Optimal Sin Taxes

Ted O’Donoghue
Department of Economics
Cornell University

and

Matthew Rabin
Department of Economics
University of California, Berkeley

September 11, 2003

Abstract

Economists currently analyze optimal commodity taxation based on the assumption that people’s choices fully reflect their own best interests. In this paper, we extend optimal-tax theory by studying the welfare effects of “sin taxes” on unhealthy items, such as fatty foods, that people may (by their own reckoning) consume too much of. We use the same assumptions and tools traditionally used in public-finance theory, except that we replace the assumption that 100% of consumers have 100% self control with the assumption that some consumers may have some degree of self-control problems. We derive “optimal” taxes that correspond to the traditional “Ramsey” taxes when people have 100% self control, but differ when some people have self-control problems. We illustrate the possibility that heavy sin taxes may be optimal even if we believe the prevalence of self-control problems is relatively small. Intuitively, imposing large sin taxes on unhealthy items while lowering taxes on other items may not hurt rational consumers by much relative to the optimal Ramsey taxes, while the same change can create significant benefits for those who over-consume unhealthy items due to self-control problems. We also demonstrate that in some instances such sin taxes may create Pareto improvements relative to the Ramsey taxes. Finally, we consider alternatives to the usual linear taxes, and show that there may exist schemes that help people with self-control problems while having little or no effect on people with full self control.

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 seminar participants at Harvard, Yale, Stanford, the 2003 USC Behavioral Public Finance Conference, and the 2003 Association of Public Economic Theory Conference at Duke, and especially Emmanuel Saez. For research assistance, we thank Christoph Vanberg. For financial support, we thank the National Science Foundation (grants SES-0214043, and SES-0214147), and Rabin thanks the Russell Sage and MacArthur Foundations.

1. Introduction

Economists currently analyze optimal commodity taxation based on the assumption that people's choices fully reflect their own best interests. In this paper, we extend optimal-tax theory by studying the welfare effects of “sin taxes” on unhealthy items, such as fatty foods, that people may (by their own reckoning) consume too much of. We use the same assumptions and tools traditionally used in public-finance theory, except that we replace the assumption that 100% of consumers have 100% self control with the assumption that some consumers may have some degree of self-control problems.\(^1\)

We derive “optimal” taxes that correspond to the traditional “Ramsey” taxes when people have 100% self control, but differ when some people have self-control problems. We illustrate the possibility that heavy sin taxes may be optimal even if we believe the prevalence of self-control problems is relatively small. Intuitively, imposing large sin taxes on unhealthy items while lowering taxes on other items may not hurt rational consumers by much relative to the optimal Ramsey taxes, while the same change can create significant benefits for those who over-consume unhealthy items due to self-control problems. We also demonstrate that in some instances such sin taxes may create Pareto improvements relative to the Ramsey taxes. Finally, we consider alternatives to the usual linear taxes, and show that there may exist schemes that help people with self-control problems while having little or no effect on people with full self control.

In Section 2, we develop a model of optimal commodity taxation — in the spirit of Ramsey (1927) and Diamond and Mirrlees (1971a,1971b) — when some agents might have self-control problems. We focus on the simple case in which agents consume one “sin good” (potato chips) that creates negative health consequences in the future, and one non-sin good (carrots) that has no future health consequences. People’s own long-run preferences trade off eating pleasure against health costs, generating an ideal mix of potato chips and carrots. But since the eating pleasures of potato chips are immediate while the health consequences are delayed, self-control problems can lead people to over-consume potato chips relative to their own long-run preferences.\(^2\)

\(^1\) Two recent papers also study the welfare effects of sin taxes: Gruber and Koszegi (2002) study optimal cigarette taxation in the presence of self-control problems, and Gruber and Mullainathan (2002) provide empirical evidence that cigarette taxation increases welfare (happiness). For an example of optimal policy analysis more generally in the face of bounded rationality, see Sheshinski (2002). Some of the ideas in this paper are discussed in O’Donoghue and Rabin (2003).

\(^2\) While our formal analysis focuses on over-consumption due to self-control problems, the general
We assume the government must raise revenue exclusively from commodity taxes on potato chips and carrots. We further assume that those setting the tax rates can observe the total economy-wide demand functions for the two goods, but cannot observe the heterogeneous distribution of tastes for the two goods and degrees of self-control problems that generate those demands. We analyze optimal taxation as a function of our beliefs about the prevalence of the underlying taste parameters and about the prevalence of self-control problems in the population.

In Section 3, we present a simple example that illustrates two key points that underlie our analysis. First, for whatever combination of potato chips and carrots that we observe an individual purchase, without further information we cannot say anything about whether that behavior comes from no self-control problems, mild self-control problems, or large self-control problems. Second, the welfare implications of tax changes vary dramatically depending on whether the person has self-control problems. Moreover, because a change in the tax burden away from healthy and towards unhealthy items has only a second-order cost to 100% rational consumers but first-order benefits to those with mild self-control problems, increasing taxes substantially on potato chips may be worth doing even if we suspect most people are very close to fully self-controlled.

In Section 4, we analyze optimal taxation in a heterogenous population using a social-welfare function that puts “equal weight” on all people. If we were confident that there were no self-control problems in the population, then the optimal taxes are merely the standard “Ramsey” taxes. Because we assume no cross-price elasticities and identical own-price elasticities, in our model the Ramsey taxes always involve equal taxes on the two goods. If instead we entertain the possibility that some people have self-control problems, then potato chips should be taxed at a higher rate than carrots. More interesting, we demonstrate with some simple numerical examples that the magnitudes of such effects may be quite large — even when we believe that on average the population is close to fully self-controlled, optimal taxation may still involve significant deviations from the Ramsey taxes.3

In Section 5, we characterize Pareto-efficient taxation. If we believe there are no self-control

3 Because people with self-control problems impose negative externalities on their future selves — dubbed “negative internalities” by Herrnstein, Loewenstein, Prelec, and Vaughan (1993) — the role that sin taxes play in our analysis is much like a Pigouvian tax to correct negative externalities (Pigou (1920); also see Sandmo (1975) for an extension to the Ramsey commodity-taxation framework). Since we assume heterogeneity in preferences and in self-control problems, these negative internalities differ across individuals. Hence, because we also assume uniform taxation, our analysis is similar in spirit to analyses of uniform taxation to correct non-uniform externalities (following Diamond (1973)).
problems in the population, the Ramsey taxes are of course Pareto-efficient; however, if there is heterogeneity in preferences for potato chips and carrots, then small-enough revenue-neutral deviations from the Ramsey taxes are also Pareto-efficient. In other words, even under a belief that all consumers are 100% self-controlled, small-enough revenue-neutral shifts in taxes towards potato chips are merely movements along the Pareto frontier; although those with a stronger preference for potato chips are harmed, those with a stronger taste for carrots are helped. If, however, we believe that some agents have self-control problems, the Pareto frontier shifts towards larger potato-chip taxes and smaller carrot taxes. In fact, the Pareto frontier need not include the Ramsey taxes, in which case imposing sin taxes may yield Pareto improvements relative to the Ramsey taxes. Intuitively, increased potato-chip taxes help counteract over-consumption among people with self-control problems, while at the same time the burden of taxation is shifted away from 100% self-controlled people. Hence, if the people who believe in 100% self control are wrong, then their Ramsey taxes may in fact be Pareto-inefficient.

While our main analysis follows the literature on optimal commodity taxation by assuming lump-sum taxation is not feasible, in Section 6 we further highlight the logic behind our results by demonstrating that our main conclusions hold even if lump-sum taxation is feasible. In Section 7, we explore even more sophisticated tax schemes, such as offering a menu of tax schedules from which consumers can choose, and selling sin licenses that are required to purchase sin goods. We conclude in Section 8 by discussing some limitations of our analysis, but also its broader implications.

2. Model

In this section, we develop a model of optimal commodity taxation when some agents might have self-control problems.

Most humans exhibit a tendency to pursue immediate gratification in a way that they themselves disapprove of in the long run.\footnote{For evidence, see for instance Ainslie (1975, 1991, 1992), Ainslie and Haslam (1992a, 1992b), Loewenstein and Prelec (1992), Thaler (1991), and Thaler and Loewenstein (1992). For a recent overview, see Frederick, Loewenstein, and O’Donoghue (2002). This tendency is often referred to as “hyberbolic discounting”.
}

Recent research on such present-biased preferences — beginning with Laibson (1997) — 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 \( u_\tau \) is her instantaneous utility in period \( \tau \). This two-parameter model is a simple modification of the standard one-parameter, exponential-discounting model. The parameter \( \delta \) represents standard time-consistent impatience; for \( \beta = 1 \) these preferences reduce to exponential discounting. The parameter \( \beta \) represents a time-inconsistent preference for immediate gratification, where \( \beta < 1 \) implies an extra bias for now over the future.\(^5\) We and other researchers often refer to \( \beta \) 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 below therefore treats this preference for immediate gratification as an error (although the main points apply for essentially any welfare criterion).

In principle, the behavior of people with such self-control problems can depend on whether they are aware vs. unaware of their future self-control problems — on whether they are “sophisticated” vs. “naive”. It turns out that this distinction is irrelevant for much of our analysis, although we return to it in Section 7 when we discuss more sophisticated tax schemes.

We analyze the implications of self-control problems in a simple consumption model of the form introduced by Ramsey (1927) and Diamond and Mirrlees (1971a,1971b). Consider a quasi-linear economy with three goods, which we refer to as potato chips, carrots, and a numeraire (perhaps best interpreted as leisure). All markets are competitive, and we normalize the price of the numeraire to be one. Potato chips and carrots are produced from the numeraire at a constant marginal cost of one. The government raises revenue by taxing potato chips and carrots, and therefore the market price of these goods will be equal to one plus the commodity tax. As in the optimal-taxation literature, we assume lump-sum taxation is not feasible (except in Section 6); and as is well known, the government cannot do better by taxing sales of the numeraire. Time is discrete, and production and consumption occur in all periods.

Consider the behavior of a single consumer who faces constant market prices \( p_x \) and \( p_y \) for potato chips and carrots, respectively (these prices incorporate any commodity taxes). Let \( x_t, y_t, \) and \( z_t \) denote, respectively, her period-\( t \) consumption of potato chips, carrots, and the numeraire. We assume the person’s instantaneous utility in period \( t \) is \( u_t \equiv \rho \ln(x_t) + \sigma \ln(y_t) + z_t - \gamma \ln(x_{t-1}) \), where \( \rho, \sigma, \gamma > 0 \) are exogenous parameters. Potato chips are a “sin” good in the sense that current consumption has health or other negative consequences in the future.\(^6\) For carrots and the

---

\(^5\) 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, Repetto, and Tobacman (1998), Angeletos, Laibson, Repetto, Tobacman and Weinberg (2001), O’Donoghue and Rabin (1999a, 1999b, 2001), Fischer (1999), Carrillo and Mariotti (2000), and Benabou and Tirole (2002).

\(^6\) The fact that the negative consequences go only one period forward is not essential.
numeraire, there are no future health consequences. In period \( t \), the person faces budget constraint \( p_x x_t + p_y y_t + z_t \leq B \), where \( B \) is the person’s endowment of the numeraire (available leisure time) in each period. We assume \( B \) is sufficiently large that \( z_t > 0 \) for all \( t \) (as will become clear, the requirement is \( B > \max\{\rho - \beta \gamma, 0\} + \sigma \)).

Assuming that \( \delta = 1 \) and \( \beta \leq 1 \), in period \( t \) the person will choose \((x_t, y_t, z_t)\) to maximize
\[
u^*(x_t, y_t, z_t) \equiv \rho \ln(x_t) + \sigma \ln(y_t) + z_t - \beta \gamma \ln(x_t) = (\rho - \beta \gamma) \ln(x_t) + \sigma \ln(y_t) + z_t.\]
The person will behave the same in all periods, and in particular will spend \( \max\{\rho - \beta \gamma, 0\} \) on potato chips and \( \sigma \) on carrots.\(^7\) Hence, in each period the person will consume \( x^* = \max\{(\rho - \beta \gamma)/p_x, 0\} \), \( y^* = \sigma/p_y \), and \( z^* = B - (\max\{\rho - \beta \gamma, 0\} + \sigma) \). By contrast, optimal behavior (that maximizes long-run well-being) maximizes \( u^{**}(x_t, y_t, z_t) \equiv (\rho - \gamma) \ln(x_t) + \sigma \ln(y_t) + z_t \). Hence, in each period optimal behavior is \( x^{**} = \max\{(\rho - \gamma)/p_x, 0\} \), \( y^{**} = \sigma/p_y \), and \( z^{**} = B - (\max\{\rho - \gamma, 0\} + \sigma) \). Observe that \( x^* > x^{**} \) whenever \( \beta < 1 \) (and \( \rho - \beta \gamma > 0 \)); the self-control problem leads to over-consumption of potato chips.\(^8\)

Consider next the behavior of a population of consumers who might differ in terms of \((\sigma, \rho, \gamma, \beta)\). Suppose the population distribution of parameters \((\sigma, \rho, \gamma, \beta)\) is given by \( F(\sigma, \rho, \gamma, \beta) \), and let \( E_F[A] \) denote the expectation of \( A \) over the distribution \( F \). We let \( \bar{\sigma}, \bar{\rho}, \bar{\gamma}, \) and \( \bar{\beta} \) denote the average value for each parameter (so that, for instance, \( E_F(\rho) = \bar{\rho} \) and \( E_F(\sigma) = \bar{\sigma} \)). To simplify our analysis, we assume that \( \beta \) and \( \gamma \) are uncorrelated, and so \( E_F(\beta \gamma) = \bar{\beta} \bar{\gamma} \). To further simplify our analysis, we assume \( \rho - \beta \gamma \geq 0 \) for all types — meaning that despite the negative consequences, all consumers choose to spend some money on potato chips. Under these assumptions, the aggregate demand functions (in per capita terms) are \( X^* = E_F[x^*] = (\bar{\rho} - \bar{\beta} \bar{\gamma})/p_x \), \( Y^* = E_F[y^*] = \bar{\sigma}/p_y \), and \( Z^* = E_F[z^*] = B - (\bar{\rho} - \bar{\beta} \bar{\gamma} + \bar{\sigma}) \).

The final piece of the model is the government’s behavior. Suppose the government imposes commodity taxes on potato chips and carrots in order to raise revenue \( R \) per capita per period. Given per-unit taxes \( t_x \) and \( t_y \), the market prices for consumers will be \( p_x = 1 + t_x \) and \( p_y = 1 + t_y \), in which case aggregate demand for potato chips and carrots will be \( X^* = (\bar{\rho} - \bar{\beta} \bar{\gamma})/(1 + t_x) \) and \( Y^* = \bar{\sigma}/(1 + t_y) \). If the government must raise revenue \( R \geq 0 \), then it must choose a tax vector \((t_x, t_y)\) such that \( t_x \left(\frac{\bar{\rho} - \bar{\beta} \bar{\gamma}}{1 + t_x}\right) + t_y \left(\frac{\bar{\sigma}}{1 + t_y}\right) = R \).

In what follows, we consider the following policy problem. We — the policy analysts — recognize

---

7 We assume that peoples’s preferences and self-control problems are deterministic; if instead we assumed that preferences or self-control problems were stochastic, we would also need to consider the insurance role of the tax system, although with log utility it wouldn’t matter.

8 To further see that \( u^{**} \) is an appropriate measure of the person’s “welfare”, note that if the person consumes \((x^*, y^*, z^*) \) in all periods — as the person does in our model — then her instantaneous utility will be exactly \( u^{**}(x^*, y^*, z^*) \) in all periods (except period 1).
the possibility that there might be population heterogeneity in tastes for the two goods and in the
degree of self-control problems; but all we can observe is the aggregate demand for the two goods.
From the observed aggregate demand, we can infer \( \bar{\sigma} \) and \( (\bar{\rho} - \bar{\beta} \bar{\gamma}) \), but not the individual parameters \( \bar{\rho}, \bar{\gamma}, \) and \( \bar{\beta} \). We then analyze optimal taxation as a function of our beliefs about these parameters,
with a particular interest in how optimal taxation depends on our beliefs about \( \bar{\beta} \), the average
degree of self-control problems in the population.\(^9\)

3. An Instructive Example

In this section, we discuss an instructive example that will help illustrate the basic intuition
behind our results. Suppose we observe that every person in the economy consumes $10 worth
of potato chips and $90 worth of carrots. Because each individual’s observed demand behavior is
determined solely by \( \sigma \) and \( \rho - \beta \gamma \), and not by the individual components of \( \rho - \beta \gamma \), this behavior
is consistent with multiple types. It could come from a “fully rational” type with \( (\sigma, \rho, \gamma, \beta) =
(90, 110, 100, 1) \); or it could come from a person with a relatively mild self-control problem with
\( (\sigma, \rho, \gamma, \beta) = (90, 105, 100, .95) \); or it could come from a person with a somewhat larger self-control
problem with \( (\sigma, \rho, \gamma, \beta) = (90, 100, 100, .9) \); and so forth. Of course, among these possibilities,
only for the first case should the person consume $10 worth of potato chips. In the second case,
she should consume only $5. In the third case, she shouldn’t consume any potato chips. This
example illustrates a first key point that underlies our analysis: If we were absolutely sure that a
person has exactly zero self-control problem, we would know that her consumption of potato chips
reflects the net benefits from consuming them. Otherwise, without further information beyond
observed behavior, we know nothing about how sensible her consumption of potato chips is.

Now consider the government’s choice of commodity taxes when it must raise $5 in revenue
from the $100 total spent on potato chips and carrots (per capita per period). If all consumers
have \( \beta = 1 \), then the optimal taxation scheme would tax the two goods equally. Specifically, the
optimal “Ramsey” taxes are \( (t^*_x, t^*_y) = \left( \frac{1}{19}, \frac{1}{19} \right) \) — about 5\% tax on both items. Any other taxes

\(^9\) An alternative policy problem would be to assume that the government knows the individual
parameters \( \bar{\sigma}, \bar{\rho}, \) and \( \bar{\gamma}, \) and to analyze optimal taxation as a function of \( \bar{\beta}; \) an analysis of this
alternative problem would yield essentially the same conclusions. We prefer our approach because
it reflects the type of situations in which real-world policymakers are likely to find themselves.
They observe some demand behavior, but cannot identify whether that behavior accurately reflects
people’s underlying preferences. Hence, they must assess the degree to which factors such as self-
control problems might be influencing demand behavior — perhaps based on evidence from the
laboratory or other domains — and then assess what is an appropriate policy given those beliefs.
that raise the same $5 revenue would make consumers worse off. But suppose the government were not absolutely sure, and in particular entertained the possibility that some consumers have self-control problems — or $\beta < 1$. Such consumers over-consume potato chips, and therefore would be better off if we were to raise the same $5 in revenue via larger taxes on potato chips — to counteract over-consumption — and smaller taxes on carrots. How are different types affected by deviations from equal taxation?

Some simple calculations are revealing. If every person consumes $10 of potato chips and $90 of carrots, we could raise the same $5 in revenue with taxes of $(t_x', t_y') = \left( \frac{2}{19}, \frac{17}{361} \right)$, or with taxes of $(t_x'', t_y'') = \left( \frac{3}{19}, \frac{4}{95} \right)$, or with taxes of $(t_x''', t_y''') = \left( \frac{4}{19}, \frac{5}{133} \right)$ — that is, we could double, triple, or quadruple the taxes on potato chips, and lower the taxes on carrots accordingly. Table 1 describes how each of the three types above would be affected by each of these tax changes. Specifically, each entry is the income transfer such that the person would be as well off with the transfer (holding taxes fixed) as she would be with the tax change (and no income transfer).\(^{10}\) Hence, negative entries mean the person is harmed by deviations from equal taxation, and positive entries mean the person benefits.

Table 1: Income Transfers That Each Type Would View as Equivalent to a Tax Change from $(t_x^*, t_y^*) = \left( \frac{1}{19}, \frac{1}{19} \right)$.

<table>
<thead>
<tr>
<th>$(\sigma, \rho, \gamma, \beta)$</th>
<th>$(t_x', t_y') = \left( \frac{2}{19}, \frac{17}{361} \right)$</th>
<th>$(t_x'', t_y'') = \left( \frac{3}{19}, \frac{4}{95} \right)$</th>
<th>$(t_x''', t_y''') = \left( \frac{4}{19}, \frac{5}{133} \right)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(90, 110, 100, 1)$</td>
<td>-0.013</td>
<td>-0.049</td>
<td>-0.103</td>
</tr>
<tr>
<td>$(90, 105, 100, .95)$</td>
<td>+0.231</td>
<td>+0.428</td>
<td>+0.596</td>
</tr>
<tr>
<td>$(90, 100, 100, .9)$</td>
<td>+0.475</td>
<td>+0.905</td>
<td>+1.295</td>
</tr>
</tbody>
</table>

Fully rational consumers with $(\sigma, \rho, \gamma, \beta) = (90, 110, 100, 1)$ are, of course, hurt by deviations from equal taxation, because their consumption is distorted towards fewer potato chips and more carrots. But the harm from this distortion is relatively small — doubling potato-chip taxes causes the same harm as taking 1 cent away from the consumer, tripling potato-chip taxes causes the same harm as taking 5 cents away, and even quadrupling potato-chip taxes causes the same harm as taking 10 cents away (recall that they spend a total of $100 on potato chips and carrots). While this example is overly simple, we suspect that a similar conclusion would emerge from more realistic analyses: According to rational-choice theory, doubling, tripling, or even quadrupling taxes on potato chips is not very damaging.\(^{11}\)

\(^{10}\) Formally, each entry $\phi$ satisfies the equation $(\rho - \gamma) \ln \left( \frac{1-\beta \sigma}{1+\beta t} \right) + \sigma \ln \left( \frac{1-\beta \gamma}{1+\beta t} \right) + \phi = (\rho - \gamma) \ln \left( \frac{1-\beta \sigma}{1+\beta t} \right) + \sigma \ln \left( \frac{1-\beta \gamma}{1+\beta t} \right)$.

\(^{11}\) Indeed, we invite the empirical program of actually quantifying, under the assumption of 100% rationality, exactly what is the harm from sin taxes.
Consumers with self-control problems, in contrast, are made better off by deviations from equal taxation because higher potato-chip taxes help counteract the over-consumption of potato chips. In fact, Table 1 reveals that these consumers may benefit by far more than fully rational consumers are harmed. For instance, the same doubling of potato-chip taxes that would be like taking 1 cent away from the \((\sigma, \rho, \gamma, \beta) = (90, 110, 100, 1)\) consumer would be like giving an extra 23 cents to the \((\sigma, \rho, \gamma, \beta) = (90, 105, 100, .95)\) consumer, and like giving an extra 48 cents to the \((\sigma, \rho, \gamma, \beta) = (90, 100, 100, .9)\) consumer. And the same quadrupling of potato-chip taxes that would be like taking 10 cents away from the \((\sigma, \rho, \gamma, \beta) = (90, 110, 100, 1)\) consumer would be like giving an extra 60 cents to the \((\sigma, \rho, \gamma, \beta) = (90, 105, 100, .95)\) consumer, and like giving an extra $1.29 to the \((\sigma, \rho, \gamma, \beta) = (90, 100, 100, .9)\) consumer.

Our point is that a revenue-neutral increase in taxes on unhealthy items that has only a second-order cost to 100% rational consumers can have a first-order benefit to those with mild self-control problems. Hence, increasing taxes substantially on these items may be worth doing even if we suspect most people are very close to fully self-controlled. The next two sections directly investigate this possibility.

4. Optimal Sin Taxes

In this section, we follow Diamond and Mirrlees (1971a,1971b) and the subsequent optimal-taxation literature in specifying a social-welfare function and analyzing optimal taxation given that social-welfare function. 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), in which case the social planner will seek to maximize \(E_F [u^{**}(x^*, y^*, z^*)]\). Note that we account for \(\beta\) when predicting people’s responses to taxes — that is, we use each consumer’s actual consumption bundle \((x^*, y^*, z^*)\) — but we, like the consumers themselves in the long run, do not incorporate \(\beta\) into the social-welfare function.

\[\text{An issue arises here, because our formulation assumes the instantaneous utility function has a coefficient of 1 on the numeraire for all agents. Because multiplicative transformations of the instantaneous utility function do not affect preferences, this assumption can be interpreted as a normalization in everything we have done to this point (and in everything we do in Section 5). But since multiplicative transformations of the instantaneous utility function for a subset of agents would alter this social-welfare function, this assumption is restrictive here. Hence, by “equal weights” we literally mean equal weights combined with an assumption that the hedonic valuation of the numeraire is the same for all agents. If instead agents differ in their hedonic valuations of the numeraire, and if we ignored this fact and continued to use the social-welfare function in the text, then we would in fact be putting higher weights on those agents with lower hedonic valuations of the numeraire.}\]
function — that is, for each consumer we use $u^{**}$ and not $u^*$.

For simplicity, we also assume that $\rho - \gamma \geq 0$ for all consumers. Substituting and taking expectations yields

$$
E_F [u^{**} (x^*, y^*, z^*)] = E_F \left[ (\rho - \gamma) \ln \left( \frac{\rho - \beta \gamma}{1 + t_x} \right) + \sigma \ln \left( \frac{\sigma}{1 + t_y} \right) + B - (\rho - \beta \gamma + \sigma) \right]
$$

$$
= \left[ (\bar{\rho} - \bar{\gamma}) \ln \left( \frac{1}{1 + t_x} \right) + \bar{\sigma} \ln \left( \frac{1}{1 + t_y} \right) \right] + E_F [A(\rho, \sigma, \gamma, \beta)]
$$

where $A(\rho, \sigma, \gamma, \beta) \equiv (\rho - \gamma) \ln(\rho - \beta \gamma) + \sigma \ln \sigma + B - (\rho - \beta \gamma + \sigma)$. Because $E_F [A(\rho, \sigma, \gamma, \beta)]$ is independent of the taxes $(t_x, t_y)$, a social planner will choose $(t_x, t_y)$ to maximize

$$
\left[ (\bar{\rho} - \bar{\gamma}) \ln \left( \frac{1}{1 + t_x} \right) + \bar{\sigma} \ln \left( \frac{1}{1 + t_y} \right) \right]
$$

subject to the revenue constraint $t_x \left( \frac{\bar{\rho} - \bar{\gamma}}{1 + t_x} \right) + t_y \left( \frac{\bar{\sigma}}{1 + t_y} \right) = R$. It is straightforward to derive that the optimal taxes are (see the Appendix for a derivation):

$$
t_x^* = \frac{R}{\bar{\rho} - \bar{\gamma} + \bar{\sigma} - R} + (1 - \bar{\beta}) \frac{\bar{\gamma}(\bar{\sigma}/(\bar{\rho} - \bar{\gamma}))}{\bar{\rho} - \bar{\gamma} + \bar{\sigma} - R}
$$

and $t_y^* = \frac{R}{\bar{\rho} - \bar{\gamma} + \bar{\sigma} - R} - (1 - \bar{\beta}) \frac{\bar{\gamma}}{\bar{\rho} - \bar{\gamma} + \bar{\sigma} - R}$.

If the government were confident that agents had no self-control problems, so $\bar{\beta} = 1$, then we would reach the standard (Ramsey) conclusion that the two goods should be taxed equally (given that our model assumes no cross-price elasticities and identical own-price elasticities). For the remainder of the paper, we shall refer to equal taxation as the “Ramsey” taxes.

If instead we entertain the possibility that some agents have self-control problems, and so $\bar{\beta} < 1$, then we should indeed raise taxes on potato chips and reduce taxes on carrots (relative

---

13 Our analysis is therefore related to the “non-welfarist” approach to policy analysis. The traditional “welfarist” approach to policy analysis uses a social-welfare function that depends on only the individuals’ utilities. The non-welfarist approach assumes that, in addition to the individuals’ utilities, policymakers have other concerns, such as horizontal equity or poverty reduction (for an example of the latter, see Kanbur, Keen, and Tuomala (1994)). In our analysis, the break between $u^*$ and $u^{**}$ is due to — and recognized by — the individuals themselves. In particular, for a person with $\beta < 1$, the utility function that determines her welfare ($u^{**}$) does not correspond to the “utility function” that rationalizes her behavior ($u^*$) — which policymakers might naively take to be what she truly cares about.
to the Ramsey taxes). Furthermore, the more prevalent we believe self-control problems are — the smaller is our belief about $\bar{\beta}$ — the more we should raise taxes on potato chips and reduce taxes on carrots. These results follow straightforwardly from the intuition in Section 3: Raising the same revenue via increased taxes on potato chips and reduced taxes on carrots helps people with self-control problems by far more than it hurts fully rational people.

We are interested in more than merely the qualitative conclusion that sin taxes might be optimal; we also demonstrate with some simple numerical examples that the magnitudes of such effects might be large. As suggested by our example in Section 3, doubling, tripling, or even quadrupling sin taxes on potato chips benefits people with mild self-control problems by far more than it harms fully rational people.

In order to analyze magnitudes, we return to our basic policy problem. We observe the aggregate demand for the two goods, from which we can infer $\bar{\sigma}$ and $(\bar{\rho} - \bar{\beta} \bar{\gamma})$, but not the individual parameters $\bar{\rho}$, $\bar{\gamma}$, and $\bar{\beta}$. We analyze optimal taxation as a function of our beliefs about $\bar{\beta}$. In order to derive the optimal taxation scheme, we still need one more set of beliefs about the underlying parameters.\(^{14}\) We shall express results in terms of our beliefs about the “average” benefit-to-cost ratio for potato chips — that is, $b \equiv \bar{\rho} / \bar{\gamma}$. In addition, it is useful to re-frame the optimal-tax equations in terms of underlying ratios. We define $r \equiv R/(\bar{\rho} - \bar{\beta} \bar{\gamma} + \bar{\sigma})$, which is the relative revenue requirement (relative to total consumer expenditures on potato chips and carrots), and we define $e \equiv (\bar{\rho} - \bar{\beta} \bar{\gamma})/\bar{\sigma}$, which is the expenditure ratio of average potato-chip consumption to average carrot consumption. We can then rewrite the equations for optimal taxes as:

$$t^*_x = \frac{r}{1-r} \left[ \frac{1}{(1-\bar{\beta})(e+1)(b-1)} \right] + (1-\bar{\beta})$$

$$t^*_y = \frac{r}{1-r} \left[ \frac{e}{(1-\bar{\beta})(e+1)(b-\bar{\beta})} \right].$$

Suppose we observe that on average people spend roughly equal amounts on potato chips and carrots, or $e = 1$. Suppose further that we believe the average immediate benefit from potato-chip consumption is roughly 10% larger than the average future cost, or $b = 1.1$. For this case, Table 2 describes the optimal taxation scheme for various values of the relative revenue requirement $r$ and beliefs about the prevalence of self-control problems $\bar{\beta}$.

\(^{14}\) Mathematically, there are four unknowns $(\bar{\sigma}, \bar{\rho}, \bar{\gamma}, \bar{\beta})$, while our observations of aggregate demand and our beliefs about $\bar{\beta}$ yield only three equations.
move closer to standard Ramsey taxation (indeed, as potato chips means that on average self-control problems are less problematic, and hence we (formally, to-costs ratio for potato chips increases, there is less need to impose large sin taxes on potato chips. These are not small changes.

Of course, the magnitudes depend on the observed expenditure ratio $e$ and on our beliefs about the benefits-to-costs ratio for potato chips $b$. The optimal-tax equations imply that as the benefits-to-costs ratio for potato chips increases, there is less need to impose large sin taxes on potato chips (formally, $dt_x^*/db < 0$ and $dt_y^*/db > 0$ whenever $\beta < 1$). Intuitively, self-control problems are less problematic for a person the larger is $\rho$ relative to $\gamma$. Hence, an increase in benefits-to-costs ratio for potato chips means that on average self-control problems are less problematic, and hence we move closer to standard Ramsey taxation (indeed, as $b \to \infty$ we converge to the Ramsey taxes). Table 3 demonstrates the magnitude of this effect by increasing $b$ from 1.1 to 2 (relative to Table 2). The optimal sin taxes are significantly dampened. For $\beta = .99$, we should now raise potato-chip taxes by only roughly $\frac{1}{2}\% - \frac{1}{2}\%$ as opposed to $5\%-6\%$, and for $\beta = .95$, we should now raise potato-chip taxes by only roughly $2\% - 4\%$ as opposed to $25\%-29\%$.

Table 2: Optimal Taxation Scheme Given $e = 1$ and $b = 1.1$.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$r = 0%$</th>
<th>$r = 5%$</th>
<th>$r = 10%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta = 1$</td>
<td>$t_x^* = 0%$</td>
<td>$t_y^* = 0%$</td>
<td>$t_x^* = 11.11%$</td>
</tr>
<tr>
<td>$\beta = .99$</td>
<td>$t_x^* = 5.00%$</td>
<td>$t_y^* = -4.55%$</td>
<td>$t_x^* = 16.16%$</td>
</tr>
<tr>
<td>$\beta = .95$</td>
<td>$t_x^* = 25.00%$</td>
<td>$t_y^* = -16.67%$</td>
<td>$t_x^* = 38.89%$</td>
</tr>
</tbody>
</table>

In this case, the possibility of self-control problems has dramatic implications. If we believe that $\beta = .99$ — which might reflect a belief that everyone has $\beta = .9$, or a belief that 10% of the population has $\beta = .9$ — then we should raise taxes on potato chips and reduce taxes on carrots by roughly 5%-6% each (relative to Ramsey equal taxation). If we believe the prevalence of self-control problems is somewhat larger, $\beta = .95$, then we should raise taxes on potato chips by roughly 25%-29% and reduce taxes on carrots by roughly 17% (relative to Ramsey equal taxation). These are not small changes.

Table 3: Optimal Taxation Scheme Given $e = 1$ and $b = 2$.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$r = 0%$</th>
<th>$r = 5%$</th>
<th>$r = 10%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta = 1$</td>
<td>$t_x^* = 0%$</td>
<td>$t_y^* = 0%$</td>
<td>$t_x^* = 11.11%$</td>
</tr>
<tr>
<td>$\beta = .99$</td>
<td>$t_x^* = 0.50%$</td>
<td>$t_y^* = -0.50%$</td>
<td>$t_x^* = 11.67%$</td>
</tr>
<tr>
<td>$\beta = .95$</td>
<td>$t_x^* = 2.50%$</td>
<td>$t_y^* = -2.38%$</td>
<td>$t_x^* = 13.89%$</td>
</tr>
</tbody>
</table>

An individual’s excess consumption on potato chips (in percentage terms) is $\frac{[(\rho - \beta \gamma) - (\rho - \gamma)]}{(\rho - \gamma)} = (1 - \beta)/(\rho/\gamma - 1)$.
For the observed expenditure ratio $e$, the optimal-tax equations imply that as potato chips become a smaller proportion of people’s consumption expenditures, so $e$ decreases, optimal sin taxes increase (formally, $dt_x^*/de < 0$ and $dt_y^*/de < 0$, but $d[t_x^* - t_y^*]/de < 0$ whenever $\bar{\beta} < 1$). Intuitively, because $\bar{\beta} < 1$ implies $t_x^* > t_y^*$, the smaller is average potato-chip consumption relative to carrot consumption, the harder it is to satisfy the revenue requirement, and hence the larger taxes must be on both goods. At the same time, because a smaller portion of the people’s budgets is spent on potato chips, the harm from sin taxes is reduced, and hence the amount by which $t_x^*$ exceeds $t_y^*$ goes up. Table 4 demonstrates the magnitude of this effect by decreasing $e$ from 1 to 0.1 (relative to Table 2). The optimal sin taxes are significantly increased. For $\bar{\beta} = .99$, we should now raise potato-chip taxes by roughly 9%-11% as opposed to 5%-6%, and for $\bar{\beta} = .95$, we should now raise potato-chip taxes by roughly 45%-51% as opposed to 25%-29%.

Table 4: Optimal Taxation Scheme Given $e = 0.1$ and $b = 1.1$.  

<table>
<thead>
<tr>
<th>$\bar{\beta}$</th>
<th>$r = 0%$</th>
<th>$r = 5%$</th>
<th>$r = 10%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$t_x^* = 0%$</td>
<td>$t_x^* = 5.26%$</td>
<td>$t_x^* = 11.11%$</td>
</tr>
<tr>
<td></td>
<td>$t_y^* = 0%$</td>
<td>$t_y^* = 5.26%$</td>
<td>$t_y^* = 11.11%$</td>
</tr>
<tr>
<td>.99</td>
<td>$t_x^* = 9.09%$</td>
<td>$t_x^* = 14.83%$</td>
<td>$t_x^* = 21.21%$</td>
</tr>
<tr>
<td></td>
<td>$t_y^* = -0.83%$</td>
<td>$t_y^* = 4.39%$</td>
<td>$t_y^* = 10.19%$</td>
</tr>
<tr>
<td>.95</td>
<td>$t_x^* = 45.45%$</td>
<td>$t_x^* = 53.11%$</td>
<td>$t_x^* = 61.62%$</td>
</tr>
<tr>
<td></td>
<td>$t_y^* = -3.03%$</td>
<td>$t_y^* = 2.07%$</td>
<td>$t_y^* = 7.74%$</td>
</tr>
</tbody>
</table>

In sum, these numerical examples demonstrate that even when we believe that on average the population is close to fully self-controlled, optimal taxation may still involve significant deviations from Ramsey taxes.

5. Pareto-Efficient Sin Taxes

In this section, we characterize Pareto-efficient sin taxes. We consider the same policy problem. We observe the aggregate demand for the two goods, from which we can infer $\bar{\sigma}$ and $(\bar{\rho} - \bar{\beta} \bar{\gamma})$, but not the individual parameters $\bar{\rho}$, $\bar{\gamma}$, and $\bar{\beta}$. We then characterize Pareto-efficient taxation as a function of our beliefs about the prevalence of self-control problems in the population ($\bar{\beta}$), and as a function of our beliefs about the “average” benefit-to-cost ratio for potato chips ($b \equiv \bar{\rho} / \bar{\gamma}$).

In order to make Pareto statements, we must have beliefs about the support of preferences, and in particular about the support of people’s “true” (or long-run) relative tastes for potato chips over carrots — which we denote by $k \equiv (\rho - \gamma)/\sigma$. Specifically, the Pareto frontier is determined by the minimum and maximum values for $k$ in the population, which we denote by $k_{\min}$ and $k_{\max}$.

The population distribution of $k$ is determined by the distributions of the underlying parameters
Given what we observe, our beliefs about the “average” relative taste \( \bar{k} \equiv (\bar{\rho} - \bar{\gamma})/\bar{\sigma} \) will depend on our beliefs about \( \bar{\beta} \) (as we discuss more below).\(^{17}\) Hence, it would make little sense to assume we have beliefs about the absolute levels of \( k_{\text{min}} \) and \( k_{\text{max}} \). We derive the Pareto frontier as a function of our beliefs about the relative spread of preferences, as reflected by \( k_{\text{min}}/\bar{k} \) and \( k_{\text{max}}/\bar{k} \).

Recall that the revenue constraint is

\[
t_x \left( \frac{\bar{\rho} - \bar{\beta} \bar{\gamma}}{1+t_x} \right) + t_y \left( \frac{\bar{\sigma}}{1+t_y} \right) = R.
\]

Using the relative revenue requirement \( r \equiv R/(\bar{\rho} - \bar{\beta} \bar{\gamma} + \bar{\sigma}) \) and the expenditure ratio \( e = (\bar{\rho} - \bar{\beta} \gamma)/\bar{\sigma} \), we can rewrite the revenue constraint as

\[
e \left( \frac{t_x}{1+t_x} \right) + \left( \frac{t_y}{1+t_y} \right) = r(e+1).
\]

This equation defines a function \( \tilde{t}_y(t_x|e,r) \) that we use in our results below.\(^{18}\) Finally, we use \( t^*(r) \equiv \frac{r \bar{\sigma}}{\bar{\rho} - \bar{\beta} \gamma} \) to denote the “Ramsey” taxes.

Proposition 1 characterizes the Pareto-efficient taxation under the assumption that people have no self-control problems (all proofs are in the Appendix):

**Proposition 1.** Suppose that we observe expenditure ratio \( e \) and have relative revenue requirement \( r \), and that we have beliefs \( k_{\text{min}}/\bar{k} \) and \( k_{\text{max}}/\bar{k} \) about the spread of preferences. If we further believe that \( \bar{\beta} = 1 \), then:

1. If \( k_{\text{min}}/\bar{k} = k_{\text{max}}/\bar{k} = 1 \), then \( (t^*(r), t^*(r)) \) is the unique Pareto-efficient tax vector; and
2. If \( k_{\text{min}}/\bar{k} < k_{\text{max}}/\bar{k} \), then there exists \( t(e, r, k_{\text{max}}/\bar{k}) \) and \( \bar{t}(e, r, k_{\text{min}}/\bar{k}) \) such that tax vector \( (t_x, \tilde{t}_y(t_x|e, r)) \) is Pareto-efficient if and only if \( t(e, r, k_{\text{max}}/\bar{k}) \leq t_x \leq \bar{t}(e, r, k_{\text{min}}/\bar{k}) \), and moreover \( t(e, r, k_{\text{max}}/\bar{k}) < t^*(r) < \bar{t}(e, r, k_{\text{min}}/\bar{k}) \).

Part 1 of Proposition 1 merely reflects the usual result that, if there is no heterogeneity in preferences, then the Ramsey taxes are the unique Pareto-efficient taxes. When there is heterogeneity in preferences, the Ramsey taxes are of course still Pareto-efficient; but Part 2 establishes that a range of taxes around the Ramsey taxes are also Pareto-efficient. Intuitively, given \( \bar{\beta} = 1 \) the Ramsey taxes are optimal from the perspective of a person with average tastes (who has \( k = \bar{k} \)). If, however, we raise the same revenue via increased potato-chip taxes and decreased carrot taxes, then while people with a large relative taste for potato chips (with \( k > \bar{k} \)) are hurt, people with a small relative taste for potato chips (with \( k < \bar{k} \)) are helped. Hence, small-enough revenue-neutral sin taxes on potato chips merely move us along the Pareto frontier. (Analogously, small-enough revenue-neutral increases in carrot taxes also move us along the Pareto frontier.)

\( \sigma, \rho, \) and \( \gamma \). Using analogous notation, \( k_{\text{min}} = (\rho_{\text{min}} - \gamma_{\text{max}})/\sigma_{\text{max}} \) and \( k_{\text{max}} = (\rho_{\text{max}} - \gamma_{\text{min}})/\sigma_{\text{min}} \).\(^{17}\) Note that \( \bar{k} \) is not the average \( k \) in the population, but rather is \( k \) evaluated at the average values for the underlying parameters.

\(^{18}\) The equation is \( \tilde{t}_y(t_x|e, r) = s/(1-s) \) where \( s \equiv r(e+1) - e(t_x/(1+t_x)) \).
We next consider how our beliefs about the Pareto frontier change if we instead believe that some consumers have self-control problems:

**Proposition 2.** Suppose that we observe expenditure ratio $e$ and have relative revenue requirement $r$, and that we have beliefs $\frac{k_{\text{min}}}{\bar{k}}$ and $\frac{k_{\text{max}}}{\bar{k}}$ about the spread of preferences. If we further believe that $\bar{\beta} < 1$ and that the average benefit-to-cost ratio for potato chips is $b$, then:

1. There exists $t^\beta(e, r, k_{\text{max}}/\bar{k}, \bar{\beta}, b)$ and $\bar{t}^\beta(e, r, k_{\text{min}}/\bar{k}, \bar{\beta}, b)$ such that tax vector $(t_x, \bar{t}_y|t_x, e, r)$ is Pareto-efficient if and only if $t^\beta(e, r, k_{\text{max}}/\bar{k}, \bar{\beta}, b) \leq t_x \leq \bar{t}^\beta(e, r, k_{\text{min}}/\bar{k}, \bar{\beta}, b)$; moreover, $t^\beta$ and $\bar{t}^\beta$ are decreasing in $\bar{\beta}$, and so for any $\beta < 1$, $t^\beta(e, r, k_{\text{max}}/\bar{k}, \bar{\beta}, b) > t(e, r, k_{\text{min}}/\bar{k})$ and $\bar{t}^\beta(e, r, k_{\text{min}}/\bar{k}, \bar{\beta}, b) > \bar{t}(e, r, k_{\text{min}}/\bar{k})$.

Proposition 2 establishes that, if our beliefs about the prevalence of self-control problems shift from $\bar{\beta} = 1$ to $\bar{\beta} < 1$, then our beliefs about the Pareto frontier shift toward larger potato-chip taxes. Intuitively, if we believed that people do not have self-control problems, we would conclude that the observed expenditure ratio of potato-chip consumption to carrot consumption reflects exactly the average relative preference for potato chips over carrots — we would conclude that $e = \bar{k}$. If instead we believe that some people have self-control problems, we would conclude that the expenditure ratio *over-states* the average relative preference for potato chips over carrots — we would conclude that $e > \bar{k}$ — and the smaller are relative tastes for potato chips, the more likely people are helped by revenue-neutral increases in potato-chip taxes. Figure 1 illustrates the results from Propositions 1 and 2. Specifically, Figure 1 depicts (i) the locus of revenue-neutral tax combinations for the case of a positive revenue requirement, which is the government’s choice set, (ii) how the Pareto frontier given $\bar{\beta} = 1$ is an interval surrounding the Ramsey tax, and (iii) how the Pareto frontier given $\bar{\beta} < 1$ is shifted toward larger potato-chip taxes.

At this point, we have established analytically that, if we believe $\bar{\beta} < 1$, then the Pareto frontier shifts toward larger potato-chip taxes. Much as in the previous section, we are also interested in the quantitative magnitudes of such effects. Table 5 describes, for various values of the expenditure ratio $e$, the spread of preferences $k_{\text{min}}/\bar{k}$ and $k_{\text{max}}/\bar{k}$, and the average ratio of benefits-to-costs for potato chips $b$, how the Pareto frontier shifts as our beliefs about $\bar{\beta}$ change. As a reference, Table 5 also presents the optimal sin tax $t^*_x$ from Section 4.
Table 5: Pareto-Efficient Potato-Chip Taxes Give $r = 5\%$.

<table>
<thead>
<tr>
<th>$e$</th>
<th>$k_{\min} / \bar{k}$</th>
<th>$k_{\max} / \bar{k}$</th>
<th>$b$</th>
<th>$\tilde{\beta}$</th>
<th>$\tilde{t}_x$</th>
<th>$\bar{t}_x$</th>
<th>$\bar{t}_x^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.75</td>
<td>1.25</td>
<td>1.1</td>
<td>.99</td>
<td>-26.63%</td>
<td>22.81%</td>
<td>5.26%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.95</td>
<td>-1.05%</td>
<td>29.82%</td>
<td>10.52%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>-12.28%</td>
<td>57.89%</td>
<td>5.26%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.99</td>
<td>-8.77%</td>
<td>68.42%</td>
<td>10.52%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.95</td>
<td>5.26%</td>
<td>110.53%</td>
<td>31.58%</td>
</tr>
<tr>
<td>1</td>
<td>0.50</td>
<td>1.50</td>
<td>1.1</td>
<td>.99</td>
<td>-13.88%</td>
<td>37.16%</td>
<td>5.26%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.95</td>
<td>24.40%</td>
<td>100.96%</td>
<td>53.11%</td>
</tr>
<tr>
<td>0.1</td>
<td>0.75</td>
<td>1.25</td>
<td>1.1</td>
<td>.99</td>
<td>-26.63%</td>
<td>100.96%</td>
<td>5.26%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.95</td>
<td>-20.26%</td>
<td>120.10%</td>
<td>14.83%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>-26.63%</td>
<td>100.96%</td>
<td>5.26%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.99</td>
<td>-20.26%</td>
<td>120.10%</td>
<td>14.83%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.95</td>
<td>5.26%</td>
<td>196.65%</td>
<td>53.11%</td>
</tr>
<tr>
<td>1</td>
<td>0.75</td>
<td>1.25</td>
<td>2</td>
<td>.99</td>
<td>-5.26%</td>
<td>22.81%</td>
<td>5.26%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.95</td>
<td>-4.84%</td>
<td>23.51%</td>
<td>5.79%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>-5.26%</td>
<td>22.81%</td>
<td>5.26%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.99</td>
<td>-3.16%</td>
<td>26.32%</td>
<td>7.89%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.95</td>
<td>-3.16%</td>
<td>26.32%</td>
<td>7.89%</td>
</tr>
<tr>
<td>1</td>
<td>0.50</td>
<td>1.50</td>
<td>2</td>
<td>.99</td>
<td>-12.28%</td>
<td>57.89%</td>
<td>5.26%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.95</td>
<td>-11.93%</td>
<td>58.95%</td>
<td>5.79%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>-10.53%</td>
<td>63.16%</td>
<td>7.89%</td>
</tr>
</tbody>
</table>

Table 5 reveals that if $\tilde{\beta} < 1$, then the Pareto frontier need not include the Ramsey taxes. For instance, consider the case where $e = 1$, $k_{\min} / \bar{k} = .75$, $k_{\max} / \bar{k} = 1.25$, and $b = 1.1$. The Ramsey taxes — the optimal taxes under the usual assumption of $\tilde{\beta} = 1$ — are 5.26%. But if policymakers are wrong, and if instead $\tilde{\beta} = .95$, then those Ramsey taxes are Pareto-inefficient. In other words, imposing revenue-neutral sin taxes on potato chips can yield a Pareto improvement relative to the Ramsey taxes. How are Pareto improvements possible? The key insight is that, since increased taxes on potato chips help counteract over-consumption among people with self-control problems, they can be made better off even when the tax change increases their burden of taxation. Hence, fully rational people can be helped as well when their burden of taxation falls.

Proposition 3 characterizes more generally when Pareto improvements are possible:

**Proposition 3.** Suppose that we observe expenditure ratio $e$ and have relative revenue requirement $r$. If we further believe that $\tilde{\beta} < 1$ and that the average benefit-to-cost ratio for potato chips is $b$, then:

1. If $k_{\max} / \bar{k} < (b - \tilde{\beta})/(b - 1)$, then the Ramsey taxes ($t^{**}(r), t^{**}(r))$ are Pareto-inefficient; and
2. Taxes ($t_x, \bar{t}_y(t_x|e, r)$) with $t_x > t^{**}(r)$ are Pareto-superior to the Ramsey taxes ($t^{**}(r), t^{**}(r))$ if $k_{\max} / \bar{k} \leq (b - \tilde{\beta})k^{c}/[(b - 1)e]$ where

\[ k^{c} = \ln \left( \frac{1 + t^{**}(r)}{1 + \bar{t}_y(t_x|e, r)} \right) / \left( -\ln \left( \frac{1 + t^{**}(r)}{1 + t_x} \right) \right). \]
Part 1 establishes when the Ramsey taxes are Pareto-inefficient. Because $\beta < 1$ implies $(b - \bar{\beta})/(b - 1) > 1$, Part 1 establishes that the Ramsey taxes are always Pareto-inefficient as long as the distribution of tastes is sufficiently tight — $k^{\text{max}}/\bar{k}$ sufficiently close to one. (In a sense, this analytical point is obvious, since if there is no heterogeneity in tastes, then for each $\beta$ there is a unique Pareto optimum, and this Pareto optimum increases as $\beta$ falls.) But Part 1 also permits us to do some numerical examples to assess the plausibility of Pareto improvements. Using similar parameter values as above: If $b = 1.1$ and $\bar{\beta} = .99$, then $k^{\text{max}}$ must be within 10% of $\bar{k}$, and if $b = 1.1$ and $\bar{\beta} = .95$, then $k^{\text{max}}$ must be within 50% of $\bar{k}$.

While Part 1 characterizes when the Ramsey taxes are Pareto-inefficient — and so small enough increases in potato-chip taxes would create Pareto improvements — Part 2 characterizes when specific larger tax changes create Pareto improvements relative to the Ramsey taxes. We again use it for some numerical examples. Consider the case where benefit-to-cost ratio for potato chips is $b = 1.1$ and $\bar{\beta} = .95$, so that $k^{\text{max}}$ within 50% of $\bar{k}$ guarantees that small enough tax increases create Pareto improvements. Suppose in addition that expenditure ratio is $e = 1$ and revenue requirement is $r = 5\%$. The Ramsey taxes are $(5.26\%, 5.26\%)$. Doubling potato-chip taxes — tax vector $(10.53\%, 0.48\%)$ — would be Pareto-superior to the Ramsey taxes as long as $k^{\text{max}}$ is within 43% of $\bar{k}$. Quadrupling potato-chip taxes — tax vector $(21.05\%, -6.88\%)$ — would be Pareto-superior to the Ramsey taxes as long as $k^{\text{max}}$ is within 32% of $\bar{k}$.

A few final comments on the implications of our analysis in this section. First, at least for small enough sin taxes on potato chips, believers in 100% rationality can object to our proposal only on distributional grounds, and not based on our proposal being Pareto inefficient. Second, we should be even more wary of the standard Ramsey taxes — and more generally of a priori assuming 100% rationality — than suggested in Section 4. We have shown that even if a large portion of the population is fully self-controlled, the Ramsey taxes may in fact be Pareto inefficient. In other words, sin taxes need not be about helping people with self-control problems at the expense of rational people. This result illustrates the more general point that rigid attachment to the assumption of 100% rationality may make us fail to see policies that could in fact help all agents. Finally, our analysis in this section suggests a new way of framing the debate between believers in 100% rationality and believers in behavioral economics — namely, let’s argue less about whose beliefs are correct, and instead, if it’s possible, find a policy that is Pareto optimal under all of our beliefs.
6. What If Lump-Sum Taxes are Feasible?

Our analysis above follows the optimal-taxation literature in assuming that lump-sum taxation is not feasible. In this section, we show that sin taxes are still optimal even when lump-sum taxation is feasible.

We investigate an environment identical to that above except that, in addition to commodity taxes $t_x$ and $t_y$, the government can also impose a lump-sum tax $\ell$ (per person). In this environment, as long as people’s endowment $B$ is sufficiently large, the individual demand functions and aggregate-demand functions are unchanged. Specifically, in each period an individual with parameters $(\sigma, \rho, \gamma, \beta)$ will consume $x^* = (\rho - \beta \gamma)/(1 + t_x)$, $y^* = \sigma/(1 + t_y)$, and $z^* = B - \ell - (\rho - \beta \gamma + \sigma)$; and thus the aggregate demands for potato chips and carrots will be $X^* = (\bar{\rho} - \bar{\beta} \bar{\gamma})/(1 + t_x)$ and $Y^* = \frac{\bar{\sigma}}{1 + t_y}$. As before, we suppose the government must raise revenue $R \geq 0$. With the introduction of lump-sum taxation, the government now must choose a tax vector $(t_x, t_y, \ell)$ such that $t_x \leq (\bar{\rho} - \bar{\beta} \bar{\gamma})/(1 + t_x) + t_y \left(\frac{\bar{\sigma}}{1 + t_y}\right) + \ell = R$.

We analyze optimal taxation given the social-welfare function from Section 4. With the introduction of lump-sum taxation, this social-welfare function becomes

$$E_F[u^*(x^*, y^*, z^*)] = E_F\left[(\rho - \gamma) \ln \left(\frac{\rho - \beta \gamma}{1 + t_x}\right) + \sigma \ln \left(\frac{\sigma}{1 + t_y}\right) + (B - \ell) - (\rho - \beta \gamma + \sigma)\right]$$

$$= \left[(\bar{\rho} - \bar{\gamma}) \ln \left(\frac{1}{1 + t_x}\right) + \bar{\sigma} \ln \left(\frac{1}{1 + t_y}\right) - \ell\right] + E_F[A(\rho, \sigma, \gamma, \beta)]$$

As before, $E_F[A(\sigma, \rho, \gamma, \beta)]$ is independent of the tax vector $(t_x, t_y, \ell)$, and so a social planner will choose $(t_x, t_y, \ell)$ to maximize

$$\left[(\bar{\rho} - \bar{\gamma}) \ln \left(\frac{1}{1 + t_x}\right) + \bar{\sigma} \ln \left(\frac{1}{1 + t_y}\right) - \ell\right].$$

It is straightforward to derive that the optimal taxes are (see the Appendix for a derivation):

$$t_x^* = \frac{\bar{\gamma}(1 - \bar{\beta})}{\bar{\rho} - \bar{\gamma}}, \quad t_y^* = 0, \quad \text{and} \quad \ell^* = R - \bar{\gamma}(1 - \bar{\beta}).$$

If the government were confident that agents had no self-control problems, so $\bar{\beta} = 1$, then the government should raise all revenue via lump-sum taxation ($\bar{\beta} = 1$ implies $t_x^* = 0$ and $\ell^* = R$). This reflects the standard conclusion that lump-sum taxation is superior to commodity taxation.

---

19 Here, $B$ must be sufficiently large relative to the government’s chosen lump-sum tax $\ell$, because a person with $(\sigma, \rho, \gamma, \beta)$ will have an interior solution when $B - \ell > \rho - \beta \gamma + \sigma$. Also, we once again assume $\rho - \gamma \geq 0$ for all agents.
Notice, however, that if we believe that some agents have self-control problems, then we should still raise some revenue via sin taxes on potato chips ($\bar{\beta} < 1$ implies $t_x^* > 0$ and $\ell^* < R$). This result further highlights the basic intuition behind all our results: If some people are prone to over-consume unhealthy items, imposing sin taxes can actually make these people better off by helping to counteract this over-consumption. Hence, even when lump-sum taxation is feasible, it may still be optimal to impose sin taxes on unhealthy items.\(^{20}\)

We are again interested in more than merely the qualitative conclusion that sin taxes are still optimal; we once again demonstrate with some simple numerical examples that optimal sin taxes may be significant. In our simple environment, when lump-sum taxation is possible, the optimal tax on potato chips depends on only our beliefs about the prevalence of self-control problems in the population, as reflected by $\bar{\beta}$, and our beliefs about the average benefit-to-cost ratio for potato chips, as reflected by $b \equiv \bar{\rho}/\bar{\gamma}$. More precisely, we can rewrite the optimal potato-chip tax as $t_x^* = (1 - \bar{\beta})/(b - 1)^{21}$ Table 6 describes optimal sin taxes for potato chips for various values of $\bar{\beta}$ and $b$.

Table 6: Optimal Sin Taxes When Lump-Sum Taxation is Feasible.

<table>
<thead>
<tr>
<th>$\bar{\beta}$</th>
<th>$b = 1.1$</th>
<th>$b = 1.5$</th>
<th>$b = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$t_x^* = 0%$</td>
<td>$t_x^* = 0%$</td>
<td>$t_x^* = 0%$</td>
</tr>
<tr>
<td>0.99</td>
<td>$t_x^* = 10%$</td>
<td>$t_x^* = 2%$</td>
<td>$t_x^* = 1%$</td>
</tr>
<tr>
<td>0.95</td>
<td>$t_x^* = 50%$</td>
<td>$t_x^* = 10%$</td>
<td>$t_x^* = 5%$</td>
</tr>
</tbody>
</table>

Much as in our earlier analysis, even if we believe the prevalence of self-control problems in the population is relatively small, the implications for sin taxes can be dramatic. Also as before, these effects are most dramatic for goods where the ratio of benefits to costs is close to one, because it is for these goods that self-control problems are most damaging.

Finally, we note that our analysis of Pareto-efficient sin taxes is also effectively unchanged when lump-sum taxation is possible. The only change is that the trade-off is now between sin taxes and lump-sum taxes. Specifically, if we believe $\bar{\beta} = 1$, then raising all revenue via lump-sum taxes is (obviously) Pareto optimal; however, reducing the per-capita lump-sum tax and raising some revenue via sin taxes is merely a movement along the Pareto frontier (as long as the sin taxes are small enough). If instead we believe that $\bar{\beta} < 1$, then raising some revenue via sin taxes may yield

\(^{20}\) Note that $t_y^* = 0$ even when $\bar{\beta} < 1$ because the two goods are unrelated (zero cross-price elasticity). If potato chips and carrots were substitutes or complements, then it would be optimal to counteract over-consumption of potato chips both directly via potato-chip taxes and indirectly via carrot taxes (analogous to indirect remedies for externalities via taxes on related goods (Green and Sheshinski, 1976)).

\(^{21}\) While the revenue requirement $R$ and the expenditure ratio $e \equiv (\bar{\rho} - \bar{\beta}\gamma)/\bar{\sigma}$ do not affect $t_x^*$, they do affect $\ell^*$.  

18
a Pareto improvement relative to raising all revenue via lump-sum taxes.

7. More Sophisticated Tax Schemes

Although our analysis above imposes the realistic constraint of using linear per-unit taxes, more sophisticated schemes may in principle be feasible. To illustrate some possibilities, let us return to our example in Section 3. Suppose we observe that every person in the economy consumes $10 worth of potato chips and $90 worth of carrots. Suppose further that we are confident that people are of two types, fully rational people with \((\sigma, \rho, \gamma, \beta) = (90, 110, 100, 1)\) and people with self-control problems with \((\sigma, \rho, \gamma, \beta) = (90, 105, 100,.95)\), but we do not know the proportions of each.

If the government must raise revenue of $5 (per capita per period), then the optimal tax schedule for a population of consumers with \((\sigma, \rho, \gamma, \beta) = (90, 110, 100, 1)\) is \((t_x^*, t_y^*) = (\frac{1}{19}, \frac{1}{19})\), and the optimal tax schedule for a population of consumers with \((\sigma, \rho, \gamma, \beta) = (90, 105, 100,.95)\) is \((\hat{t}_x^*, \hat{t}_y^*) = (100\%, 0\%)\). When both types coexist, any single tax schedule makes both types worse off relative to their respective optima. But suppose consumers were offered the choice between \((t_x^*, t_y^*)\) and \((\hat{t}_x^*, \hat{t}_y^*)\) (let’s ignore implementation issues for the moment). What would they choose?

It is clear that the \((\sigma, \rho, \gamma, \beta) = (90, 110, 100, 1)\) consumers would choose taxes \((t_x^*, t_y^*)\) for the same reasons as we derived them as optimal. Determining what the \((\sigma, \rho, \gamma, \beta) = (90, 105, 100,.95)\) consumers would choose requires more effort. If they are choosing taxes that will apply in the future, their choice will depend on their beliefs about future behavior, and therefore we must make an assumption about their prediction of their own future preferences. Two common assumptions in the literature are that people fully predict their future self-control problem — they are “sophisticates” — or that they (incorrectly) predict they will have no self-control problem in the future — they are “naifs”.\(^2^2\) While these two types often behave quite differently, in this context it happens that they will make the same choice: Both will choose the \((\hat{t}_x^*, \hat{t}_y^*)\) schedule over the \((t_x^*, t_y^*)\) schedule. Sophisticates have precisely the same preferences as “we” do over their future behavior (which serves as a reminder that “our” preferences should be equal to the long-run preferences of the agents themselves), and hence will choose this tax schedule precisely because it optimally overcomes their self-control problem. Naifs, while behaving the same in the end, will think very differently. They think that in the future they will be fully self-controlled — and not want to eat many potato chips.

\(^2^2\) See O’Donoghue and Rabin (1999b,2001) for a discussion of sophistication and naivete, and an approach to modeling the intermediate cases of “partial naivete".
Hence, they also view \((\hat{t}_x^*, \hat{t}_y^*)\) as better than \((t_x^*, t_y^*)\).

In fact, it’s relatively easy to derive that whenever all consumers have identical demands, we can get perfect sorting into the type-specific optimal tax schemes.

**Proposition 4.** Suppose every agent spends \(\pi_x\) on potato chips and \(\pi_y\) on carrots, so that the expenditure ratio is \(e = \pi_x / \pi_y\). Then if we offer a tax menu in which consumers choose in advance any taxes from the set \(\{(t_x, \tilde{t}_y(t_x|e, t)) \mid t_x \geq t\}\), then every type will choose its type-specific optimal taxes.

This extreme result of perfect sorting into the type-specific optimal tax schemes is clearly not robust to more heterogeneity in preferences. But in general there may be room for improvement over the imposition of a single tax schedule. A bigger issue is whether such schemes can be implemented in the marketplace. Perhaps we could require consumers to choose one of two electronic cards needed for all purchases that would determine their tax schedules (which is becoming more feasible as more and more consumer transactions are carried out with cards of various stripes). Alternatively, perhaps we could require consumers to buy non-refundable coupons in advance that give them the right to purchase these items. If these coupons could be bought in any quantity, they would merely require a bit of foresight for 100% self-controlled consumers. The rest of us, whether sophisticated or naive, may benefit — e.g., those whose New Year’s resolutions are to eat fewer calories or quit smoking would not buy the coupons for potato chips and cigarettes.23

We can even speculate about more efficient mechanisms. Suppose it were feasible to charge well-calibrated *sin licenses* whereby people pay a one-time (or few-time) fee for the right to purchase an item. In the example above, for instance, we might raise the same amount of revenue by charging a little less than $5 per day for a license for either potato chips or carrots, and a little more than $5 per day for a license for both goods. The \((\sigma, \rho, \gamma, \beta) = (90, 110, 100, 1)\) consumers would purchase the license for both items, and end up consuming \((x, y, z) \approx (10, 90, B - 105)\). The \((\sigma, \rho, \gamma, \beta) = (90, 100, 100, .9)\) consumers would purchase solely the carrot license, and end up consuming \((x, y, z) \approx (0, 90, B - 95)\). Both types are better off than they would be under any per-unit taxation scheme — indeed, in this example, such licenses achieve the first best. (For fully rational consumers, we achieve the first best merely because the licensing scheme is equivalent to lump-sum taxation; for consumers with self-control problems, in addition the licensing scheme eliminates undesirable consumption of potato chips.)

Applying these principles outside of our framework, suppose that instead of charging a $2-per-

---

23 Of course, all this depends on somehow preventing arbitrage....
pack tax on cigarettes, we charged $5000 for a picture I.D. that allows that person to purchase up to 2500 packs tax free — and made it illegal to purchase cigarettes without this I.D. Such a policy change could help all types of consumers. All the 18-year-olds who are rationally deciding to become lifetime nicotine addicts would purchase the license. The 18-year-olds who instead end up paying $5000 in taxes for a lifetime habit they did not identify as optimal when they started would not buy the licenses. (If there were concerns that this scheme would prevent optimal experimentation, we could also issue a one-time “learner’s permit” allowing a person to purchase up to 10 packs of cigarettes.)

8. Conclusion

In this section, we conclude by discussing some limitations of our analysis, and also its broader implications.

Our analysis makes numerous simplifying assumptions, many of which we know to be important. We certainly do not expect the reader to take this model seriously as specifying optimal sin taxes. But we hope that by deriving clear conclusions in this simplified environment we are generating ideas that can be applied in a more serious analysis of optimal taxation.

We assume throughout log utility and log health costs. We do so primarily for tractability, and in particular to yield closed-form solutions, but these assumptions are potentially quite restrictive. To check, we have numerically analyzed other functional forms, and have reached two main conclusions. First, it seems somewhat robust that fully rational agents are little affected by even seemingly large deviations from the Ramsey taxes. Second, what is not robust is that small self-control problems necessarily cause significant harm. For instance, if potato chips have log benefits and linear costs, a small overweighting of those costs causes very little change in behavior, and hence very little harm (e.g., on the order of a couple cents). If self-control problems do not cause significant harm, then of course imposing sin taxes is not valuable. But in cases where they do, our conclusions still hold.

We have also ignored the possibility of substitute sin goods. If we impose sin taxes on potato chips, people might substitute out of potato chips and into Twinkies. If such goods are taxable, our analysis extends in a straightforward way. 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 policymakers fully recognize their existence, then substitute but non-taxable sins merely put limits on the effectiveness of our policy. In our framework, for instance, if $u^* = (\rho - \beta \gamma) \ln(x + \theta w) + \ldots$
\[ \sigma \ln y + z, \text{ where } w \text{ is a non-taxable sin good with market price } p_w, \text{ then a constraint on policy would be that we must have } 1 + t_x \leq p_w/\theta, \text{ because otherwise people would buy } w \text{ rather than } x. \text{ If instead policymakers naively ignore the existence of substitute and non-taxable sins, then imposing sin taxes may inadvertently do more harm than good.} \]

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. Our optimal taxes in this paper are no more a limit on choices than are the traditional Ramsey taxes — they both change choice sets, but neither reduces choice sets. Moreover, the more sophisticated schemes we discuss in Section 7 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.\(^\text{24}\)

The reader may be worried that the road to paternalism is paved with a surfeit of new regulations designed to counteract the variety of errors that people might or might not have. In many instances, however, paternalism merely means incorporating additional considerations when we set existing regulations. In the present paper, we are not proposing that the government implement new taxes; commodities are already being taxed. All we are proposing is that we introduce a new consideration — that taxes on sin goods may actually help some consumers make better decisions — into the calculus of optimal taxation.

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.\(^\text{25}\)

\(^{24}\) 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)).

\(^{25}\) See for instance O’Donoghue and Rabin (1999a, 2001) who discuss “cautious paternalism”; Camerer, Issacharo, Loewenstein, O’Donoghue, and Rabin (2003), who explore “asymmetric paternalism”; Sunstein and Thaler (2003), who investigate “libertarian paternalism”; and Choi, Laib-
While the focus on minimal interventions is a nice 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 sin taxes may appear to be more heavy-handed and invasive than some of the other cautiously paternalistic policies that we and others have advocated, our analysis reveals that in fact even relatively large sin 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.

son, Madrian, and Metrick (2003), who discuss “benign paternalism”.

23
Appendix: Derivations and Proofs

Derivation of optimal taxes \((t_x^*, t_y^*)\) for Section 4: The social planner chooses \((t_x, t_y)\) to maximize \([\tilde{p} - \gamma \ln \left(\frac{1}{1+t_x}\right) + \tilde{\sigma} \ln \left(\frac{1}{1+t_y}\right)]\) subject to the revenue constraint \(t_x \left(\frac{\tilde{p} - \beta \tilde{y}}{1+t_x}\right) + t_y \left(\frac{\tilde{\sigma} t}{1+t_x}\right) = R\).  

The revenue constraint implicitly defines a function \(t_y^R(t_x)\), and differentiating the revenue constraint yields

\[
\frac{\tilde{p} - \beta \tilde{y}}{(1+t_x)^2} + \frac{\tilde{\sigma}}{(1+t_y)^2} \frac{dt_y^R}{dt_x} = 0 \quad \text{or} \quad \frac{dt_y^R}{dt_x} = -\left(\frac{\tilde{p} - \beta \tilde{y}}{\tilde{\sigma}}\right) \left(1 + \frac{1}{1+t_x}\right)^2.
\]

Using this function, the social planner will choose \(t_x\) to maximize \([\tilde{p} - \gamma \ln \left(\frac{1}{1+t_x}\right) + \tilde{\sigma} \ln \left(\frac{1}{1+t_y(t_x)}\right)]\).  

The first-order condition is

\[
-\frac{\tilde{p} - \gamma}{1+t_x} + \frac{-\tilde{\sigma}}{1+t_y} \frac{dt_y^R}{dt_x} = -\frac{\tilde{p} - \gamma}{1+t_x} + \frac{\tilde{\sigma}}{1+t_y} \left(\frac{\tilde{p} - \beta \tilde{y}}{\tilde{\sigma}}\right) \left(1 + \frac{1}{1+t_x}\right)^2 = 0,
\]

which yields \((1 + t_y)/(1 + t_x) = (\tilde{p} - \gamma)/(\tilde{p} - \beta \tilde{y})\). Combining this condition with the revenue constraint and solving for \(t_x\) and \(t_y\) yields the optimal taxes.

Proof of Proposition 1: We first establish that for any beliefs, the Pareto frontier is an interval (possibly degenerate) on the locus of revenue-neutral tax combinations (which is the choice set for policymakers). The revenue constraint is

\[
e \left(\frac{t_x}{1+t_x}\right) + \left(\frac{t_y}{1+t_y}\right) = r(e + 1),
\]

and we know \(e\) and \(r\). As in the text, this equation defines a function \(\tilde{t}_y(t_x)\) — to simplify notation we suppress the arguments \(e\) and \(r\) — where the locus of revenue-neutral tax combinations is \(\{(t_x, \tilde{t}_y(t_x)) \mid t_x \in \mathbb{R}\}\). Hence, we can frame the policy problem as merely a choice of \(t_x\). It is straightforward to derive that \(d\tilde{t}_y/dt_x = -e(1 + \tilde{t}_y(t_x))^2/(1 + t_x)^2 < 0\). Note that the Ramsey tax \(t^*\) — again to simplify notation we suppress the argument \(r\) — is the \(t_x\) that satisfies \(\tilde{t}_y(t_x) = t_x\).

Consider type \((\sigma, \rho, \gamma, \beta)\), who has \(k = (\rho - \gamma)/\sigma\). As a function of \(t_x\), her (per-period) utility is

\[
U(t_x) = (\rho - \gamma) \ln \left(\frac{\sigma}{1+t_x}\right) + \tilde{\sigma} \ln \left(\frac{1}{1+t_y(t_x)}\right) + [B - (\rho - \beta \gamma + \sigma)].
\]

Differentiating yields

\[
\frac{dU}{dt_x} = -\frac{\rho - \gamma}{1+t_x} + \frac{-\tilde{\sigma}}{1+t_y(t_x)} \frac{d\tilde{t}_y}{dt_x} = -\frac{\rho - \gamma}{1+t_x} + \frac{-\tilde{\sigma}}{1+t_y(t_x)} (-e) \left(1 + \tilde{t}_y(t_x)\right)^2
\]

\[
= \frac{\sigma}{1+t_x} \left[-k + e \left(\frac{1 + \tilde{t}_y(t_x)}{1+t_x}\right)\right].
\]

d\tilde{t}_y/dt_x < 0 implies that for any \(k\), \(U\) is quasi-concave in \(t_x\). In addition, defining \(k^c(t_x) \equiv e \left(1 + \tilde{t}_y(t_x)/(1+t_x)\right)\), \(dU/dt_x > 0\) for all types with \(k < k^c(t_x)\), \(dU/dt_x < 0\) for all types with \(k > k^c(t_x)\), and \(dU/dt_x = 0\) for type \(k = k^c(t_x)\). It follows that \(t_x\) is Pareto-efficient if and only if \(k_{\text{min}} \leq k^c(t_x) \leq k_{\text{max}}\). Moreover, because \(k^c\) is strictly decreasing in \(t_x\), there exist \(\bar{t}\) and \(\tilde{t}\) such that \(t_x\) is
Pareto-efficient if and only if \( t \leq t_x \leq \bar{t} \), where \( t \) satisfies \( k^c(t) = k^{\text{max}} \) and \( \bar{t} \) satisfies \( k^c(\bar{t}) = k^{\text{min}} \).

Obviously, if \( k^{\text{min}} = k^{\text{max}} \) then \( \underline{t} = \bar{t} \) and there is a unique Pareto optimum.

Now suppose we believe \( \bar{\beta} = 1 \). If we believe \( \bar{\beta} = 1 \), then we’ll infer from what we observe that \( \bar{k} \equiv \frac{\bar{\rho} - \bar{\gamma}}{\bar{\sigma}} = e \), and therefore that \( k^c(t_x) = \bar{k} \left( 1 + \frac{1 + t_x}{1 + t_x} \right) \). It follows that \( \underline{t} \) satisfies \( \frac{1 + \tilde{t}_y(t_x)}{1 + t} = k^{\text{max}} / \bar{k} \) and \( \bar{t} \) satisfies \( \frac{1 + \tilde{t}_y(t_x)}{1 + t} = k^{\text{min}} / \bar{k} \). If \( k^{\text{max}} / \bar{k} = k^{\text{min}} / \bar{k} = 1 \), then the unique Pareto-efficient \( t_x \) must satisfy \( \frac{1 + \tilde{t}_y(t_x)}{1 + t} = 1 \), which holds only for \( t_x = t^{**} \) (because \( t_x = t^{**} \) implies \( \tilde{t}_y(t_x) = t_x \)). If instead \( k^{\text{max}} / \bar{k} > k^{\text{min}} / \bar{k} \), which requires \( k^{\text{max}} / \bar{k} > 1 > k^{\text{min}} / \bar{k} \), then \( \underline{t} \) satisfies \( \frac{1 + \tilde{t}_y(t_x)}{1 + t} = k^{\text{min}} / \bar{k} < 1 \), which requires \( \bar{t} > t^{**} \). The result follows.

\[ \square \]

**Proof of Proposition 2:** The first two paragraphs in the proof of Proposition 1 hold for any \( \bar{\beta} \).

It follows that there exist \( \underline{t} \) and \( \bar{t} \) such that \( t_x \) is Pareto-efficient if and only if \( \underline{t} \leq t_x \leq \bar{t} \), where \( \underline{t} \) satisfies \( k^c(\underline{t}) = k^{\text{max}} \) and \( \bar{t} \) satisfies \( k^c(\bar{t}) = k^{\text{min}} \). It remains to prove that \( \underline{t} \) and \( \bar{t} \) are both decreasing in \( \bar{\beta} \).

Given beliefs \( \bar{\beta} \leq 1 \) and \( \bar{\rho} / \bar{\gamma} = b \), and given that we observe \( e \) which we know to equal \( \frac{\bar{\rho} - \bar{\gamma}\bar{\beta}}{\bar{\sigma}} \), we’ll infer that \( \bar{k} \equiv \frac{\bar{\rho} - \bar{\gamma}}{\bar{\sigma}} = \frac{b - 1}{b - \bar{\beta}} e \), and therefore that \( k^c(t_x) = \frac{b - \bar{\beta}}{b - 1} \bar{k} \left( 1 + \frac{1 + t_x}{1 + t_x} \right) \). Hence, \( \bar{t} \) satisfies \( k^{\text{min}} / \bar{k} = \frac{b - \bar{\beta}}{b - 1} \left( \frac{1 + \tilde{t}_y(t_x)}{1 + t} \right) \), and holding fixed our beliefs about \( k^{\text{min}} / \bar{k} \), increasing \( \bar{\beta} \) requires decreasing \( \bar{t} \). Similarly, \( \underline{t} \) satisfies \( k^{\text{max}} / \bar{k} = \frac{b - \bar{\beta}}{b - 1} \left( \frac{1 + \tilde{t}_y(t_x)}{1 + t} \right) \), and holding fixed our beliefs about \( k^{\text{max}} / \bar{k} \), increasing \( \bar{\beta} \) requires decreasing \( \underline{t} \). The result follows.

\[ \square \]

**Derivation for Table 5:** We build on and use the same notation as in the proofs of Propositions 1 and 2. From the revenue constraint, it is straightforward to derive that

\[
\tilde{t}_y(t_x) = \frac{r(e + 1) - \frac{t_x}{1 + t_x} e}{1 + \frac{t_x}{1 + t_x} e - r(e + 1)}
\]

and therefore

\[
1 + \tilde{t}_y(t_x) = \frac{1}{1 + \frac{t_x}{1 + t_x} e - r(e + 1)} = \frac{1 + t_x}{t_x(1 - r)(e + 1) + (1 - r(e + 1))}
\]

From the proof of Proposition 2, \( \bar{t} \) satisfies \( k^{\text{min}} / \bar{k} = \frac{b - \bar{\beta}}{b - 1} \left( \frac{1 + \tilde{t}_y(t_x)}{1 + t} \right) \), and substituting for \( 1 + \tilde{t}_y(t) \) yields \( k^{\text{min}} / \bar{k} = \frac{b - \bar{\beta}}{b - 1} \left( \frac{1}{t_x(1 - r)(e + 1) + (1 - r(e + 1))} \right) \) or

\[
\bar{t} = \frac{\frac{b - \bar{\beta}}{b - 1} \left( k^{\text{min}} / \bar{k} \right)}{(1 - r)(e + 1)} - (1 - r(e + 1)).
\]
An analogous derivation yields

\[ t = \frac{\left(\frac{b-\bar{\beta}}{b-1}\right) \left(\frac{1}{k_{\text{max}}/k}\right) - (1 - r(e + 1))}{(1 - r)(e + 1)}. \]

**Proof of Proposition 3:** (1) It follows from the proof of Proposition 2 that the Ramsey taxes \((t^*, t^*)\) are Pareto-inefficient if and only if \(t > t^*\). Because \(t\) satisfies \(k_{\text{max}}/k = \frac{b-\bar{\beta}}{b-1} \left(\frac{1+\tilde{t}_y(t_x)}{1+t_x}\right)\), and since \(\left(1+\tilde{t}_y(t_x)\right)\) is decreasing in \(t_x\), \(t > t^*\) if and only if \(k_{\text{max}}/k < \frac{b-\bar{\beta}}{b-1} \left(\frac{1+\tilde{t}_y(t^*)}{1+t^*}\right)\). Given \(\tilde{t}_y(t^*) = t^*\), the result follows.

(2) Consider how type \((\sigma, \rho, \gamma, \beta)\) — who has \(k = (\rho - \gamma)/\sigma\) — feels about taxes \((t_x, \tilde{t}_y(t_x))\) relative to taxes \((t^*, t^*)\), where \(t_x > t^*\). Using the (per-period) utility function \(U\) from the proof of Proposition 1,

\[ U(t_x) - U(t^*) = (\rho - \gamma) \ln \left(\frac{1+t^*}{1+t_x}\right) + \sigma \ln \left(\frac{1+t^*}{1+\tilde{t}_y(t_x)}\right) \]

\[ = \sigma \left[k \ln \left(\frac{1+t^*}{1+t_x}\right) + \ln \left(\frac{1+t^*}{1+\tilde{t}_y(t_x)}\right)\right]. \]

Defining \(k^c = \ln \left(\frac{1+t^*}{1+\tilde{t}_y(t_x)}\right)/\left(-\ln \left(\frac{1+t^*}{1+t_x}\right)\right), U(t_x) > U(t^*)\) for all types with \(k < k^c\), \(U(t_x) < U(t^*)\) for all types with \(k > k^c\), and \(U(t_x) = U(t^*)\) for type \(k = k^c\). Hence, if \(k_{\text{max}} > k^c\), then type \(k_{\text{max}}\) strictly prefers \(t^*\) to \(t_x\), and so \((t_x, \tilde{t}_y(t_x))\) is not Pareto-superior to \((t^*, t^*)\). In contrast, if \(k_{\text{max}} \leq k^c\), then type \(k_{\text{max}}\) weakly prefers \(t_x\) to \(t^*\), and all other types, who all have \(k < k_{\text{max}}\), strictly prefer \(t_x\) to \(t^*\), and so \((t_x, \tilde{t}_y(t_x))\) is Pareto-superior to \((t^*, t^*)\).

Finally, from the proof of Proposition 2, given beliefs \(\bar{\beta} \leq 1\) and \(\bar{\rho}/\bar{\gamma} = b\), and given that we observe \(e\), we infer that \(\bar{k} \equiv \bar{\beta}/\bar{\gamma} = \frac{b-1}{b-\bar{\beta}} e\). We can then convert \(k_{\text{max}} \leq k^c\) into \(k_{\text{max}}/\bar{k} \leq \frac{b-\bar{\beta}}{b-1} k^c/e\), and the result follows.

\[\blacksquare\]

**Derivation of optimal taxes \((t^*_x, t^*_y, \ell^*)\) for Section 6:** The social planner chooses \((t_x, t_y, \ell)\) to maximize \([\hat{\rho} - \hat{\gamma}] \ln \left(\frac{1}{1+t_x}\right) + \hat{\sigma} \ln \left(\frac{1}{1+t_y}\right) - \ell\) subject to the revenue constraint \(t_x (\hat{\rho}/\hat{\gamma}) + t_y (\hat{\sigma}/\hat{\gamma}) + \ell = R\). The Lagrangian is

\[ L = \left[(\hat{\rho} - \hat{\gamma}) \ln \left(\frac{1}{1+t_x}\right) + \hat{\sigma} \ln \left(\frac{1}{1+t_y}\right) - \ell\right] + \lambda \left[R - t_x \left(\frac{\hat{\rho} - \hat{\gamma}}{1+t_x}\right) - t_y \left(\frac{\hat{\sigma}}{1+t_y}\right) - \ell\right], \]

and the first-order conditions are

\[ \frac{\partial L}{\partial t_x} = -\left(\frac{\hat{\rho} - \hat{\gamma}}{1+t_x}\right) - \lambda \frac{\hat{\rho} - \hat{\gamma}}{(1+t_x)^2} = 0 \]

\[ \frac{\partial L}{\partial t_y} = -\left(\frac{\hat{\sigma}}{1+t_y}\right) - \lambda \frac{\hat{\sigma}}{(1+t_y)^2} = 0 \]

\[ \frac{\partial L}{\partial \ell} = -1 - \lambda = 0. \]
The first and third conditions together imply that $1 + t_x^* = (\bar{\rho} - \bar{\beta} \gamma)/(\bar{\rho} - \bar{\gamma})$ or $t_x^* = \bar{\gamma}(1 - \bar{\beta})/(\bar{\rho} - \bar{\gamma})$. The second and third conditions together imply that $1 + t_y^* = 1$ or $t_y^* = 0$. Finally, plugging $t_x^* = \bar{\gamma}(1 - \bar{\beta})/(\bar{\rho} - \bar{\gamma})$ and $t_y^* = 0$ into the revenue constraint yields $\ell^* = R - \bar{\gamma}(1 - \bar{\beta})$.

Proof of Proposition 4: Suppose we know that a person is of type $(\sigma, \rho, \gamma, \beta)$, and we want to choose person-specific taxes that raise revenue $R$ (per-period) from this person. It is straightforward to derive that the optimal taxes for type $(\sigma, \rho, \gamma, \beta)$ maximize

$$
(\rho - \gamma) \ln \left(\frac{1}{1 + t_x}\right) + \sigma \ln \left(\frac{1}{1 + t_y}\right)
$$

subject to $t_x \left(\frac{\rho - \beta \gamma}{1 + t_x}\right) + t_y \left(\frac{\sigma}{1 + t_y}\right) = R$ (the derivation is much the same as that in Section 4).

Now, we are given that we observe expenditure ratio $\pi_x/\pi_y$. Hence, the set of taxes that each person is choosing from is \{$(t_x, t_y) \mid t_x \left(\frac{\pi_x}{1 + t_x}\right) + t_y \left(\frac{\pi_y}{1 + t_y}\right) = R$\}. Also note that the premise is that any type $(\sigma, \rho, \gamma, \beta)$ who exists in the population has $\rho - \beta \gamma = \pi_x$ and $\sigma = \pi_y$.

Consider first any type $(\sigma, \rho, \gamma, \beta)$ with $\beta = 1$. Substituting for $\pi_x$ and $\pi_y$, for this person the set of available taxes in terms of her own parameters is \{$(t_x, t_y) \mid t_x \left(\frac{\rho - \gamma}{1 + t_x}\right) + t_y \left(\frac{\sigma}{1 + t_y}\right) = R$\}. This person correctly perceives that in each period she will spend $\rho - \gamma$ on good $x$ and $\sigma$ on good $y$. Hence, her (per-period) utility as a function of her tax choice is

$$
(\rho - \gamma) \ln \left(\frac{\rho - \gamma}{1 + t_x}\right) + \sigma \ln \left(\frac{\sigma}{1 + t_y}\right) + [B - (\rho - \gamma + \sigma)].
$$

It follows that, in terms of her own parameters $(\sigma, \rho, \gamma, \beta)$, her choice of taxes will maximize

$$
(\rho - \gamma) \ln \left(\frac{1}{1 + t_x}\right) + \sigma \ln \left(\frac{1}{1 + t_y}\right)
$$

subject to $t_x \left(\frac{\rho - \gamma}{1 + t_x}\right) + t_y \left(\frac{\sigma}{1 + t_y}\right) = R$, and so she chooses her type-specific optimal taxes.

Next consider any type $(\sigma, \rho, \gamma, \beta)$ with $\beta < 1$. Substituting for $\pi_x$ and $\pi_y$, for this person the set of available taxes in terms of her own parameters is \{$(t_x, t_y) \mid t_x \left(\frac{\rho - \beta \gamma}{1 + t_x}\right) + t_y \left(\frac{\sigma}{1 + t_y}\right) = R$\}. The analysis here is slightly complicated because we must determine what the person believes about her future spending. As in O’Donoghue and Rabin (2001), a person’s beliefs about her future self-control problem, which we denote by $\hat{\beta}$, might differ from her actual future self-control problem $\beta$. She might be “sophisticated” and be fully aware of her future self-control problem, in which case $\hat{\beta} = \beta$; she might be “naive” and be fully unaware of her future self-control problem, in which case $\hat{\beta} = 1$; or she might be aware of her future self-control problem but underestimate its magnitude, in which case $\hat{\beta} \in (\beta, 1)$. The key point here is, assuming there do not exist any commitment devices, a type with beliefs $\hat{\beta}$ perceives (incorrectly when $\hat{\beta} > \beta$) that her future (per-period) spending on good $x$ will be $\rho - \hat{\beta} \gamma$. Hence, the perceived (per-period) utility as a function of her tax choice for
type \((\sigma, \rho, \gamma, \beta)\) with \(\beta < 1\) who has beliefs \(\hat{\beta}\) is

\[
(\rho - \gamma) \ln \left( \frac{\rho - \hat{\beta} \gamma}{1 + t_x} \right) + \sigma \ln \left( \frac{\sigma}{1 + t_y(t_x)} \right) + \left[ B - (\rho - \hat{\beta} \gamma + \sigma) \right]
\]

\[
= (\rho - \gamma) \ln \left( \frac{1}{1 + t_x} \right) + \sigma \ln \left( \frac{1}{1 + t_y(t_x)} \right) + A(\rho, \sigma, \gamma, \hat{\beta}).
\]

Because \(\hat{A}(\rho, \sigma, \gamma, \hat{\beta})\) is independent of the taxes \((t_x, t_y)\), \(\hat{\beta}\) is irrelevant to the person’s tax choice (of course, this conclusion depends heavily on log utility). It follows that, in terms of her own parameters \((\sigma, \rho, \gamma, \beta)\), her choice of taxes will maximize \([ (\rho - \gamma) \ln \left( \frac{1}{1 + t_x} \right) + \sigma \ln \left( \frac{1}{1 + t_y} \right) \] subject to \(t_x \left( \frac{\rho - \beta \gamma}{1 + t_x} \right) + t_y \left( \frac{\sigma}{1 + t_y} \right) = R\), and so she chooses her type-specific optimal taxes.

\[\blacksquare\]
References


Fischer, Carolyn (1999). “Read This Paper Even Later: Procrastination with Time-Inconsistent


