reduced consumption distortions are larger than the costs of a negative net income transfer. In particular, if their consumption responds very little to a tax increase, then the main effect of the tax increase will be the negative net income transfer, in which case they will be worse off. By guaranteeing that people with self-control problems have a sufficiently strong responsiveness to taxes, condition (*A*) is sufficient to conclude that small taxes create Pareto improvements.

Part 2 considers the case in which there is heterogeneity in both tastes and self-control problems. In this case, we cannot directly apply the logic above, because some people with self-control problems might consume *less* than some fully rational people – if they have weaker tastes for potato chips – and hence some fully rational people may have a negative net income transfer. Even so, the logic applies on average: on average fully rational people receive positive net income transfers, and therefore small taxes make fully rational people on average better off. Similarly, condition (*A*) guarantees that, for people with self-control problems, on average the benefits from reduced consumption distortions are larger than the costs of the negative net income transfer. Hence, condition (*A*) is sufficient to conclude that small taxes create quasi-Pareto improvements. Finally, whether small taxes can yield full Pareto improvements depends on whether the fully rational person that most likes potato chips is made better off or worse off, which in turn depends on whether that person's potato-chip consumption is smaller or larger than average.

The question remains how restrictive is condition (*A*). While it is a bit unclear how to give a general assessment, it holds in essentially all examples that we have considered. For instance, when the costs are linear or convex quadratic – when $c_{xx} \geq 0$ and $c_{xxx} = 0$ – then condition (*A*) holds as long as $v_{xxx} \geq 0$, which in turn holds for most commonly used functional forms. For instance, $v_{xxx} \geq 0$ for any benefit function that has non-increasing absolute risk aversion, including the CRRA and CARA functional forms, and also for a quadratic benefit function. In more intuitive terms, the requirement is that people with self-control problems are sensitive to tax changes. Whether they are is an empirical question. But if they respond very little or not at all to a tax increase, then the tax would merely redistribute income away from these types without any corresponding benefits.

Our analysis in this section demonstrates that paternalistic policies – by which we mean policies aimed at helping people overcome their errors – need not take the form of helping the irrational to the detriment of the rational. Rather, in some instances, paternalistic policies can make everyone better off, or at least can help people while making fully rational people on average better off. The intuition is straightforward: If a policy can make irrational people strictly better off, then there is scope to make fully rational people better off as well by reallocating resources from irrational people to rational people. This intuition is clearly quite general. What is somewhat special to taxes on sin goods is that the same policy that provides help to irrational people can lead naturally to the compensating reallocation of resources to rational people.

5. An example

Our formal results in Sections 3 and 4 are based on marginal arguments. To give a sense for the potential calibrational importance of these results, in this section we consider a specific example for which we can perform some back-of-the-envelope numerical calculations of the optimal taxes. These calculations reveal that, even if the prevalence of self-control problems in the population is relatively small, optimal taxes can still be significant.

Specifically, we assume that the future costs from consumption are linear in the amount consumed — that is, we assume $c(x;\gamma)=\gamma x$. The assumption of linearity will aid in interpreting our numerical calculations below. In particular, with linear costs, γ represents the magnitude of the future health cost relative to the cost of production. We also assume that the benefits from consumption take the CRRA functional form — that is, we assume $v(x;\rho)=\rho x^{1-r}/(1-r)$. We assume for simplicity that all consumers have the same r; however, we choose r to match a reasonable market elasticity of demand, as we discuss below.

With CRRA benefits and linear costs, an individual's demand becomes

$$x^*(t)=\frac{\rho^{1/r}}{(\beta\gamma+1+t)^{1/r}},$$

and the social-welfare function $\Omega(t)$ from Section 3 becomes:

$$\Omega(t)=E_F[\rho^{1/r}]E_F\left[\frac{1}{1-r}\left(\frac{1}{\beta\gamma+1+t}\right)^{\frac{1-r}{r}}-(\gamma+1)\left(\frac{1}{\beta\gamma+1+t}\right)^{\frac{1}{r}}\right]+I.$$

As this formula reveals, the combination of CRRA benefits and linear costs implies that the distribution of ρ is irrelevant for the optimal tax that maximizes $\Omega(t)$. In other words, the optimal tax will depend exclusively on the distributions of γ and β, and on r.[22]

To perform our numerical calculations, we assume specific distributions and then calculate the optimal tax numerically. Since our goal is to investigate how the optimal tax depends on the distribution of self-control problems in the population, we consider various distributions of β. We are particularly interested in whether "small" self-control problems can have a significant impact on optimal taxes. Hence, we consider values of β that are relatively close to one — specifically,

[22] This conclusion requires our assumption that ρ is not correlated with β or γ. For a derivation of $\Omega(t)$, see the Appendix.

we permit people to have $\beta \in \{1, 0.99, 0.95, 0.90\}$, and consider several distributions over these four values.

With linear health costs, recall that γ reflects the magnitude of health costs relative to the costs of production. For simplicity, we assume that the future health costs are the same for everyone. But we still must choose a reasonable value for γ. Unfortunately, we are unaware of any estimates in the literature for the future health costs from snack foods. As an admittedly somewhat arbitrary benchmark, we use the numbers from Gruber and Koszegi (2004) for cigarette consumption. Their back-of-the-envelope calculation for the cost in terms of life-years lost per pack of cigarettes is $35.64. This cost is on the order of ten times the production costs for cigarettes, suggesting $\gamma=10$. Although our model is not really applicable to addictive products such as cigarettes, we believe that Gruber and Koszegi's conclusion that the health costs could be an order of magnitude larger than production costs is equally plausible for snack foods such as potato chips. Hence, in our numerical calculations below, we start by assuming $\gamma=10$. But to highlight the importance of γ for the magnitude of the optimal tax, we also consider $\gamma=2$.

The remaining parameter is r, which reflects the curvature of the benefits function. We choose this parameter to yield a reasonable market elasticity of demand. For this target, we use two estimates in the literature for the elasticity for potato chips. Kuchler et al. (2005) use the AC Nielsen Homescan panel data – in which households scan their purchases at home – and estimate an elasticity for potato chips of −0.45. This elasticity is quite similar to the usual estimated elasticities for cigarettes and alcohol. Katchova et al. (2005) use aggregate price and quantity data, and estimate an elasticity of potato chips of −1.07. Hence, in our numerical calculations, we use a target elasticity of both $\varepsilon_D=-0.5$ and $\varepsilon_D=-1.0$.

Table 1 presents our numerical calculations. For each γ and distribution of β, we choose r to yield the target elasticity of demand when $t=0\%$ (the elasticity of demand depends on the tax). We

Table 1
Optimal taxes for different populations

Health cost and elasticity	Proportion of population with:				r	t^* (%)	$\underline{t}$ (%)
	$\beta=1$	$\beta=0.99$	$\beta=0.95$	$\beta=0.9$			
$\gamma=10$ and $\varepsilon_D=-0.5$	1/2	1/2	0	0	0.18	5.15	5.00
	1/2	0	1/2	0	0.19	28.53	24.81
	1/2	0	0	1/2	0.19	63.71	48.48
	1/4	1/4	1/4	1/4	0.19	49.26	39.26
	1/2	1/4	1/8	1/8	0.18	29.26	23.20
$\gamma=10$ and $\varepsilon_D=-1.0$	1/2	1/2	0	0	0.09	5.28	4.99
	1/2	0	1/2	0	0.09	31.68	24.34
	1/2	0	0	1/2	0.10	72.72	45.83
	1/4	1/4	1/4	1/4	0.10	56.41	38.01
	1/2	1/4	1/8	1/8	0.09	37.58	24.55
$\gamma=2$ and $\varepsilon_D=-0.5$	1/2	1/2	0	0	0.67	1.01	1.00
	1/2	0	1/2	0	0.68	5.21	5.00
	1/2	0	0	1/2	0.69	10.82	9.97
	1/4	1/4	1/4	1/4	0.69	8.52	7.98
	1/2	1/4	1/8	1/8	0.68	4.65	4.36
$\gamma=2$ and $\varepsilon_D=-1.0$	1/2	1/2	0	0	0.33	1.02	1.00
	1/2	0	1/2	0	0.33	5.34	4.99
	1/2	0	0	1/2	0.34	11.31	9.93
	1/4	1/4	1/4	1/4	0.34	8.84	7.96
	1/2	1/4	1/8	1/8	0.34	4.90	4.43

then fix that r, and solve numerically for the optimal tax t^*, which we report in Column 7 of Table 1. See the Appendix for more details about these calculations.[23]

Table 1 demonstrates that the existence of self-control problems can have dramatic implications for optimal taxation. For instance, the first five rows apply for $\gamma=10$ and $\varepsilon_D=-0.5$. If half the population is fully self-controlled while the other half the population has a very small present bias of $\beta=0.99$, then the optimal tax is 5.15%. If instead the half the population with self-control problems has a somewhat larger present bias of $\beta=0.90$ – which is still a smaller present bias (larger β) than often discussed in the literature – the optimal tax is 63.71%. Table 1 also reveals the role of the future health costs and the role of the elasticity of demand for optimal taxes. The magnitude of the optimal tax is quite sensitive to the magnitude of the health costs. When we compare $\gamma=10$ to $\gamma=2$, the optimal tax is roughly 5–7 times larger when $\gamma=10$. In contrast, the magnitude of the optimal tax is not much influenced by the elasticity of demand — for both $\gamma=10$ and $\gamma=2$, changing the elasticity of demand from $\varepsilon_D=-0.5$ to $\varepsilon_D=-1.0$ has a very small impact on the optimal tax. While it is an open empirical question exactly by how much the existence of self-control problems would alter optimal taxes, these numerical examples highlight that we should not presume the effect to be small.

We can also use this example to address the magnitude of Pareto-efficient taxes. With CRRA benefits and linear costs, condition A from Proposition 4 is satisfied. Hence, it follows from Proposition 4 that, if there is no heterogeneity in (ρ, γ), the minimum Pareto-efficient tax is larger than 0%; and even if there is heterogeneity in (ρ, γ), the minimum quasi-Pareto-efficient tax is larger than 0%. Our interest here is *how much* larger these taxes are. Because, for standard Pareto efficiency, the distribution of ρ plays a crucial role, and because we do not see a natural way to choose this distribution, we focus on quasi-Pareto efficiency. Recall from Definition 1 that, if we let G denote the distribution of (ρ, γ) and $\mathbf{B}$ denote the set of βs in the population, a tax t is quasi-Pareto-superior to a tax t' if $E_G[\hat{u}^{**}(t)] \geq E_G[\hat{u}^{**}(t')]$ for all $\beta \in \mathbf{B}$ and $E_G[\hat{u}^{**}(t)] > E_G[\hat{u}^{**}(t')]$ for some $\beta \in \mathbf{B}$. In our example here,

$$E_G[\hat{u}^{**}(t)] = E_G[\rho(x^*(t))^{1-r}/(1-r) - \gamma x^*(t) - (1+t)x^*(t) + \ell(t) + I].$$

Much as for the optimal tax that maximizes $\Omega(t)$, the combination of CRRA benefits and linear costs implies that the distribution of ρ is irrelevant for quasi-Pareto efficiency (for details, see the Appendix). Hence, for each case in Table 1, we can use this formula to solve numerically for the minimum quasi-Pareto-efficient tax $\underline{t}$, which we present in Column 8 of Table 1. For instance, consider again the case where $\gamma=10$ and $\varepsilon_D=-0.5$. When half the population has $\beta=1$ while the other half has $\beta=0.99$, any tax smaller than 5.00% is quasi-Pareto-inefficient. In other words, for any tax $t<5.00\%$, increasing the tax to 5.00% on average helps the people with $\beta=0.99$ *and* also on average helps the people with $\beta=1$. Similarly, when half the population has $\beta=1$ while the other half has $\beta=0.9$, any tax smaller than 48.48% is quasi-Pareto-inefficient — for any tax $t<48.48\%$, increasing the tax to 48.48% on average helps the people with $\beta=0.9$ *and* also on average helps the people with $\beta=1$.

Table 1 implies that, even if we mostly care about the average welfare of the fully self-controlled, the existence of small self-control problems in a portion of the population can have a significant impact on optimal taxes. In particular, if all we cared about is maximizing the average

[23] It is straightforward to confirm that, for the relevant range of taxes, the condition from Part 1 of Proposition 2 holds. Hence, throughout Table 1, an increase in the prevalence of self-control problems – in the sense of first-order stochastic dominance – leads to a larger optimal tax.

welfare of the fully self-controlled, then we would choose to implement the minimum quasi-Pareto-efficient tax $\underline{t}$. And much as for the optimal tax t^* that maximizes $\Omega(t)$, in our numerical examples, $\underline{t}$ can be significantly larger than 0%.[24]

6. Alternative models

Our main conclusions are driven by two crucial features of our model: (i) a person's behavior (x^*) does not maximize her welfare (u^{**}), and in particular the person consumes more potato chips than she herself would like; and (ii) the person's consumption of potato chips is sensitive to the market price, and hence a potato-chip tax can help to counteract her over-consumption. In our model, these features are driven by a time-inconsistent preference for immediate gratification. In this section, we discuss to what extent these features could arise from other behavioral models.

We first note that our model in this paper can literally be reinterpreted in terms of other sources of over-consumption. For instance, a person might under-appreciate the severity of future health costs, in which case the $\beta<1$ would reflect the extent of this under-appreciation. Alternatively, a person might have an irrational (incorrect) optimism that the negative health consequences will not occur for her, in which case the $\beta<1$ would reflect the extent of this optimism.

The two crucial features of our model can also arise in other behavioral models of intertemporal choice. Consider, for instance, the models of cue-triggered visceral factors (Loewenstein, 1996; Bernheim and Rangel, 2004) or dual motivations (Loewenstein and O'Donoghue, 2005). In such models, behavior depends on both "cognitive" motivations and "visceral" motivations — variously labeled cognitive vs. emotional, deliberative vs. affective, cold vs. hot, and so forth. For many sin goods, a natural assumption is that visceral motivations are primarily influenced by the consumption benefits. If so, and if we take cognitive motivations to reflect welfare, then such models generate exactly the type of over-consumption that has been our focus.[25] To illustrate (using our notation), it might be that cognitive motivations (and welfare) are reflected by our welfare function $u^{**}(x, z)=v(x;\rho)-c(x;\gamma)+z$, but, due to the visceral focus on consumption benefits, behavior is derived from decision utility $(1+\phi)v(x;\rho)-c(x;\gamma)+z$, where ϕ reflects the magnitude of the visceral motivations. Although this approach reflects an overweighting of immediate benefits as opposed to an underweighting of future costs, the policy conclusions would be much the same.

There is, however, an important sense in which a dual-motivations approach might yield different conclusions. Once again, not only do our conclusions require over-consumption, but they also require that consumption is sensitive to the market price. For over-consumption that is driven by visceral motivations, there may be reasons to believe that consumption is not very price-sensitive. Indeed, Bernheim and Rangel (2004, in press) argue that, for many addictive goods, consumption is driven by a kind of short-circuiting of rational decision-making, where visceral motivations (the "hot state") take over and are not price-sensitive at all. If so, then sin taxes may not be optimal, because they might merely make addicts pay a higher price without changing their

[24] There is also a maximum quasi-Pareto-efficient tax, which is determined by how taxes affect those with the smallest β in the population. But since our goal is to demonstrate that the existence of small self-control problems in a portion of the population can have a significant impact on optimal taxes even if we mostly care about the average welfare of the fully self-controlled, we do not investigate the maximum quasi-Pareto-efficient tax.

[25] Because they interpret visceral motivations (the "hot state") as a short-circuiting of rational decision-making, Bernheim and Rangel (2004) indeed argue that only cognitive motivations (the "cold state") are relevant for welfare. Loewenstein and O'Donoghue (2005) discuss reasons why visceral motivations might also be (to some extent) relevant.

consumption. Bernheim and Rangel in fact show that, because of such effects, in some circumstances it could conceivably be optimal to subsidize an addictive good.

Gul and Pesendorfer (2001) develop an alternative model of self-control problems that does not yield over-consumption, and hence yields different conclusions. In their model, self-control problems are driven by temptation disutility—specifically, when making consumption decisions, people experience disutility when they forgo the most tempting option currently available. Gul and Pesendorfer motivate this model as an alternative explanation for people making ex-ante commitments—from a prior perspective, commitments can be valuable if they alter the most tempting option that will be available when it is time to consume. Gul and Pesendorfer (in press) extend this model to addiction, and they conclude that a tax can only reduce welfare. To illustrate their logic (using our notation), a person's behavior might be derived from decision utility $u^{**}(x, z) - [v(\hat{x}; \rho) - v(x; \rho)]$, where $\hat{x}$ is the most tempting level of potato-chip consumption and the bracketed term reflects the disutility from forgoing this option. Gul and Pesendorfer further assume that the person's welfare function corresponds to this decision utility. Hence, under the plausible assumption that most tempting level of potato-chip consumption $\hat{x}$ is not determined by what one can afford but rather by some satiation point, a tax on potato chips merely distorts potato-chip consumption without any benefits.[26]

Although Gul and Pesendorfer reach a different conclusion from ours, this conclusion requires the assumption that temptation disutility merits full normative weight. If instead the temptation disutility were given less-than-full normative weight – e.g., if the person's welfare function were $u^{**}(x, z) - \alpha[v(\hat{x}; \rho) - v(x; \rho)]$ for some $\alpha \in [0, 1)$ – then our conclusions would once again hold. In particular, anyone with $\alpha < 1$ would over-consume (by her own reckoning), and as long as the person was price-sensitive, sin taxes could improve her welfare.

7. Discussion

In this section, we discuss the broader implications of our analysis, and also its limitations. This paper is part of a very recent literature that addresses public-policy implications of research in behavioral-economics. Because much of the behavioral-economics literature describes the ways in which people make errors that lead them not to behave in their own best interests, it suggests the possible desirability of designing paternalistic policies that help people make better choices. But opening this door raises a number of concerns.

Economists (and others) often equate "paternalism" with restrictions on choices. We do not. By "paternalism", we mean that we are concerned that people might not be behaving in their own best interests and we are designing policy with an eye towards how that policy might help people make better choices. The taxes that we discuss are no more a limit on choices than are any traditional taxes. Because the prescribed taxes change relative prices, they *change* choice sets relative to the no-tax case, but do not *reduce* choice sets. Moreover, the more sophisticated schemes we discuss below involve the *expansion* of choice sets — illustrating how in some instances the best way to help consumers make better choices is to make new options available.[27]

[26] Formally, the assumption about $\hat{x}$ means that $v(x; \rho)$ is maximized for some finite $\hat{x}$. And on the distortion, our tax-and-lump-sum-transfer scheme will merely reduce $u^{**}(x, z) + v(x; \rho)$ without changing $v(\hat{x}; \rho)$.

[27] This point is certainly implicit in the literature on self-control problems, where it is often discussed how the creation of commitment technologies can make people better off (see for instance Laibson (1997)).

A major worry with regard to paternalism is that most adults in most situations make better choices for themselves than the government or others would make for them. Most behavioral-economists, ourselves included, agree. As a result, there has been considerable emphasis in the literature on searching for minimally interventionist policies that help people who make errors while having little effect on those who are fully rational.[28]

While the focus on minimal interventions is a natural place to start, we believe economists should study "optimal paternalism" using the standard methods of economic theory: write down assumptions about the distribution of rational and irrational types of agents, about the available policy instruments, and about the government's information about agents, and then investigate which policies achieve the "best" outcomes. In other words, economists ought to treat the analysis of optimal paternalism as a mechanism-design problem when some agents might be boundedly rational. Our analysis in this paper illustrates the value of this approach. While heavy taxes may appear to be more heavy-handed and invasive than other cautiously paternalistic policies that we and others have advocated, our analysis reveals that in fact even relatively large taxes are unlikely to cause much harm to 100% self-controlled agents. Hence, even when we believe only a small proportion of the population makes errors, optimal policy might involve seemingly large deviations from the policy that would be optimal if everyone were fully rational. Furthermore, our analysis of Pareto-efficient taxes reveals that imposing taxes may not even involve trading off benefits for people who make errors against costs for fully rational people. In some instances, everyone can benefit.

To what extent should the government get involved in providing commitment devices to counteract self-control problems — after all, why couldn't the private market provide any needed commitment devices? There are reasons to believe that, in fact, the government may play a very special role. One reason to be cautious in presuming that the private market will solve self-control problems is that people may be unaware of their own need for commitment; it may be hard to sell people a service they do not think they need. It may also simply be impossible for the private market to provide the needed commitment devices. The same consumer who wants a commitment device to apply for some future decision may also want to get around that commitment device when that future decision arrives. If it is profitable for firms to provide ways to get around earlier commitments, then the earlier commitments will never be taken in the first place. Imagine if our example of taxes were left to the private market. In principle, a person might sign a contract with a firm that says the firm will charge her a price above cost for potato chips. When she is craving potato chips, however, nothing stops another firm from offering her potato chips at cost. The special role of the government is that a government-imposed per-unit tax requires *all firms* to charge the higher price.[29]

Our analysis has numerous limitations. For instance, we have ignored the possibility of substitute goods. If taxes are imposed on potato chips, people might substitute out of potato chips and into Twinkies. If such goods are taxable–and carefully taxed at the appropriate rate–our

[28] See for instance O'Donoghue and Rabin (1999b, 2001) who discuss "cautious paternalism"; Camerer et al. (2003), who explore "asymmetric paternalism"; Sunstein and Thaler (2003a,b), who investigate "libertarian paternalism"; and Choi et al. (2003), who discuss "benign paternalism".

[29] More generally, as an alternative to loose intuitions for how markets might deal with self-control problems, explicit models can allow economists to study carefully how market reactions compare to government intervention. Indeed, some researchers have begun this process — for instance, DellaVigna and Malmendier (2004) and Koszegi (in press) explore more systematically the types of situations in which the market is likely to be able or unable to provide commitment devices.

analysis extends in a straightforward way.[30] But if such goods are not taxable–for instance, if we start increasing taxes on alcohol, people might substitute into marijuana – a problem arises. If policy-makers fully recognize their existence, then substitute but non-taxable sins merely put limits on the effectiveness of policy. In our framework, for instance, if $u^*=v(x+w;\ \rho)-c(x+w;\ \gamma)+z$, where w is a non-taxable sin good with market price p_w, then a constraint on policy would be that we must have $1+t_x \leq p_w$, because otherwise people would buy w rather than x. If instead policymakers naively ignore the existence of substitute and non-taxable sins, then imposing taxes may inadvertently do more harm than good. This is especially a concern if substitute sins have larger health costs—for instance, cigarette taxes might lead people to substitute into black-market, unfiltered cigarettes.

We have also limited attention to uniform linear taxes. Especially because such taxes generally cannot implement the first-best, it is natural to consider whether more sophisticated schemes can do better. In particular, we might take advantage of the fact that people with self-control problems would like to behave themselves in the future. For instance, a policy might attempt to sort types via tax menus wherein each consumer chooses in advance her per-unit tax and the associated lump-sum transfer. For some initial thoughts on such mechanisms, see O'Donoghue and Rabin (2003, in press).

A closely related issue is whether there exist superior policy instruments besides taxes. Given that the problem is over-consumption, perhaps a superior policy instrument would be to impose quantity restrictions—that is, a maximum quantity that people are permitted to consume. In some instances, quantity restrictions could be effective—e.g., if we knew everyone's ideal consumption, we could just set the maximum quantity equal to it. But more generally, price commitments (taxes) have a major advantage relative to quantity commitments. Specifically, if people experience day-to-day variation in their tastes, there is value to having some flexibility to react to this day-to-day variation. When there is a commitment to a higher price (as with a tax), the person can still buy more when her tastes are high and buy less when her tastes are low. Under a quantity restriction, only the latter flexibility is possible.

Despite these limitations, we hope that the insights from our analysis in this simplified environment can be an early step to a more general analysis of optimal taxation when not all consumers are 100% self-controlled.

Acknowledgments

We thank Robert Hall and Robert Barro for helpful discussions of a related shorter paper presented at the AEA meetings in January 2003. For helpful comments, we thank Jonathan Gruber and an anonymous referee, and also Steve Coate, Botond Koszegi, Emmanuel Saez, and seminar participants at Harvard, Yale, Stanford, Cornell, Vanderbilt, North Carolina State, the USC Behavioral Public Finance Conference, the 2003 Association of Public Economic Theory Conference at Duke, the 2004 ASSA meetings, and the Cornell-LSE-MIT Conference on Behavioral Public and Development Economics. For research assistance, we thank Christoph Vanberg and Chris Cotton. For financial support, we thank the National Science Foundation (grants SES-0214043 and SES-0214147), and Rabin thanks the Russell Sage and MacArthur Foundations.

[30] There is of course a substantial practical issue of whether real-world governments can accurately assess which goods should be taxed and which should not.

Appendix A. Proofs and derivations

Preliminary results: we first derive some basic results that will be useful throughout the proofs. Recall that our propositions assume $\beta \leq 1$, and we assume throughout that $v_{xx} - c_{xx} < 0$ (see footnote 8). From the text, $\Omega(t) = E_F[\hat{u}(t)] + I$, where

$$\hat{u}(t) \equiv v(x^*(t); \rho) - c(x^*(t); \gamma) - x(t).$$

As long as v and c are thrice differentiable, Ω is continuous and twice differentiable, where

$$\frac{d\Omega}{dt} = E_F\left[\frac{d\hat{u}}{dt}\right] = E_F\left[\frac{d\hat{u}}{dx}\frac{dx^*}{dt}\right].$$

In addition, for each (ρ, γ, β), $x^*(t)$ satisfies $v_x(x^*(t); \rho) - \beta c_x(x^*(t); \gamma) - (1+t) = 0$, from which one can derive:

$$\frac{dx^*}{dt} = \frac{-1}{-[v_{xx}(x^*(t); \rho) - \beta c_{xx}(x^*(t); \gamma)]} < 0$$

$$\frac{dx^*}{d\beta} = \frac{-c_x(x^*(t); \gamma)}{-[v_{xx}(x^*(t); \rho) - \beta c_{xx}(x^*(t); \gamma)]} < 0$$

$$\frac{dx^*}{d\rho} = \frac{v_{x\rho}(x^*(t); \gamma)}{-[v_{xx}(x^*(t); \rho) - \beta c_{xx}(x^*(t); \gamma)]} > 0$$

$$\frac{dx^*}{d\gamma} = \frac{-\beta c_{x\gamma}(x^*(t); \gamma)}{-[v_{xx}(x^*(t); \rho) - \beta c_{xx}(x^*(t); \gamma)]} < 0$$

$$\frac{d\hat{u}}{dx} = v_x(x^*(t); \rho) - c_x(x^*(t); \gamma) - 1 = t - (1-\beta)c_x(x^*(t); \gamma)$$

Proof of Proposition 1. (1) If everyone has $\beta = 1$, then everyone has $d\hat{u}/dx = t$, and thus

$$\frac{d\Omega}{dt} = E_F\left(t\frac{dx^*}{dt}\right) = tE_F\left(\frac{dx^*}{dt}\right).$$

Because $E_F(dx^*/dt) < 0$ for all t, Ω is quasiconcave in t. In particular, $d\Omega/dt > 0$ for $t < 0\%$, $d\Omega/dt = 0$ for $t = 0\%$, and $d\Omega/dt < 0$ for $t > 0\%$, and so the optimal tax $t^* = 0\%$.

(2) Suppose instead that everyone has $\beta \leq 1$ and some people have $\beta < 1$. As in part 1, everyone with $\beta = 1$ has $d\hat{u}/dt = t(dx^*/dt) \geq 0$ for all $t \leq 0\%$. Everyone with $\beta < 1$ has $d\hat{u}/dt = [t - (1-\beta)c_x(x^*(t); \gamma)](dx^*/dt)$. By assumption, $c_x(x^*(t); \gamma) > 0$, which implies $[t - (1-\beta)c_x(x^*(t); \gamma)] < 0$ for any $t \leq 0\%$, which in turn implies $d\hat{u}/dt > 0$ for any $t \leq 0\%$. Because $d\Omega/dt = E_F[d\hat{u}/dt]$, it follows that $d\Omega/dt > 0$ for all $t \leq 0\%$, and thus the optimal tax $t^* > 0\%$.

(Note: When some people have $\beta<1$, Ω is not necessarily quasiconcave in t. But our proof establishes that, while there may be multiple local maxima, all local maxima have $t>0\%$, and therefore any global maximum has $t>0\%$.) □

Proof of Proposition 2. To clarify notation, define $F^0(\rho, \gamma, \beta)\equiv G(\rho, \gamma)H^0(\beta)$ and $F^1(\rho, \gamma, \beta)\equiv G(\rho, \gamma)H^1(\beta)$, and then $\Omega^0(t)=E_{F^0}[\hat{u}(t)]+I$ and $\Omega^1(t)=E_{F^1}[\hat{u}(t)]+I$.

(1) For any (ρ, γ) and t, if $d\hat{u}/dt$ is larger for smaller β, then $H^1(\beta)\geq H^0(\beta)$ for all β implies

$$E_{H^0}[d\hat{u}/dt] < E_{H^1}[d\hat{u}/dt].$$

For any t, if, for all (ρ, γ), $d\hat{u}/dt$ is larger for smaller β, then

$$\begin{array}{lll} E_G\ (E_{H^0}[d\hat{u}/dt]) & < E_G(E_{H^1}[d\hat{u}/dt]) & \Leftrightarrow \\ E_{F^0}[d\hat{u}/dt] & < E_{F^1}[d\hat{u}/dt] & \Leftrightarrow \\ d\Omega^0/dt & < d\Omega^1/dt. & \end{array}$$

Hence, if for all (ρ, γ) and $t\leq t_0^*$, $d\hat{u}/dt$ is larger for smaller β, then $d\Omega^1/dt>d\Omega^0/dt$ for all $t\leq t_0^*$. It follows that $t_1^*>t_0^*$.

(2) Because $H^1(\beta|\beta\geq\beta_0)=H^0(\beta)$ for all $\beta\geq\beta_0$,

$$\Omega^0(t) = E_{H^0}[E_G(\hat{u}(t))] + I = E_{H^1|\beta\geq\beta_0}[E_G(\hat{u}(t))] + I.$$

Hence,

$$\begin{aligned} \Omega^1(t) &= \Pr_{H^1}(\beta\geq\beta_0)E_{H^1|\beta\geq\beta_0}[E_G(\hat{u}(t))] + \Pr_{H^1}(\beta<\beta_0)E_{H^1|\beta<\beta_0}[E_G(\hat{u}(t))] + I \\ &= \Pr_{H^1}(\beta\geq\beta_0)\Omega^0(t) + \Pr_{H^1}(\beta<\beta_0)E_{H^1|\beta<\beta^0}[E_G(\hat{u}(t))] + (1-\Pr_{H^1}(\beta\geq\beta_0))I. \end{aligned}$$

We first prove that $\Omega^1(t_0^*)\geq\Omega^1(t)$ for any $t<t_0^*$. Because t_0^* is optimal given Ω^0, $\Omega^0(t_0^*)\geq\Omega^0(t)$ for any $t<t_0^*$. For any (ρ, γ), because $dx^*/d\beta<0$ and $dx^{**}/d\beta=0$, if $x^*(t_0^*)>x^{**}$ for $\beta=\beta_0$ then $x^*(t_0^*)>x^{**}$ for all $\beta<\beta_0$. Moreover, because, for any t, $x^*(t)>x^{**}$ implies $v_x(x^*(t);\ \rho)-c_x(x^*(t);\ \gamma)-1<0$, and because $dx^*/dt<0$, $x^*(t_0^*)>x^{**}$ implies $d\hat{u}/dt=[v_x(x^*(t);\ \rho)-c_x(x^*(t);\ \gamma)-1]\ \frac{dx^*}{dt}>0$ for all $t\leq t_0^*$. It follows that for all (ρ, γ) and $\beta<\beta_0$, $\hat{u}(t_0^*)>\hat{u}(t)$ for all $t<t_0^*$. And thus $\Omega^1(t_0^*)\geq\Omega^1(t)$ for any $t<t_0^*$.

We next prove that $d\Omega^1/dt|_{t=t_0^*}>0$.

$$\frac{d\Omega^1}{dt} = \Pr_{H^1}(\beta\geq\beta_0)\frac{d\Omega^0}{dt} + \Pr_{H^1}(\beta<\beta_0)E_{H^1|\beta<\beta_0}\left[E_G\left(\frac{d\hat{u}}{dt}\right)\right].$$

Because t_0^* is optimal given Ω^0, $d\Omega^0/dt|_{t=t_0^*}=0$. And from above, for all (ρ, γ) and $\beta<\beta_0$, $x^*(t_0^*)>x^{**}$ implies $d\hat{u}/dt|_{t=t_0^*}>0$. It follows that $d\Omega^1/dt|_{t=t_0^*}>0$.

Finally, the combination of $\Omega^1(t_0^*)\geq\Omega^1(t)$ for any $t<t_0^*$ and $d\Omega^1/dt|_{t=t_0^*}>0$ implies that $t_1^*>t_0^*$. □

Proof of Proposition 3. From the text,

$$\hat{u}^{**}(t) = [v(x^*(t);\rho)-c(x^*(t);\gamma)-x^*(t)] + [I+\ell(t)-tx^*(t)],$$

where $\ell(t)\equiv tX^*(t)=tE_F[x^*(t)]$.

(1) If there is no heterogeneity in (ρ, γ) or in β, then $x^*(t)$ is the same for everyone, and so $\ell(t)-tx^*(t)=0$ for everyone. Hence, everyone has the same

$$\hat{u}^{**}(t) = v(x^*(t);\rho)-c(x^*(t);\gamma)-x^*(t)+I,$$

and, given $\beta=1$, this $\hat{u}^{**}(t)$ is maximized at $t=0\%$. It follows that $t=0\%$ is the unique Pareto-efficient tax.

(2) Let $\rho^{\max}$ and $\rho^{\min}$ be the maximum and minimum ρ in the population, and let $\gamma^{\max}$ and $\gamma^{\min}$ be the maximum and minimum γ in the population. We refer to a person who has $\rho=\rho^{\max}$ and $\gamma=\gamma^{\min}$ as a type (A) person, and we refer to a person who has $\rho=\rho^{\min}$ and $\gamma=\gamma^{\max}$ as a type (B) person. Because $dx^*/d\rho>0$ and $dx^*/d\gamma<0$, for any t, type (A) has the largest $x^*(t)$ in the population, and therefore has $x^*(t)>X^*(t)$. Analogously, type (B) has the smallest $x^*(t)$ in the population, and therefore has $x^*(t)<X^*(t)$.

We can rewrite

$$\hat{u}^{**}(t) = [v(x^*(t);\rho)-c(x^*(t);\gamma)-x^*(t)]+[I+t(X^*(t)-x^*(t))].$$

Because, given $\beta=1$, $v(x^*(t);\rho)-c(x^*(t);\gamma)-x^*(t)$ is maximized at $t=0$, type (A) has $\hat{u}^{**}(0)>\hat{u}^{**}(t)$ for any $t>0$, while type (B) has $\hat{u}^{**}(0)>\hat{u}^{**}(t)$ for any $t<0$. Moreover, for any (ρ, γ), $\beta=1$ implies $v_x(x^*(0);\rho)-c_x(x^*(0);\gamma)-1=0$, and thus

$$\left.\frac{d\hat{u}^{**}}{dt}\right|_{t=0\%} = X^*(0)-x^*(0).$$

It follows that there exists a $t''<0\%$ such that type (A) has $d\hat{u}^{**}/dt<0$ for all $t\in(t'', 0\%]$, and there exists a $t'>0\%$ such that type (B) has $d\hat{u}^{**}/dt>0$ for all $t\in[0\%, t')$.

Consider some $t_0\in[0\%, t')$. Any $t<t_0$ cannot be Pareto-superior to t_0 because type (B) has $\hat{u}^{**}(t)<\hat{u}^{**}(t_0)$. For any $t>t_0$, all types have

$$v(x^*(t);\rho)-c(x^*(t);\gamma)-x^*(t) < v(x^*(t_0);\rho)-c(x^*(t_0);\gamma)-x^*(t_0).$$

At the same time, because $X^*(t)=E_G[x^*(t)]$, it is impossible for all types to have $X^*(t)-x^*(t)>X^*(t_0)-x^*(t_0)$. Hence, there must be some type that has $\hat{u}^{**}(t)<\hat{u}^{**}(t_0)$, and so t cannot be Pareto-superior to t_0. It follows that any $t_0\in[0\%, t')$ is Pareto-efficient.

An analogous argument (using type (A)) establishes that any $t_0\in(t'', 0\%]$ is Pareto-efficient. □

Proof of Proposition 4. We first prove that condition (A) implies that $d\hat{u}^{**}/dt|_{t=0\%}$ is larger for smaller β.

$$\left.\frac{d\hat{u}^{**}}{dt}\right|_{t=0\%} = [v_x(x^*(0);\rho)-c_x(x^*(0);\gamma)-1]\frac{dx^*}{dt}+X^*(0)-x^*(0).$$

To ease notation in what follows, we suppress the arguments in the derivatives of v and c.

$$\frac{d\left[\left.\frac{d\hat{u}^{**}}{dt}\right|_{t=0\%}\right]}{d\beta} = [v_{xx}-c_{xx}]\frac{dx^*}{dt}\frac{dx^*}{d\beta}+[v_x-c_x-1]\frac{d\left[\frac{dx^*}{dt}\right]}{d\beta}-\frac{dx^*}{d\beta}.$$

From our preliminary results, $dx^*/dt = 1/(v_{xx} - \beta c_{xx})$, $dx^*/d\beta = c_x/(v_{xx} - \beta c_{xx})$, and $v_x - c_x - 1 = t - (1-\beta)c_x = -(1-\beta)c_x$ at $t = 0\%$. Differentiating dx^*/dt,

$$\frac{d\left[\frac{dx^*}{dt}\right]}{d\beta} = \frac{-\left[(v_{xxx} - \beta c_{xxx})\frac{dx^*}{d\beta} - c_{xx}\right]}{(v_{xx} - \beta c_{xx})^2} = \frac{dx^*}{dt}\frac{dx^*}{d\beta}\left[\frac{c_{xx}}{c_x} - \frac{(v_{xxx} - \beta c_{xxx})}{(v_{xx} - \beta c_{xx})}\right].$$

Hence,

$$\frac{d\left[\left.\frac{d\hat{u}^{**}}{dt}\right|_{t=0\%}\right]}{d\beta} = \frac{dx^*}{dt}\frac{dx^*}{d\beta}\left[(v_{xx} - c_{xx}) - \left((1-\beta)c_x\left[\frac{c_{xx}}{c_x} - \frac{(v_{xxx} - \beta c_{xxx})}{(v_{xx} - \beta c_{xx})}\right]\right) - (v_{xx} - \beta c_{xx})\right]$$
$$= \frac{dx^*}{dt}\frac{dx^*}{d\beta}(1-\beta)\left[\frac{c_x(v_{xxx} - \beta c_{xxx})}{(v_{xx} - \beta c_{xx})} - 2c_{xx}\right].$$

Because $dx^*/d\beta < 0$ and $dx^*/dt < 0$, it follows that $d\hat{u}^{**}/dt|_{t=0\%}$ is larger for smaller β if

$$v_{xxx} - \beta c_{xxx} \geq \frac{2c_{xx}}{c_x}(v_{xx} - \beta c_{xx}) \text{ for all } x.$$

(1) If there is no heterogeneity in (ρ, γ), then $dx^*/d\beta < 0$ implies that $x^*(0)$ is smallest for people with $\beta = 1$, and so people with $\beta = 1$ have $x^*(0) < X^*(0)$. Since people with $\beta = 1$ also have $v_x(x^*(0);\ \rho) - c_x(x^*(0);\ \gamma) - 1 = 0$, they have $d\hat{u}^{**}/dt|_{t=0\%} > 0$. Moreover, since $d\hat{u}^{**}/dt|_{t=0\%}$ is larger for smaller β, everyone with $\beta < 1$ also has $d\hat{u}^{**}/dt|_{t=0\%} > 0$. Hence, everyone in the population has $d\hat{u}^{**}/dt|_{t=0\%} > 0$, and so there exists $t' > 0\%$ such that all taxes $t \in (0\%, t')$ are Pareto-superior to $t = 0\%$.

(2) We first prove the latter statement. Much as in the proof for part 1, if $\max_{(\rho,\ \gamma)\in\Gamma,\ \beta=1} x^*(0) < X^*(0)$, then for all (ρ, γ), people with $\beta = 1$ have $d\hat{u}^{**}/dt|_{t=0\%} > 0$. And since $d\hat{u}^{**}/dt|_{t=0\%}$ is larger for smaller β, for all (ρ, γ), everyone with $\beta < 1$ also has $d\hat{u}^{**}/dt|_{t=0\%} > 0$. Again, since everyone in the population has $d\hat{u}^{**}/dt|_{t=0\%} > 0$, there exists $t' > 0\%$ such that all taxes $t \in (0\%, t')$ are Pareto-superior to $t = 0\%$.

To prove the former statement, we prove that, for all β, $d[E_G[\hat{u}^{**}(t)]]/dt|_{t=0\%} > 0$. Note that

$$\left.\frac{d\ [E_G[\hat{u}^{**}(t)]]}{dt}\right|_{t=0\%} = E_G\left[\left.\frac{d\hat{u}^{**}}{dt}\right|_{t=0\%}\right].$$

People with $\beta = 1$ have $d\hat{u}^{**}/dt|_{t=0\%} = X^*(0) - x^*(0)$, and so $E_G[d\hat{u}^{**}/dt|_{t=0\%}] = X^*(0) - E_G[x^*(0)]$. Because $dx^*/d\beta < 0$, $d[E_G[x^*(0)]]/d\beta = E_G[d[x^*(0)]/d\beta] < 0$. Hence, $E_G[x^*(0)]$ is smallest for $\beta = 1$, and so $X^*(0) > E_G[x^*(0)]$. It follows that $d[E_G[\hat{u}^{**}(t)]]/dt|_{t=0\%} > 0$ for $\beta = 1$.

Consider any $\beta < 1$. Because, for all (ρ, γ), $d\hat{u}^{**}/dt|_{t=0\%}$ is larger for smaller β, it follows that $d[E_G[\hat{u}^{**}(t)]]/dt|_{t=0\%}$ is larger for smaller β. Given $d[E_G[\hat{u}^{**}(t)]]/dt|_{t=0\%} > 0$ for $\beta = 1$, we have $d[E_G[\hat{u}^{**}(t)]]/dt|_{t=0\%} > 0$ for all β. The result follows. □

Details for numerical calculations in Section 5:

First, we derive analytically an equation for $d\Omega(t)/dt$. It is straightforward to derive that, with CRRA benefits and linear costs, an individual's demand becomes

$$x^*(t) = \frac{\rho^{1/r}}{(\beta\gamma + 1 + t)^{1/r}}.$$

Hence, the social-welfare function $\Omega(t)$ from Section 3 becomes:

$$\begin{aligned}\Omega(t) &= E_F[\rho(x^*(t))^{1-r}/(1-r)-\gamma x^*(t)-x^*(t)+I] \\ &= E_F\left[\frac{\rho}{1-r}\left(\frac{\rho^{1/r}}{(\beta\gamma+1+t)^{1/r}}\right)^{1-r}-(\gamma+1)\left(\frac{p^{1/r}}{(\beta\gamma+1+t)^{1/r}}\right)\right]+I \\ &= E_F[\rho^{1/r}]E_F\left[\frac{1}{1-r}\left(\frac{1}{\beta\gamma+1+t}\right)^{\frac{1-r}{r}}-(\gamma+1)\left(\frac{1}{\beta\gamma+1+t}\right)^{\frac{1}{r}}\right]+I.\end{aligned}$$

Note that maximizing $\Omega(t)$ is equivalent to maximizing $\hat{\Omega}(t)\equiv\Omega(t)/E_F[\rho^{1/r}]$, and:

$$\begin{aligned}\frac{d\hat{\Omega}(t)}{dt} &= E_F\left[\frac{-1}{r}\left(\frac{1}{\beta\gamma+1+t}\right)^{1/r}+\frac{(\gamma+1)}{r}\left(\frac{1}{\beta\gamma+1+t}\right)^{1/r+1}\right] \\ &= E_F\left[\frac{(1-\beta)\gamma-t}{r(\beta\gamma+1+t)^{1/r+1}}\right].\end{aligned}$$

Next, we derive analytically an equation for the elasticity of demand. Let $q\equiv 1+t$ denote the market price, and so $x^*(t)=\rho^{1/r}/(\beta\gamma+q)^{1/r}$. Market demand is thus

$$X^* = E_F[x^*(t)] = E_F[\rho^{1/r}]E_F[(\beta\gamma+q)^{-1/r}],$$

and therefore the elasticity of demand is

$$\begin{aligned}\varepsilon_D &= \frac{dX^*}{dq}\frac{q}{X^*} = E_F[\rho^{1/r}]E_F[(-1/r)(\beta\gamma+q)^{-1/r-1}]*\frac{q}{E_F[\rho^{1/r}]E_F[(\beta\gamma+q)^{-1/r}]} \\ &= \left(\frac{-q}{r}\right)\frac{E_F[(\beta\gamma+q)^{-1/r-1}]}{E_F[(\beta\gamma+q)^{-1/r}]}.\end{aligned}$$

We can now conduct our numerical calculations for the optimal tax. Specifically, we fix a γ (that is the same for everyone) and a distribution of β. We then set $t=0\%$, or $q=1$, and choose r so that ε_D is equal to its target value. Finally, we fix that r, and find numerically the t^* such that $d\hat{\Omega}(t)/dt>0$ for all $t<t^*$ and $d\hat{\Omega}(t)/dt<0$ for $t>t^*$.[31]

Finally, we search for the minimum quasi-Pareto-efficient tax. To do so, note that

$$\begin{aligned}E_G[\hat{u}^{**}(t)] &= E_G[\rho(x^*(t))^{1-r}/(1-r)-\gamma x^*(t)-(1+t)x^*(t)+\ell(t)+I] \\ &= E_G[\rho^{1/r}]\left[\frac{1}{1-r}\left(\frac{1}{\beta\gamma+1+t}\right)^{(1-r)/r}-(\gamma+1+t)\left(\frac{1}{\beta\gamma+1+t}\right)^{1/r}\right]+\ell(t)+I.\end{aligned}$$

Because $\ell(t)=t\ E_F[x^*(t)]=t\ E_G[\rho^{1/r}]\ E_F[(\beta\gamma+1+t)^{-1/r}]$, and because γ is the same for everyone, it follows that $E_G[\hat{u}^{**}(t)]\geq E_G[\hat{u}^{**}(t')]$ if and only if $\hat{u}^{***}(t)\geq\hat{u}^{***}(t')$, where

[31] One can show that $\hat{\Omega}(t)$ is concave for all $t<(1+\gamma)r$. For all cases in Table 1, t^* is well within this range. Moreover, for the cases in Table 1, one can also show that $d\hat{\Omega}(t)/dt<0$ for all $t>(1+\gamma)r$. It follows that this technique does indeed identify the global optimum.

$\hat{u}^{***}(t) \equiv t\frac{1}{1-r}(\beta\gamma+1+t)^{-1/r+1} - (\gamma+1+t)(\beta\gamma+1+t)^{-1/r} + tE_F[(\beta\gamma+1+t)^{-1/r}]$. Hence, we merely compute numerically $\hat{u}^{***}(t)$ for $\beta=1$, and search for the minimum t such that $\hat{u}^{***}(t) > \hat{u}^{***}(t+0.0001)$.[32]

References

Ainslie, George, 1991. Derivation of 'Rational' Economic Behavior from Hyperbolic Discount Curves. American Economic Review 81 (2), 334–340.

Ainslie, George, 1992. Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person. Cambridge University Press, New York.

Ainslie, George, Haslam, Nick, 1992a. Self-control. In: Loewenstein, George, Elster, Jon (Eds.), Choice Over Time. Russell Sage Foundation, New York, pp. 177–209.

Ainslie, George, Haslam, Nick, 1992b. Hyperbolic discounting. In: Loewenstein, George, Elster, Jon (Eds.), Choice Over Time. Russell Sage Foundation, New York, pp. 57–92.

Angeletos, George-Marios, Laibson, David, Repetto, Andrea, Tobacman, Jeremy, Weinberg, Stephen, 2001. The hyperbolic buffer stock model: calibration, simulation, and empirical evaluation. Journal of Economic Perspectives 15 (3), 47–68.

Benabou, Roland, Tirole, Jean, 2002. Self-Confidence and Personal Motivation. Quarterly Journal of Economics 117 (3), 871–915.

Bernheim, B. Douglas, Rangel, Antonio, 2004. Addiction and cue-triggered decision processes. American Economic Review 94 (5), 1558–1590.

Bernheim, B. Douglas, Rangel Antonio, in press. From neuroscience to public policy: a new economic view of addiction. Swedish Economic Policy Review.

Besley, Timothy, 1988. A simple model for merit good arguments. Journal of Public Economics 35, 371–383.

Camerer, Colin, Issacharoff, Samuel, Loewenstein, George, O'Donoghue, Ted, Rabin, Matthew, 2003. Regulation for conservatives: behavioral economics and the case for 'asymmetric paternalism'. University of Pennsylvania Law Review 151 (3), 1211–1254.

Carrillo, Juan, Mariotti, Thomas, 2000. Strategic ignorance as a self-disciplining device. Review of Economic Studies 67, 529–544.

Choi, James, Laibson, David, Madrian, Brigitte, Metrick, Andrew, 2003. Optimal defaults. American Economic Review (Papers and Proceedings) 93 (2), 180–185.

DellaVigna, Stefano, Malmendier, Ulrike, 2004. Contract design and self-control: theory and evidence. Quarterly Journal of Economics 119, 353–402.

Diamond, Peter A., 1973. Consumption externalities and imperfect corrective pricing. Bell Journal of Economics 4, 526–538.

Diamond, Peter A., Mirrlees, James A., 1971a. Optimal taxation and public production: Part I. Production efficiency. American Economic Review 61 (1), 8–27.

Diamond, Peter A., Mirrlees, James A., 1971b. Optimal taxation and public production: Part II. Tax rules. American Economic Review 61 (3), 261–278.

Fischer, Carolyn, 1999. Read this paper even later: procrastination with time-inconsistent preferences. Resources for the Future Discussion Paper 99-20.

Frederick, Shane, Loewenstein, George, O'Donoghue, Ted, 2002. Time discounting and time preference: a critical review. Journal of Economic Literature 40 (2), 351–401.

Gruber, Jonathan, Koszegi, Botond, 2001. Is addiction 'rational'? Theory and evidence. Quarterly Journal of Economics 116, 1261–1303.

Gruber, Jonathan, Koszegi, Botond, 2004. Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes. Journal of Public Economics 88, 1959–1987.

Gruber, Jonathan, Mullainathan, Sendhil, 2005. Do cigarette taxes make smokers happier? B.E. Journals: Advances in Economic Analysis and Policy 5 (Article 4).

Gul, Faruk, Pesendorfer, Wolfgang, 2001. Temptation and self-control. Econometrica 69, 1403–1435.

Gul, Faruk, Pesendorfer, Wolfgang, in press. Harmful addiction. Review of Economic Studies.

[32] One can show that, in general, $d\hat{u}^{***}/dt$ is smaller for larger β. Hence, if $\hat{u}^{***}(t+0.0001) > \hat{u}^{***}(t)$ for the $\beta=1$ type, then $\hat{u}^{***}(t+0.0001) > \hat{u}^{***}(t)$ for all types with $\beta<1$ as well. Moreover, one can also show that, for all cases in Table 1, $\hat{u}^{***}(t)$ is quasi-concave. It follows that this technique does indeed identify the minimum quasi-Pareto-efficient tax.

Herrnstein, Richard, Loewenstein, George, Prelec, Drazen, Vaughan Jr., William, 1993. Utility maximization and melioration: internalities in individual choice. Journal of Behavioral Decision Making 6, 149–185.

Kahneman, Daniel, 1994. New challenges to the rationality assumption. Journal of Institutional and Theoretical Economics 150, 18–36.

Katchova, Ani, Sheldon, Ian, Miranda, Mario, 2005. A dynamic model of oligopoly and oligopsony in the U.S. potato-processing industry. Agribusiness 21 (3), 409–428.

Koszegi, Botond, in press. On the feasibility of market solutions to self-control problems. Swedish Economic Policy Review.

Kuchler, Fred, Tegene, Abebayehu, Harris, J. Michael, 2005. Taxing snack foods: manipulating diet quality or financing information programs? Review of Agricultural Economics 27 (1), 4–20.

Laibson, David, 1997. Hyperbolic discounting and golden eggs. Quarterly Journal of Economics 112 (2), 443–477.

Laibson, David, 1998. Life-cycle consumption and hyperbolic discount functions. European Economic Review 42, 861–871.

Laibson, David, Repetto, Andrea, Tobacman, Jeremy, 1998. Self-control and saving for retirement. Brookings Papers on Economic Activity 1, 91–196.

Loewenstein, George, 1996. Out of control: visceral influences on behavior. Organizational Behavior and Human Decision Processes 65, 272–292.

Loewenstein, George, O'Donoghue, Ted, 2005. "Animal Spirits: Affective and Deliberative Processes in Economic Behavior" Mimeo, Cornell University.

Loewenstein, George, Prelec, Drazen, 1992. Anomalies in intertemporal choice: evidence and an interpretation. Quarterly Journal of Economics 107 (2), 573–597.

Musgrave, Richard A., 1959. The Theory of Public Finance. McGraw-Hill, New York.

O'Donoghue, Ted, Rabin, Matthew, 1999a. Doing it now or later. American Economic Review 89 (1), 103–124.

O'Donoghue, Ted, Rabin, Matthew, 1992b. "Procrastination in preparing for retirement". In: Henry, Aaron (Ed.), Behavioral Dimensions of Retirement Economics. Brookings Institution Press and Russell Sage Foundation, Washington DC and New York, pp. 125–156.

O'Donoghue, Ted, Rabin, Matthew, 2001. Choice and procrastination. Quarterly Journal of Economics 116 (1), 121–160.

O'Donoghue, Ted, Rabin, Matthew, 2003. Studying optimal paternalism, illustrated by a model of sin taxes. American Economic Review (Papers and Proceedings) 93 (2), 186–191.

O'Donoghue, Ted, Rabin, Matthew, in press. Optimal taxes for sin goods. Swedish Economic Policy Review.

Phelps, E.S., Pollak, Robert A., 1968. On second-best national saving and game-equilibrium growth. Review of Economic Studies 35, 185–199.

Pigou, Arthur C., 1920. The Economics of Welfare. Macmillan and Company, London.

Ramsey, Frank P., 1927. A contribution to the theory of taxation. Economic Journal 37 (145), 47–61.

Sheshinski, Eytan (2002). "Bounded Rationality and Socially Optimal Limits on Choice in a Self-Selection Model." Mimeo, Hebrew University.

Sunstein, Cass, Thaler, Richard H., 2003a. Libertarian paternalism. American Economic Review (Papers and Proceedings) 93 (2), 175–179.

Sunstein, C., Thaler, R.H., 2003b. Libertarian paternalism is not an oxymoron. University of Chicago Law Review 70, 1159–1202.

Thaler, Richard H., 1991. Some empirical evidence on dynamic inconsistency. Quasi Rational Economics. Russell Sage Foundation, New York, pp. 127–133.

Thaler, Richard H., Loewenstein, George, 1992. Intertemporal choice. In: Thaler, R.H. (Ed.), The Winner's Curse: Paradoxes and Anomalies of Economic Life. Free Press, NewYork, pp. 92–106.