Abstract

We calculate the socially optimal level of illiquidity in an economy populated by households with taste shocks and (naive) present bias (i.e., $\beta$-$\delta$ preferences). The government chooses mandatory contributions to respective spending/savings accounts, each with a different pre-retirement withdrawal penalty. Penalties collected by the government are redistributed through the tax system. When households have homogeneous present bias, the social optimum is well approximated by a single account with a Pigouvian correction: the proportional penalty for pre-retirement consumption is $1 - \beta$. When households have heterogeneous present bias, the social optimum is well approximated by a three-account system: (i) a completely liquid account, (ii) a completely illiquid account, and (iii) an account with an $\approx 10\%$ early withdrawal penalty. In some ways this resembles the U.S. system, which includes completely liquid accounts, completely illiquid Social Security, and 401(k)/IRA accounts with a $10\%$ early withdrawal penalty. The social optimum is also well approximated by an even simpler two-account system – (i) a completely liquid account and (ii) a completely illiquid account – which is the most common retirement system in the world today.
1 Introduction

How much liquidity should be built into a socially optimal savings system? On the one hand, flexibility allows households to consume in ways that reflect their idiosyncratic preferences—i.e., households can respond to normatively legitimate taste shocks. However, liquidity allows households with self-control problems (and other types of biases or errors) to overconsume.

If some illiquidity is optimal, how should it be implemented? Possible forms of illiquidity include a perfectly illiquid retirement claim (like a typical defined-benefit pension or the U.S. Social Security system) or a partially liquid account (like an IRA or 401(k) plan, which allow penalty-based early withdrawals). In some cases, an optimal system would combine different types of illiquid accounts.

In the practical policy domain, there is a partial consensus on some of these questions. Almost all developed countries have some form of compulsory savings that is completely illiquid (e.g., Social Security in the United States).

Nevertheless, there are substantial differences among retirement savings systems. For example, in most developed countries, defined-contribution (DC) savings accounts have mandatory contributions and balances that are completely illiquid before age 55 (Beshears, et al., 2015). By contrast, in the United States, DC contributions are voluntary, certain types of withdrawals are allowed without penalty, and, for IRAs, withdrawals may be made for any reason if a 10% penalty is paid. Liquidity engenders significant pre-retirement “leakage”: for every $1 contributed to the DC retirement accounts of U.S. households under age 55, $0.30 to $0.40 simultaneously flows out of the 401(k)/IRA system, not counting rollovers or loans (Argento, Bryant, and Sabelhaus, 2014).\(^1\) From a theoretical perspective, it is not clear whether allowing such leakage is good or bad from the perspective of overall social welfare. Nevertheless, most media coverage bemoans leakage.\(^2\)

\(^1\)On a dollar-weighted basis, about half of these withdrawals are made in a way that allows the household to avoid the 10% penalty.

Our paper evaluates the optimality of an $N$-account retirement savings system comprised of liquid, partially illiquid, and/or completely illiquid accounts. The illiquidity is obtained with linear penalties for early withdrawals. Within this framework, we focus on systems with two accounts and systems with three accounts. However, we show that such two- and three-account systems come extremely close to delivering the welfare obtainable from a fully general (non-linear) mechanism. We do this by finding an upper bound for social welfare and show that our two- and three-account systems nearly obtain this bound.

We study preferences that include both normatively legitimate taste shifters and normatively undesirable self-control problems. The self-control problems are modeled as the consequence of present bias (Phelps and Pollak, 1968, Laibson, 1997): i.e., a discount function with weights $\{1, \beta \delta, \beta \delta^2, \ldots, \beta \delta^T\}$, where the degree of present bias is $1 - \beta$. Our model is an aggregate version (which adds interpersonal transfers) of the flexibility/commitment framework of Amador, Werning, and Angeletos (2006; hereafter referred to as AWA).3

There is a growing literature that studies how the introduction of present bias effects retirement savings and how governments should optimally respond to the distortions introduced by present bias.4 Our model is most closely related to the model of Moser and Olea de Souza e Silva (2017), who also generalize AWA by allowing for mechanisms with inter-household transfers. In their model, households have unobservable earnings ability and unobservable $\beta$, whereas we study the case of unobservable taste shocks (with exogenous earnings) and unobservable $\beta$. Moser and Olea de Souza e Silva (2017) find that second-best optimal savings institutions have many of the properties of the U.S. retirement savings system, including forced savings, a result that also emerges in AWA and in our paper. Like Moser and Olea de

---

3 Halac and Yared (2014) study the commitment vs. flexibility tradeoff with persistent shocks and show that the second-best optimal mechanism features history dependence. Bond and Sigurdsson (2018) study the commitment vs. flexibility tradeoff in three periods, identifying conditions that produce a first-best allocation.

4 For example Laibson, Repetto, and Tobacman (1997, 2003) study the design of U.S. 401(k)’s, Galperti (2015) studies optimal screening among agents with different levels of present bias, Pauluszynski and Yu (2019) study the effects of preference heterogeneity across educational groups, Yu (2021) studies screening between sophisticates and naives, Pavoni and Hakki (2017) study optimal lifecycle taxation, Maxted (2021) identifies isomorphisms between optimal policies with time consistent and present-biased agents (in economies in which agents are always in the interior of their action space).
Souza e Silva, we find that optimal savings mechanisms are characterized by more mandatory savings than currently exists in the U.S. system. Our paper contributes to this literature by showing how highly simplified retirement savings systems (e.g., two- and three-account systems with linear early-withdrawal penalties) come very close to generating welfare levels that arise under the fully general optimized non-linear mechanism.\textsuperscript{5} This adds to a literature that has shown that simple mechanisms often provide good welfare approximations to vastly more complicated optimal mechanisms.\textsuperscript{6}

Finally, a large literature studies how firms attempt to exploit agents with present bias.\textsuperscript{7} By contrast, our paper studies how a benign planner would set up a socially optimal pension scheme.

We divide our analysis into the cases of homogeneous present bias and heterogeneous present bias. In the homogeneous case all agents have the same degree of present bias—in other words, the same value of \( \beta \). In the case of homogeneous \( \beta \), our model implies that a single partially illiquid account with (Pigouvian) early-withdrawal penalty \( \pi \approx 1 - \beta \) approximates the welfare obtained by the optimal general mechanism.

We then relax the homogeneity assumption, and consider an economy in which agents have heterogeneous present bias. In this heterogeneous-preference case, we find that completely illiquid savings accounts play an important role in improving welfare. Specifically, the social optimum is well-approximated by a three-account system with a perfectly liquid savings account, a partially illiquid savings account (with an early-withdrawal penalty of approximately 13%), and a completely illiquid savings account. Even more strikingly, the social optimum is also well-approximated by an even simpler two-account system with a completely liquid savings account and a completely illiquid savings account. In both the two- and three-

\textsuperscript{5} There is a large literature on optimal taxation when consumers have present bias, including Laibson, Repetto and Tobacman (1998), Gruber and Kőszegi (2001, 2004), O’Donoghue and Rabin (2006), Lockwood (2016), Farhi and Gabaix (2018), Allcott, Lockwood and Taubinsky (forthcoming). See Bernheim and Taubinsky (2018) for a review of the literature on behavioral public economics.

\textsuperscript{6} For example, see Reichelstein (1992), Bower (1993), Sappington and Weisman (1996), Gasmi et al. (1999), McAfee (2002), Rogerson (2003), and Chu and Sappington (2007).

\textsuperscript{7} For example, see Dellavigna and Malmendier (2004, 2006), Heidhues and Kőszegi (2010), Sulka (2020), and several literature reviews: Heidhues and Kőszegi (2018), Ericson et al. (2019), and Cohen et al. (2020).
account systems the completely illiquid savings account receives a substantial mandatory contribution from the household—enough to almost smooth consumption between working life and retirement even if all other wealth is consumed during working life. The completely illiquid savings account caters to the households with the low \( \beta \) values. Fully illiquid savings generates large welfare gains for these low-\( \beta \) agents, and these welfare gains swamp the welfare losses of the high-\( \beta \) agents (who are made only slightly worse off by shifting some of their wealth from completely liquid accounts to completely illiquid accounts).

To the extent that there is a role for partially illiquid accounts in the heterogeneous-\( \beta \) economy, we find that such accounts should have low early-withdrawal penalties—in most calibrations, the penalty is slightly above 10%. This implies that the partially illiquid accounts look much like a typical 401(k) account in the U.S. Moreover, these partially illiquid accounts display a high level of leakage in equilibrium. In other words, early withdrawals (i.e., pre-retirement withdrawals) are commonplace. This leakage is a double-edged sword: it results in part from legitimate taste shocks and in part from self-control problems (i.e., low \( \beta \)). The costs of the partially illiquid account to low-\( \beta \) types (who end up paying most of the early-withdrawal penalties) and benefits to high-\( \beta \) types (who are net recipients of these penalties) are nearly offsetting, although the net effect for all of society is slightly positive.

Section 2 describes the planner’s problem—i.e., account allocations and early-withdrawal penalties that maximize social welfare subject to information asymmetries between the planner and the households. Section 2 presents the benchmark case against which we will compare all other cases: the laissez faire policy of completely liquid accounts.

Sections 3 and 4 analyze the case of homogeneous present bias. In Section 3 we follow AWA and assume that inter-household transfers are not permitted: i.e., any revenue collected by the government must be destroyed rather than being redistributed, which is referred to in the literature as the assumption of ‘money burning’. Section 4 analyzes the solution to the planner’s problem when inter-household transfers are admitted.

Section 5 analyzes the solution to the planner’s problem in the case of inter-household
transfers and heterogeneous present-bias. This is the most realistic benchmark that we study. We show that the fully optimal retirement savings system is well-approximated by two- and three-account systems, which have a completely liquid account and a (well-funded) completely illiquid account. In the three-account system, welfare is nearly identical to welfare in the two-account system and welfare in the fully general, non-linear system. Accordingly, little can be gained by advancing from two accounts to an arbitrary number of accounts \((N > 2)\). In addition, we find that the optimal partially illiquid account in the three-account system is characterized by a high rate of leakage.

Section 6 presents robustness analysis. In Section 7, we conclude and discuss the many strong assumptions that we make in our model and resulting questions of generalizability. Three online appendices are used for proofs, including our explanation of how to calculate what turns out to be a tight upper bound on the welfare that can be obtained from any \(N\)-account mechanism as well as the optimal general mechanism (Appendix 2).

## 2 Model

We study a two-period model of consumption for a continuum of households with unit mass. Households are indexed by a taste shock \(\theta\) and a present bias \(\beta\). In period 1, a household consumes \(c_1(\theta, \beta)\). In period 2, a household consumes \(c_2(\theta, \beta)\). One can think of period 1 as working life and period 2 as retirement. We will sometimes refer only to \(c_1\) and \(c_2\) for notational simplicity. When we use this short-hand notation, the dependence on \(\theta\) and \(\beta\) is implied.

In this model, we give households access to \(N\) savings accounts with initial mandatory balances \((x_n)_n^{N}\) and linear early-withdrawal penalties \((\pi_n)_n^{N}\) (which will usually turn out to be positive—i.e., we will usually find \(\pi_n > 0\)—but our framework does allow for negative penalties). We allow households to withdraw from these accounts in any order; in equilibrium, they will choose to withdraw from the low-penalty accounts first. This \(N\)-account
framework is a special case of the general mechanism; the $N$-account system is equivalent to a budget set that is piecewise linear and convex, whereas the general mechanism imposes neither of these restrictions. We show that the welfare that arises from the $N$-account framework with $N \leq 2$ is very close to the welfare for the optimal general (non-linear) mechanism. In Appendix C, we present a method for characterizing the optimal general (non-linear) mechanism. We choose to focus most of our non-appendix text on the $N$-account framework because of its similarity to the actual retirement savings systems that are currently in use around the world.

2.1 Preferences of households

We now describe the preferences of households. Let $\theta u_1(c_1)$ represent the utility flow in period 1, where $\theta$ is a stochastic taste shifter that is realized in period 1.\textsuperscript{8} Let $u_2(c_2)$ represent the utility flow in period 2. Without loss of generality, there is no taste shifter in period 2.\textsuperscript{9}

From the perspective of the self in period 1, utility flows in period 2 are discounted by a standard (normative) discount factor $\delta$, as well as an additional discount factor, $\beta$, that represents present bias, with

$$0 < \delta \leq 1,$$
$$0 < \beta \leq 1.$$  

The agent has no present bias when $\beta = 1$.

Putting these elements together, preferences in period 1 are given by

$$\theta u_1(c_1) + \beta \delta u_2(c_2),$$

\textsuperscript{8}See Atkeson and Lucas (1992) for an earlier use of such taste shifters. There are also other ways of modeling taste shifters. For example, one could assume that the utility function is $u(c - \theta)$, where $\theta$ is an additive taste shifter inside the utility function. Analyzing this case is beyond the scope of the current paper, but is part of our ongoing work.

\textsuperscript{9}Specifically, including a multiplicative taste-shifter in period 2 would not change any of our results as long as the taste shifter had expected value of one.
where $\theta$ is a stochastic taste shifter, $u_t : (0, \infty) \to \mathbb{R}$ is the period-$t$ utility function, $c_t$ is period-$t$ consumption, $\beta$ is the present bias discount factor, and $\delta$ is the standard discount factor.\footnote{This representation of household preferences can be generalized by including a second independent stochastic taste shifter (with mean 1, which is realized in period 2) that multiplies period 2’s utility function.}

### 2.2 Information structure

We assume that households are naive in the sense that they don’t anticipate their own present bias and hence won’t use commitment strategies (see Strotz, 1956; ODonoghue and Rabin, 1999a, 1999b). The assumption of partial- or full naivite is broadly supported by the empirical literature (see reviews in Ericson and Laibson 2020; Cohen et al 2021), although there are a range of results with some papers finding evidence of almost perfect sophistication, especially among experienced agents (e.g., see Alcott et al 2021). The assumption of naivite simplifies the analysis considerably by eliminating the opportunity for screening in a hypothetical ‘pre-period’.\footnote{Galperti (2015) studies screening in a contracting setting where agents are sophisticated, have private information about their degree of present bias, and contract with a firm. See also Moser and de Souza e Silva (2019) and Yu (2019).}

We assume that taste shifters, $\theta$, and present bias, $\beta$, are private information of each household in the economy. The social planner knows the aggregate distributions of these (independent) parameters. We denote the distribution function of $\theta$ by $F(\cdot)$ and denote the distribution function of $\beta$ by $G(\cdot)$.

### 2.3 Preferences of the social planner

The social planner and the household (with taste shifter $\theta$) have nearly identical preferences over consumption in periods 1 and 2. The only difference is that the social planner does not endorse present bias, implying that the planner’s preferences for an individual household are summarized by

$$\theta u_1(c_1) + \delta u_2(c_2).$$
The social planner chooses policies that maximize the social objective function:

\[
\int_{\theta, \beta} \left( \theta u_1(c_1(\theta, \beta)) + \delta u_2(c_2(\theta, \beta)) \right) dF(\theta) dG(\beta).
\]  

(1)

Note that the social planner must take account of the (endogenous) equilibrium policy functions of the households, \(c_1\) and \(c_2\). The social planner can influence these policy functions (as we explain below), but can’t control them directly because the social planner doesn’t directly observe \(\theta\) and \(\beta\) for each household (though the revelation principle will apply in our analysis). Finally, the social planner must set up a mechanism that uses total resources less than or equal to the aggregate endowment \(Y\).

Equation (1) implies that the planner has two motives in changing the allocations that would emerge under a laissez faire system. First, the planner would like to generate more savings, because only households, and not the planner, have present bias. Second, the planner would like to generate inter-personal reallocations from agents with low \(\theta\) values to agents with high \(\theta\) values. The first motive is an inter-temporal reallocation (within a single household) and the second motive is an inter-personal redistribution.

2.4 Timing

**Time 0:** The planner sets up \(N\) accounts with interest rate \(R\), where \(N\) is a constraint that we discuss in the next section. Each of the \(N\) accounts is characterized by two variables: an initial allocation \(x_n\) and a linear withdrawal penalty \(\pi_n\), which applies only to withdrawals in period 1. Because penalties apply only in period 1, we refer to them as early-withdrawal penalties. Specifically, if a consumer withdraws \(\omega\) dollars from account \(n\) in period 1 with withdrawal penalty \(\pi_n\), then the consumer receives \((1 - \pi_n)\omega\) dollars.\(^{12}\) Without loss of generality, we assume that there are no withdrawal penalties in period 2. From the planner’s perspective, the choice variables are the allocations to the \(N\) accounts, \((x_n)_{n=1}^{N}\), and the early

\(^{12}\)The framework admits negative penalties for period 1 consumption (i.e., subsidies for period 1 consumption).
(i.e., period-1) withdrawal penalties on those accounts, \((\pi_n)_{n=1}^N\).

In this framework, a completely liquid account has \(\pi_n = 0\), a partially liquid account has an early-withdrawal penalty such that \(0 < \pi_n < 1\), and a completely illiquid account has an early-withdrawal penalty \(\pi_n = 1\).

The planner must satisfy intertemporal, economy-wide budget balance. We state the budget constraint in two equivalent ways. First, the integral of equilibrium consumption over states must equal the overall resources in the economy, \(Y\). Our framework assumes a continuum population of consumers (with measure one), so integrating over taste-parameters, \(\theta\) and \(\beta\), is the same as integrating over consumers. Assuming a linear storage technology, the aggregate budget constraint can be written:

\[
\int \left( c_1(\theta, \beta) + \frac{c_2(\theta, \beta)}{R} \right) dF(\theta) dG(\beta) \leq Y.
\]

An equivalent way of describing budget balance is to relate allocations to resources. Allocations are the accounts given to each consumer. Resources are both the initial endowment and the revenue raised from penalties paid in equilibrium. Let \(\omega_n(\theta, \beta)\) be equilibrium period-1 withdrawals from account \(n\) for consumers with taste shifter \(\theta\) and present bias \(\beta\). Then the household-level budget constraints can be written:

\[
c_1(\theta, \beta) = \sum_{n=1}^N (1 - \pi_n) \omega_n(\theta, \beta),
\]

\[
c_2(\theta, \beta) = R \sum_{n=1}^N \left( x_n - \omega_n(\theta, \beta) \right).
\]

The societal budget constraint can be written

\[
\sum_{n=1}^N x_n \leq Y + \sum_{n=1}^N \left( \pi_n \int \omega_n(\theta, \beta) dF(\theta) dG(\beta) \right).
\]

Note that the account allocates resources from two sources: the societal endowment, \(Y\), and
the revenue from penalties paid from early withdrawals:

\[
\sum_{n=1}^{N} \left( \pi_{n} \int \omega_{n}(\theta, \beta) \, dF(\theta) \, dG(\beta) \right).
\]

**Time 1:** Self 1 maximizes her perceived welfare from the perspective of time 1 (which includes present bias). This will generate withdrawals from the accounts established at date 0. Her consumption will be \(c_{1}(\theta, \beta)\).

**Time 2:** Self 2 spends any remaining funds in the accounts. Her consumption will be \(c_{2}(\theta, \beta)\).

### 2.5 Summary of the \(N\)-account mechanism-design problem

We can now jointly express both the planner’s problem and the consumer’s problem. We begin with the consumer’s problem, since consumer behavior is an input to the planner’s problem. In essence, the consumer has only one decision to make.

In period 1, the consumer with parameters \(\theta\) and \(\beta\) faces the problem

\[
\max_{(\omega_{n})_{n=1}^{N}} \theta \, u_{1}(c_{1}) + \beta \, \delta \, u_{2}(c_{2}),
\]

where consumption is given by

\[
c_{1} = \sum_{n=1}^{N} (1 - \pi_{n}) \, \omega_{n},
\]

\[
c_{2} = R \sum_{n=1}^{N} (x_{n} - \omega_{n}).
\]

Conditional on the policy vectors \((x_{n})_{n=1}^{N}\) and \((\pi_{n})_{n=1}^{N}\), this generates consumption levels \(c_{1}(\theta, \beta)\) and \(c_{2}(\theta, \beta)\), where we have suppressed the dependency on \((x_{n})_{n=1}^{N}\) and \((\pi_{n})_{n=1}^{N}\).
In period 0, the planner faces the problem

$$\max_{(x_n)_{n=1}^N, (\pi_n)_{n=1}^N} \int \left( \theta u_1(c_1(\theta, \beta)) + \delta u_2(c_2(\theta, \beta)) \right) dF(\theta) dG(\beta)$$

subject to the constraints that (i) $c_1(\theta, \beta)$ and $c_2(\theta, \beta)$ are given by the consumer’s problem (equations 2-4) and (ii) economy-wide budget balance is satisfied:

$$\int \left( c_1(\theta, \beta) + \frac{c_2(\theta, \beta)}{R} \right) dF(\theta) dG(\beta) \leq Y.$$  

In other words, the planner chooses the account allocation vector, $(x_n)_{n=1}^N$, and the penalty vector, $(\pi_n)_{n=1}^N$, to maximize social surplus (equation 5) subject to the constraints that agents will exhibit present bias in their choices (equations 2-4) and that total consumption does not exceed social resources (equation 6). Although we assume the planner implements the $N$-account allocation through involuntary contributions, the planner could implement the same allocation under voluntary contributions through appropriate use of contribution subsidies (e.g., matching contributions).\textsuperscript{13} We choose to use an involuntary framing in our model presentation because it is without loss of generality and also notationally simpler (it avoids the need for matching notation) and almost all developed countries do have some involuntary retirement savings (e.g., Social Security in the United States, superannuation in Australia, the Central Provident Fund in Singapore, and the public pension system in Sweden, to pick just a few illustrative examples).\textsuperscript{14}

The $N$-account problem summarized here is a restricted version of a completely general (non-linear) mechanism-design problem (our solution method is explained in Appendix C). We compare our results to the solution of the general mechanism-design problem below.

\textsuperscript{13}For example, if the planner sets an account-specific match threshold of $z$ (i.e., the maximum voluntary contribution that can be matched) and an account-specific match rate of $m$ (i.e., the match per dollar of voluntary contributions), then for all $m$ greater than some match rate $m^*$, the equilibrium account contribution will produce a total account balance of $x = (1 + m)z$.

\textsuperscript{14}Some of these systems are funded, some of are unfunded, and some are hybrid. The key unifying feature (for the purposes of our model) is that they are all involuntary.
2.6 Laissez faire reference case: $\pi = 0$

In the analysis that follows, we will always compare social welfare to a reference case in which there are no early-withdrawal penalties—in other words, the agent has access to only one account ($x_1 = Y$), and this account has no penalty for early withdrawal ($\pi_1 = 0$). This is a pure laissez-faire system, in which the government does nothing to distort the decisions of households.

3 Optimal Liquidity with Homogeneous Present Bias and No Inter-household Transfers

In this section, we consider a first deviation from the (laissez faire) reference case. We allow the government to intervene by setting up multiple accounts and imposing early-withdrawal penalties, but we do not allow any inter-household transfers. This is equivalent to saying that any penalty revenue that is collected must be discarded/burned (instead of being transferred to other households through the government budget constraint). Such money burning is a case of theoretical interest and it has been characterized by AWA. This restriction on inter-household transfers is equivalent to assuming that

$$\sum_{n=1}^{N} x_n = Y.$$ 

In other words, the sum of the resources allocated to households (account by account) will equal the total sum of resources in society, which is $Y = 1$. (In the next section, we eliminate the money-burning restriction and accordingly allow inter-household transfers to occur through the tax/penalty system.)

In this section, we assume that all agents share a common value of $\beta$—i.e., a common degree of present bias. Hence, the distribution function $G$ is degenerate.

With the assumption of no inter-household transfers, our problem can be expressed using
our standard notation with the aggregate budget constraint replaced by a household-level budget constraint:

$$c_1 + \frac{c_2}{R} \leq Y$$

for each household. (7)

To simplify notation, we henceforth we set $\delta = 1$, $R = 1$ and $Y = 1$.\(^{15}\)

We now formulate a generalization of a theorem by AWA (2006).

We begin by denoting the support of the taste shifter $\theta$ by $\Theta = [\bar{\theta}, \bar{\theta}]$, where $0 < \underline{\theta} < \bar{\theta} < \infty$. We denote the distribution function of $\theta$ by $F : (0, \infty) \rightarrow [0, 1]$; we denote the density function of $\theta$ by $F' : (0, \infty) \rightarrow [0, \infty)$; and, following AWA (2006), we define a function $\Gamma : (0, \infty) \rightarrow \mathbb{R}$ by the formula

$$\Gamma(\theta) = (1 - \beta) \theta F'(\theta) + F(\theta).$$

Next, we define the “pooling type” $\theta_1$ to be the minimum $\theta \in (0, \bar{\theta})$ such that

$$\frac{1}{\bar{\theta}} \int_{t}^{\bar{\theta}} \Gamma(s) ds \geq 1 \text{ for all } t \in [\theta, \bar{\theta}).$$

Notice that $\theta_1$ is well-defined. Indeed, $\theta_1 > 0$ and if we denote by $\Theta_1$ the set of all $\theta \in (0, \bar{\theta})$ such that $\frac{1}{\bar{\theta}} \int_{t}^{\bar{\theta}} \Gamma(s) ds \geq 1 \text{ for all } t \in [\theta, \bar{\theta})$, then $\Theta_1$ is the non-empty half-open interval $[\theta_1, \bar{\theta})$. However, it is entirely possible that $\theta_1$ is a “hypothetical” type, in the sense that $\theta_1 < \underline{\theta}$.\(^{16}\)

Our candidate for an optimal consumption allocation is then obtained by requiring that:

(i) all types in the “separating interval” $\Theta_S = \{\theta \mid \theta \in \Theta, \ \theta < \theta_1\}$ choose freely from the unconstrained budget line, namely the set of all $(c_1, c_2)$ such that $c_1 \geq 0$, $c_2 \geq 0$ and $c_1 + c_2 = \frac{Y}{1 - \beta}$.

\(^{15}\)This involves no loss of generality because the utility function can be rescaled.

\(^{16}\)It is helpful to compare our definition of $\theta_1$ with AWA’s (2006) definition of $\theta_p$. AWA define $\theta_p$ to be the minimum value of $\theta \in (\underline{\theta}, \bar{\theta})$ such that $\int_{t}^{\bar{\theta}} (1 - \Gamma(s)) ds \leq 0$ for all $t \in [\theta, \bar{\theta})$. Hence $\theta_p$ and $\theta_1$ are related by the formula $\theta_p = \max \{\theta_1, \underline{\theta}\}$. Hence AWA’s Proposition 3 holds when $\theta_p > \underline{\theta}$, in which case, $\theta_p = \theta_1$. AWA’s Proposition 3 does not, however, hold when $\theta_p = \underline{\theta}$. To see why, consider the following counterexample. Suppose that $\bar{\theta} - \underline{\theta}$ is small and that $1 - \beta$ is large. Then offering different consumption bundles to different $\theta$ is not a priority for the planner, but preventing overconsumption is. So the planner will want to choose a pooling type strictly less than $\underline{\theta}$.
1; and (ii) all types in the “pooling interval” \( \Theta_P = \{ \theta \mid \theta \in \Theta, \ \theta \geq \theta_1 \} \) receive the allocation that the (possibly hypothetical) type \( \theta_1 \) would choose freely from the unconstrained budget line. Notice that \( \Theta_S \) may be empty, but that \( \Theta_P \) never is.

If this construction is to work, then we need to ensure that all the allocated consumption bundles lie in the interior of the unconstrained budget line. If \( \theta_1 > \underline{\theta} \), then this will be the case if and only if: (i) the most patient of the relevant types, namely \( \underline{\theta} \), would choose \( c_1 > 0 \) from the unconstrained budget line; and (ii) the least patient of the relevant types, namely \( \theta_1 \), would choose \( c_2 > 0 \) from the unconstrained budget line. If \( \theta_1 \leq \underline{\theta} \), then the only relevant type is the pooling type \( \theta_1 \), and we need only require that this type chooses both \( c_1 > 0 \) and \( c_2 > 0 \) from the unconstrained budget line.\(^{17}\)

Finally, we need to ensure that the Lagrange multiplier used in the sufficiency argument is non-negative. To that end, we assume that \( \Gamma \) is non-decreasing on the separating interval \( \Theta_S = [\underline{\theta}, \theta_1] \).\(^{18}\) Notice that, if \( \theta_1 \leq \underline{\theta} \), then \( \Theta_S \) is empty; so in that case this assumption places no restriction on \( \Gamma \).

We now enumerate all of our assumptions.

**A1** \( u_1, u_2 \) are twice continuously differentiable, with \( u_1', u_2' > 0 \) and \( u_1'', u_2'' < 0 \).

**A2** \( u_1'(0+) = u_2'(0+) = \infty \).

**A3** \( F' \) is a function of bounded variation.\(^{19}\)

**A4** \( \Gamma \) is non-decreasing on the separating interval \( \Theta_S = [\underline{\theta}, \theta_1] \).

---

\(^{17}\)A simple sufficient condition ensuring that all the allocated consumption bundles lie in the interior of the unconstrained budget line is therefore that \( u_1'(0+) = u_2'(0+) = +\infty \).

\(^{18}\)It is helpful to compare our Assumption A4 with AWA’s (2006) Assumption A. AWA assume that \( \Gamma \) is non-decreasing on the interval \( [\underline{\theta}, \theta_P] \).

\(^{19}\)Intuitively speaking, \( F' \) is a function of bounded variation if there exists a bounded Borel measure \( F'' \) on \( (0, \infty) \) such that \( F' \) is the distribution function of \( F'' \). For example, if \( F'' \) assigns mass 1 to the point 1 and mass -1 to the point 2 (and assigns no mass to any other point) then \( F' \) will be the density of the uniform distribution on \( [1, 2] \). More generally: (i) the truncation to the interval \( [\underline{\theta}, \theta] \) of the densities of most named distributions are functions of bounded variation; and (ii) any step function, the support of which is contained in \( [\underline{\theta}, \theta] \), is a function of bounded variation. See Appendices A.3 and A.4 for a detailed discussion of functions of bounded variation.
A5 $0 < \beta < 1$.

**Theorem 1** (Cf. Proposition 3 of AWA (2006).) Suppose that $\beta$ is the same for all households. Suppose further that inter-household transfers are not possible. Then welfare is maximized by dividing the endowment between two accounts: a completely liquid account (that can be used in both period 1 and period 2) and a completely illiquid account (that can be used only in period 2). In particular, types in the separating interval $\Theta_S$ – which consists of those $\theta \in \Theta$ such that $\theta < \theta_1$, and which will be empty if $\theta_1 \leq \theta$ – choose $c_1$ strictly less than the balance of the liquid account; and types in the pooling interval $\Theta_P$ – which consists of those $\theta \in \Theta$ such that $\theta \geq \theta_1$, and which is never empty – set $c_1$ equal to the balance of the liquid account (and therefore set $c_2$ equal to the balance of the completely illiquid account).\(^{20}\)

In other words, in the case of homogeneous $\beta$, no inter-household transfers and a weak restriction on the distribution function of the taste shifter $\theta$, the socially optimal allocation is achieved with only two accounts: one account that is completely liquid, and a second account that is completely illiquid in period 1 and completely liquid in period 2. Additional accounts (with intermediate levels of liquidity) do not have any value.

This theorem embeds two cases: in one case ($\theta_1 > \theta$), some types are separated and some types are pooled; and in the other case ($\theta_1 \leq \theta$), all agents are pooled. We emphasize that, in the second case, it is entirely possible that $\theta_1 < \theta$. In other words, the pooling type $\theta_1$ is a hypothetical type that is not a member of the population $\Theta$. Either way, all types $\theta \in \Theta$ with $\theta \geq \theta_1$ pool on the choice that type $\theta_1$ would make from the unconstrained budget line.

The key difference between our analysis and that of AWA (2006) is that their analysis covers the case $\theta_1 > \theta$, whereas our analysis holds for all values of $\theta_1$.\(^{21}\)

---

\(^{20}\)In particular, no money burning arises in equilibrium. See Ambrus and Egorov (2013) for cases (that do not satisfy our assumptions) in which money burning arises.

\(^{21}\)There is another important difference between our analysis and AWA’s. The original AWA proof shows that the two-account system is optimal in the class of continuous incentive-compatible consumption allocations, whereas our proof shows that the two-account system is optimal in the class of all incentive-compatible consumption allocations. This is potentially important because many incentive-compatible consumption allocations are in fact discontinuous. For example, suppose that there is a type $\theta_2 \in (\theta, \bar{\theta})$ and two consumption bundles $\mathbf{c}$ and $\bar{\mathbf{c}}$ such that all types in $[\theta, \theta_2)$ choose $\mathbf{c}$ and all types in $(\theta_2, \bar{\theta}]$ choose $\bar{\mathbf{c}}$. Then there is a jump in the allocation at $\theta_2$. 
In summary, Theorem 1 implies that no gain in welfare is achieved by increasing the number of accounts beyond \( N = 2 \) in the \( N \)-account mechanism-design problem (equations 2-6). But the theorem relies on two strong assumptions – homogeneous \( \beta \) and no inter-household transfers. We next analyze the model in the case in which the latter assumption does not hold.

4 Optimal Liquidity with Homogeneous Present Bias and Inter-Household Transfers

In the previous section, we analyzed the case in which the government had a limited set of tools: the government could not make inter-household transfers. We now study the case in which the government can make inter-household transfers. Specifically, we now replace household-by-household budget balance (Equation 7) with overall budget balance (Equation 6). With overall budget balance, we will show that a combination of a perfectly liquid and a perfectly illiquid account is not sufficient to maximize social surplus. We continue to make assumptions A1-A5. To these assumptions we add:

**A6** \( F' \) is bounded away from 0 on \((\theta, \overline{\theta})\).\(^{\text{22}}\)

**Theorem 2** Suppose that inter-household transfers are possible. A two-account system with one completely liquid account and one completely illiquid account does not maximize welfare.

Intuitively, when inter-household transfers are possible (in the interior case, with partial separation), we can use an incentive compatible mechanism to redistribute \( c_1 \) away from low-\( \theta \) types (i.e., households with low marginal utility, ceteris paribus). This theorem is proven in Appendix B.

\(^{\text{22}}\)In particular, both the right-hand limit \( F'_L(\theta) \) of \( F' \) at \( \theta \) and the left-hand limit \( F'_L(\overline{\theta}) \) of \( F' \) at \( \overline{\theta} \) are strictly positive.
4.1 Optimal policy with quasi-linear utility

Theorem 2 tells us what is not socially optimal (namely a two-account system comprised of a completely liquid account and a completely illiquid account), but it does not provide guidance on what is optimal. To gain intuition about socially optimal mechanisms, it is helpful to begin by studying the quasi-linear limit case of our model. Specifically, we consider the case in which utility is linear in period 2, i.e. \( u_2(c_2) = c_2 \). In this way, we obtain a useful exact result that captures the intuition behind the general case in which utility is concave in both periods.\(^{23}\)

**Theorem 3** Suppose that all households have the same value of \( \beta \). Suppose that inter-household transfers are possible. Assume that utility is strictly concave in the first period, linear in the second period, and the solution is interior. Then the socially optimal retirement system is a one-account system with a Pigouvian tax on consumption in period 1:

\[
\pi = 1 - \beta.
\]

This one-account system is also first-best efficient.

**Proof.** The wedge between the welfare criterion of the planner and the choice-function of the agent, which is generated by present bias \( \beta < 1 \), can be exactly offset by the early-withdrawal penalty \( \pi = 1 - \beta \). This Pigouvian tax corrects the negative internality generated by overconsumption. With this penalty, the household’s (present-biased) Euler Equation reduces to:

\[
(1 - \pi) \theta u'_1(c_1) = \beta \theta u'_1(c_1) = \beta u'_2(c_2).
\]

Crossing out identical terms, we obtain

\[ \theta u'(c_1) = u'_2(c_2), \]

which is the planner’s Euler Equation (if the planner observed \( \theta \)).

To this point, the argument does not rely on quasi-linearity, which we now deploy to prove that the resulting allocation is also first-best. At the margin, all agents are doing some consumption in period 2 (because we assume an interior solution), so for all households the value of a marginal dollar of wealth is \( u'_2(c_2) = 1 \). Accordingly, social welfare cannot be raised by changing the level of inter-household transfers. ■

Quasi-linear utility in period 2 implies that all agents have the same period-2 marginal utility (regardless of their period-2 consumption). Because marginal transfers to period 2 have the same marginal value for all agents, and because all agents have the same degree of present bias, a homogeneous Pigouvian correction achieves the first best allocation. This is not true in the general case in which the utility function is concave in both periods. However, the special case of quasi-linear utility turns out to be a good proxy for the general case. We next study that case.

### 4.2 Optimal policy with \( N \) accounts

We now return to the general case in which the utility functions in periods 1 and 2, namely \( u_1 \) and \( u_2 \), are both strictly concave (as opposed to the quasi-linear case, in which \( u_2 \) is linear). We continue to assume that inter-household transfers are possible. Theorem 2 establishes that a perfectly liquid account and a perfectly illiquid account do not jointly obtain the social optimum for this case. We now study other account structures using simulation results. Each simulation has a different assumption on the number of accounts and the scope that the planner has to set withdrawal penalties on those accounts. In our benchmark simulations, we make the following functional form assumptions.
S1. The utility functions in periods 1 and 2 are \( u_1(c) = u_2(c) = \ln(c) \);

S2. The density of the multiplicative taste shocks is a truncated normal distribution. Specifically: we start with a normal distribution (mean \( \mu = 1 \) and standard deviation \( \sigma = 0.25 \)); truncate it at the symmetrically placed points \( 1 - \chi \) and \( 1 + \chi \) (where \( \chi = 2/3 \), resulting in a distribution with support \([1 - \chi, 1 + \chi]\)); and rescale it so that it integrates to one.

Assumption S1 implies that the coefficient of relative risk aversion is one, a magnitude that often (approximately) emerges in estimates of lifecycle savings models\(^{24}\). In Section 6, we show that our paper’s findings are robust to variation in this assumption. Assumption S2 implies that a one standard deviation taste shock will induce marginal utility in period 1 to change by \( \pm 24.2\% \).\(^{25}\) We view this as a plausible assumption given the many uninsurable shocks that buffet households, but we are not aware of formal estimates of this parameter. In Section 6, we show that our paper’s findings are robust to variation in the assumed value of \( \sigma \).

We begin with Table 1, which reports the improvement in total welfare for different systems of accounts where the planner chooses the optimal \( x_n \) and \( \pi_n \). The entries in this table are percentage welfare improvements with the laissez-faire case as the benchmark. Specifically, each entry tells us how much social welfare improves expressed as the equivalent percentage improvement in the societal resource endowment; this is typically referred to as a money metric welfare criterion. We use this welfare reporting framework throughout the rest of the paper (with the laissez-faire case as our benchmark in all analyses).

The first row of Table 1 reports the case of one (flexible) account. We refer to the account as flexible to emphasize that the planner sets the penalty-level, \( \pi_1 \), for this account (as well as the mandatory initial balance \( x_1 \)). The second row of Table 1 reports the case of two (flexible) accounts: now the planner sets \( \pi_1 \) and \( \pi_2 \) (as well as \( x_1 \) and \( x_2 \)).

\(^{24}\)For example, see Gourinchas and Parker (1999) and Laibson et al (2021).

\(^{25}\)This is slightly less than \( \sigma = 0.25 \) because of the truncation of the (deep) tails of the distribution of taste shocks.
Table 1: The welfare gain from four mechanisms (namely a single account with an endogenous penalty (row 1); two accounts with endogenous penalties (row 2); the general mechanism (relaxed), which allows for an arbitrary non-linear budget set and does not impose the monotonicity restriction (row 3); and a two-account system with one completely liquid account and one completely illiquid account (row 4)) calculated for 10 different values of $\beta$ (namely 0.1, 0.2, ..., 1.0) in the homogeneous-$\beta$ model. The welfare gain is calculated as the percentage increase in household wealth that would produce the same average welfare in the laissez-faire case. Welfare is calculated using the planner’s welfare criterion (i.e., without present bias in the welfare objective).

<table>
<thead>
<tr>
<th></th>
<th>Value of $\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td>1 Flexible</td>
<td>69.66</td>
</tr>
<tr>
<td>2 Flexible</td>
<td>71.65</td>
</tr>
<tr>
<td>General Mechanism (Relaxed)</td>
<td>71.67</td>
</tr>
<tr>
<td>1 Liquid, 1 Illiquid</td>
<td>71.63</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Value of $\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.6</td>
</tr>
<tr>
<td>1 Flexible</td>
<td>2.79</td>
</tr>
<tr>
<td>2 Flexible</td>
<td>2.86</td>
</tr>
<tr>
<td>General Mechanism (Relaxed)</td>
<td>2.88</td>
</tr>
<tr>
<td>1 Liquid, 1 Illiquid</td>
<td>2.54</td>
</tr>
</tbody>
</table>

The third row of Table 1 reports an upper bound for welfare improvements that can be achieved with the completely general non-linear mechanism. This general non-linear mechanism does not restrict the planner to use a finite set of accounts that each have a linear early-withdrawal penalty. The general non-linear mechanism allows the planner to offer households a non-linear budget set from which each household can pick a consumption pair, $(c_1, c_2)$. This non-linear mechanism is solved numerically using the differential equations.

Our calculation of welfare improvements in the general non-linear mechanism is an upper bound on welfare, because we omit a monotonicity restriction when we calculate the optimum, leading us to refer to this as a ‘relaxed’ case of the general mechanism. In practice, this monotonicity restriction has little or even no effect on welfare. See Appendix C for a complete description of our analysis of the general non-linear mechanism. We know that our reported upper bound for welfare improvements in the general non-linear mechanism is tight.
in practice because it is nearly identical to the welfare improvements that we calculate for highly restricted finite-account cases with a small number of linear accounts (e.g., $N = 2$), as we will show below.\footnote{There is one final sense in which our general non-linear mechanism is an upper bound on welfare because it relaxes restrictions imposed by the $N$-account framework. Specifically, the $N$-account framework requires the budget set to be convex (so that low-penalty accounts are depleted before high-penalty accounts), whereas the general non-linear mechanism does not impose convexity.}

The fourth row of Table 1 reports the case of two accounts: a completely liquid account and a completely illiquid account.

The columns of Table 1 represent different cases of homogeneous $\beta$, starting with $\beta = 0.1$ and progressing to $\beta = 1.0$.\footnote{There is a growing literature on estimation of present bias (e.g., Dellavigna and Paserman, 2005; Shapiro, 2005; Dellavigna and Malmendier, 2006; Gine, Karlan, and Zinman, 2010; Meier and Sprenger, 2010; Augenblick, Niederle, and Sprenger, 2015; see Cohen, et al., forthcoming for a review of this literature).}

Table 1 reveals that a simple single-account system generates most of the obtainable welfare gains. For example, consider the column $\beta = 0.6$ (a natural value for a homogeneous calibration in light of current estimates in the empirical literature—see Cohen, et al., forthcoming). In this column, one flexible account generates a social-welfare gain equal to 2.794\% of the endowment (relative to the laissez-faire reference case). Two flexible accounts generate a social-welfare gain equal to 2.860\% of the endowment. The general non-linear mechanism generates a welfare gain that is weakly bounded above by 2.881\% of the endowment.

This analysis also reveals another important feature of the homogeneous case: the optimal penalties are essentially Pigouvian corrections to present bias. We can see this in Figure 1, where we report the optimal penalty for the one-account case, which turns out to be nearly identical to $(1 - \beta)$, both of which are plotted in Figure 1. This near-Pigouvian result echoes the exact Pigouvian correction that arises in the quasi-linear case (subsection 4.1).\footnote{Similar Pigouvian taxes also arise in the cases with more than one account. For example, with $\beta = 0.6$ and two accounts, the penalties on those two accounts are respectively 0.32 and 0.42, straddling the exact Pigouvian correction of $1 - \beta = 0.4$.}

Finally, note that an exact Pigouvian correction (which did arise in the quasi-linear case) is not generally socially optimal, because there is an inter-household redistributive motive for the planner when both $u_1$ and $u_2$ are concave. In this general case, the planner would
Figure 1: The optimal penalty $\pi^*$ and the notional Pigouvian tax $1 - \beta$ as a function of $\beta$ in the case in which: (i) the population has homogeneous $\beta$; (ii) the planner is confined to a mechanism with a single account, with penalty $\pi$. Note that $\pi^*$ is always lower than $1 - \beta$. In particular, $\pi^*$ is negative at $\beta = 1$. This is due to the redistributitional motives of the planner: she wishes to redistribute from types with low $\theta$ to types with high $\theta$.

like to reallocate resources from low-$\theta$ types to high-$\theta$ types. This redistributive motive is reflected in the fact that the socially optimal penalties in the one-account case (for any given value of $\beta$) are all strictly below the corresponding value of $(1 - \beta)$. Intuitively, the households who will be paying the penalties are those households with the higher $\theta$ values. To transfer resources to these households, the planner lowers the socially optimal penalty below the $(1 - \beta)$ benchmark value. However, as one can see in Figure 1, this downward adjustment is small in magnitude. Accordingly, the Pigouvian correction is the dominant force in these simulations.
5 Optimal Liquidity with Heterogeneous Present Bias and Inter-Household Transfers

In this section, we continue to allow inter-household transfers. In addition, we now relax the assumption that consumers have homogeneous $\beta$. As in the previous section, we begin with the quasi-linear case and then provide quantitative simulations.

5.1 Optimal policy with heterogeneous present bias and quasi-linear utility

The efficacy of Pigouvian taxation generalizes to an economy with heterogeneous present bias and quasi-linear utility, even though the planner does not directly observe each household’s $\beta$. To prove this, we exploit the revelation principle and study mechanisms in which agents reveal their intertemporal preferences (between periods 1 and 2). Note that a household’s preferences are

$$\theta u_1(c_1) + \beta u_2(c_2).$$

This representation is observationally equivalent (in choices) to the preferences,

$$\frac{\theta}{\beta} u_1(c_1) + u_2(c_2).$$

Accordingly, intertemporal preferences have a sufficient statistic, $\phi = \theta / \beta$, and the revelation principle can be implemented with the variable $\phi$. We will study general (non-linear) mechanisms in which the agents each report $\phi$ and receive a consumption pair $(c_1, c_2)$ which depends on their reported value of $\phi$. Now that we study a heterogeneous population of $\beta$-types, we amend our assumptions (A1-A6) by requiring that the support of $\beta$ be bounded strictly below by 0 and weakly above by 1 (generalizing A5), and omitting assumption A4 (which assumes homogeneous $\beta$).
Theorem 4 Suppose that inter-household transfers are possible. Assume that utility is strictly concave in the first period and linear in the second period, that the solution is interior and that $E[\theta \mid \phi]$ is non-decreasing in $\phi = \theta / \beta$. Then the optimal allocation is characterized by

$$E[\theta \mid \phi] u'_1(c_1(\phi)) = 1,$$

and the implied (local) marginal penalty rate for period 1 withdrawals is

$$\pi(\phi) = E[1 - \beta \mid \phi].$$

Note that this penalty is an ‘average Pigouvian correction,’ in the sense that the marginal dollar of consumption in period 1 is penalized with the conditional expected value of $1 - \beta$, where the conditioning is done with respect to the (truthfully) reported value of $\phi$. The proof of Theorem 4 is in Appendix D.

Theorem 4 tells us that an optimal mechanism will generate penalties ranging from $\pi = 1 - \overline{\beta}$ to $\pi = 1 - \underbar{\beta}$. Intuitively, households who report the lowest feasible value of $\phi = \theta / \beta$ are revealing that they have the highest feasible value of $\beta = \overline{\beta}$ and the lowest feasible value of $\theta$. They will consume relatively little in period 1 and will face a penalty of $\pi = 1 - \overline{\beta}$ for this consumption, which is an exact Pigouvian correction because their $\overline{\beta}$ value has been fully revealed. An identical argument applies for households who report the highest feasible value of $\phi = \theta / \beta$. They have revealed their $\beta = \overline{\beta}$ value, and accordingly will receive an exact Pigouvian correction embodied by a penalty of $\pi = 1 - \overline{\beta}$.

5.2 Optimal policy with $N$ accounts

We now switch from the case of quasi-linear utility to the case in which the consumer has log utility in both periods. In the current and the following subsections, we study optimal
Table 2: The welfare gain from five mechanisms (namely a single account with an endogenous penalty (row 1); two accounts with endogenous penalties (row 2); the general mechanism (relaxed), which allows for an arbitrary non-linear budget set and does not impose the monotonicity restriction (row 3); a two-account system with one completely liquid account and one completely illiquid account (row 4); and a three-account system with one completely liquid account, one account with an endogenous penalty and one completely illiquid account (row 5)) in the heterogeneous-β model (with β distributed uniformly between 0.2 and 1).

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>Welfare Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Flexible</td>
<td>3.569</td>
</tr>
<tr>
<td>2 Flexible</td>
<td>6.136</td>
</tr>
<tr>
<td>General Mechanism (Relaxed)</td>
<td>6.144</td>
</tr>
<tr>
<td>1 Liquid, 1 Illiquid</td>
<td>6.105</td>
</tr>
<tr>
<td>1 Liquid, 1 Flexible, 1 Illiquid</td>
<td>6.137</td>
</tr>
</tbody>
</table>

Table 2 reveals that a one-account system no longer obtains most of the feasible welfare gains: one flexible account generates a social-welfare gain of only 3.569% of the endowment.
(Following the approach in Table 1, all welfare gains in Table 2 are reported relative to the laissez faire reference case.) With just one account, the system cannot generate the necessary heterogeneity in marginal penalties for early consumption. However, a two-account system does get very close to this efficiency bound: two flexible accounts generate a social-welfare gain equal to 6.136% of the endowment. (For the two account case, we find that one penalty is close to zero and the other is close to one.) The general mechanism generates a welfare gain that is bounded above by 6.144% of the endowment. A completely liquid and a completely illiquid account generate a welfare gain of 6.105% of the endowment. Finally, the three-account system (one completely liquid, one partially liquid and one completely illiquid) generates a welfare gain of 6.137% of the endowment. The (money-metric) differences among the mechanisms with more than a single account are small in economic magnitude. To summarize, a very simple two-account system—one perfectly liquid and one perfectly illiquid—generates approximately optimal welfare gains. Such a two-account system is commonplace in most countries in the developed world (Beshears, et al., 2015).

To gain intuition for this result, we report a related set of analyses in Figure 2. Here, we study a two-account system. One account is completely liquid (i.e., \( \pi_1 = 0 \)) and the other account has varying illiquidity (i.e., \( \pi_2 \) varies). As we vary the penalty \( \pi_2 \) from 0 to 1, we re-optimize the allocations \( x_1 \) and \( x_2 \) to the liquid and the partially illiquid accounts. The horizontal axis shows the penalty \( \pi_2 \), and the vertical axis shows the average welfare of the cross sections of the population obtained by fixing \( \beta \) at one the five values 0.2, 0.4, 0.6, 0.8 and 1.0. In this context it should be emphasized that all households are treated identically ex ante and, therefore, receive the same allocations and face the same early-withdrawal penalties. Also, the allocations are chosen to maximize the welfare of the population as a whole. The welfare of a given \( \beta \) cross section is then obtained by averaging over the welfare of the households with that particular value of \( \beta \).

For the \( \beta = 0.2 \) households, welfare as perceived by the planner rises dramatically as the early-withdrawal penalty increases (Figure 2). Indeed, as \( \pi_2 \) rises from 0 to 1, the increase in
Figure 2: The welfare of various $\beta$ cross sections of the population as a function of $\pi_2$ in the case in which: (i) the population has heterogeneous $\beta$; (ii) the planner is confined to a mechanism with two accounts, with penalties $\pi_1$ and $\pi_2$ respectively; (iii) $\pi_1 = 0$ (i.e., the first account is completely liquid); (iv) the account allocations are chosen to maximize the welfare of the population as a whole. Note that the cross section of the population with $\beta = 0.82$ (not shown) is almost indifferent between the system with $\pi_2 = 0$ and the system with $\pi_2 = 1$. 
welfare for the $\beta = 0.2$ households is equivalent, using a money metric, to a 30% increase of wealth (the money-metric welfare improvement is equivalent to half of the vertical distance on Figure 2). From the planner’s perspective, there is a substantial gain from discouraging these low-$\beta$ households from excessive consumption in period 1.

Households with other $\beta$ values experience increasing and then decreasing welfare as $\pi_2$ increases from 0 to 1. However, conditional on $\beta$, all households experience a rise in expected welfare as $\pi_2$ rises from zero. For low-$\beta$ households, this rise occurs because higher penalties prevent low-$\beta$ households from overconsuming in period 1. For high-$\beta$ households, this rise occurs because higher penalties generate larger cross-subsidies from low-$\beta$ households to high-$\beta$ households. Specifically, these cross-subsidies occur because higher penalty revenue relaxes the planner’s budget constraint, thereby enabling the planner to give agents higher endowments in period 1. High-$\beta$ households are net recipients of cross-subsidies because they tend to make smaller early withdrawals and, therefore, pay fewer penalties than low-$\beta$ households (look ahead to Figure 4).

Unlike the welfare of low-$\beta$ households, which rises monotonically as $\pi_2$ rises, the welfare of high-$\beta$ households eventually peaks and thereafter falls with $\pi_2$. This single-peaked property arises because, while initial rises in $\pi_2$ simply result in greater revenue from the early-withdrawal penalties paid by low-$\beta$ households, later rises tend to eliminate early withdrawals altogether. Hence the cross-subsidy to high-$\beta$ households first rises and then falls. By the time $\pi_2$ reaches 1, the cross-subsidy has been completely eliminated, and high-$\beta$ households are now facing a binding constraint (if they have a sufficiently high $\theta$ value) that limits their ability to adjust consumption in period 1, so high-$\beta$ households are slightly worse off on average than they were when $\pi_2$ was 0. On a money-metric basis, the $\beta = 1$ households experience a welfare loss equivalent to 0.23% of their income as the planner moves from $\pi_2 = 0$ to $\pi_2 = 1$ in Figure 2. However, this welfare loss is swamped by the welfare gain experienced by the $\beta = 0.2$ types (which is two orders of magnitude larger).

Figure 3 shows the welfare of the population as a whole as a function of the early-
withdrawal penalty $\pi_2$. It confirms that—as one would expect—the enormous welfare gains for low-$\beta$ households swamp the modest welfare losses for high-$\beta$ households, an example of asymmetric paternalism (Camerer, et al., 2003). Although it appears that total social welfare rises monotonically and asymptotes, social welfare actually reaches a global maximum at $\pi_2 = 0.85$ and then falls very slightly. However, the fall in welfare between $\pi_2 = 0.85$ and $\pi_2 = 1$ is insignificant: it is 0.00002% of wealth using a money metric. Accordingly, the social optimum is effectively obtained with one completely liquid account and one completely illiquid account.

Figure 4 reports the gross penalties paid by households with different values of $\beta$ (again integrating over $\theta$). As anticipated above, the penalties are hump shaped in $\pi_2$, with the hump occurring for larger values of $\pi_2$ the lower the value of $\beta$. For example, for $\beta = 1$, the hump peaks at $\pi_2 = 0.07$. For $\beta = 0.2$, the hump peaks at $\pi_2 = 0.51$. Intuitively, high-$\beta$ households stop making early withdrawals even when the penalty is low (because they value utils in period 2 almost as much as they value utils in period 1). By contrast, low-$\beta$ households do not stop making early withdrawals until the penalty reaches a relatively high level.

5.3 A three-account system that approximates the U.S. retirement savings system

The fifth row in Table 2 reports the welfare gains for a three-account system ($N = 3$). We will see that this analysis reproduces some of the features of the U.S. system.

We constrain the first account to be completely liquid ($\pi_1 = 0$) and the third account to be completely illiquid ($\pi_3 = 1$). Think of this third account—the illiquid account—as Social Security or a defined-benefit pension. The planner picks the penalty on the “middle” account ($0 < \pi_2 < 1$) and the values of the three endowments ($x_1$, $x_2$ and $x_3$) to optimize social welfare (while satisfying the budget constraint). The “middle” account turns out to have an optimal penalty of $\pi_2 = 0.13$, which is close to the actual penalty associated with a
Figure 3: The welfare of the population as a whole as a function of $\pi_2$ in the case in which: (i) the population has heterogeneous $\beta$; (ii) the planner is confined to a mechanism with two accounts, with penalties $\pi_1$ and $\pi_2$, respectively; (iii) $\pi_1 = 0$ (i.e., the first account is completely liquid); (iv) the account allocations are chosen to maximize the welfare of the population as a whole. Note that: (i) while this is not immediately apparent from the figure, the function in question is non-monotone; (ii) the optimal penalty $\pi_2^*$ is approximately 85%; (iii) $\pi_2^*$ yields a proportional increase of approximately 0.00002% in money-metric welfare relative to the case in which $\pi_2 = 1$ (i.e., the case in which the second account is completely illiquid). In particular, the welfare cost of setting the penalty on the second account too low far exceeds that of setting it too high.
Figure 4: The total penalties paid by various $\beta$ cross sections of the population as a function of $\pi_2$ in the case in which: (i) the population has heterogeneous $\beta$; (ii) the planner is confined to a mechanism with two accounts, with penalties $\pi_1$ and $\pi_2$, respectively; (iii) $\pi_1 = 0$ (i.e., the first account is completely liquid); (iv) the account allocations are chosen to maximize the welfare of the population as a whole.

401(k) or IRA account, namely 0.10.

Adding this optimized “middle” account to the constrained two-account system (row 4 in Table 2) only slightly raises consumer welfare relative to the two-account system with a completely liquid and a completely illiquid account. Comparing rows 4 and 5 of Table 2, we see that the addition of the middle account increases social welfare by $6.137\% - 6.105\% = 0.032\%$ of wealth (using a money metric).

Our simulations reveal that the middle account is characterized by a very high degree of leakage in equilibrium. Ninety percent of the assets in the middle account are withdrawn to fund consumption in period 1. Figure 5 disaggregates this result, by plotting the cumulative distribution function of the ratio $c_2/c_1$. Figure 5 shows that 76% of households choose the
 maximal value of $c_1$ (corresponding to a full withdrawal of the funds in the partially liquid account), generating a consumption ratio of $c_2/c_1 = 0.94$. Another 22% of households withdraw at least some (though not all) of the funds in the partially liquid account. Another 1% of households choose to save all of the funds in the partially liquid account for retirement (but no more than that). Finally, 1% of households choose to save some of the funds in the completely liquid account for retirement (in addition to all of the funds in the partially liquid account).

In summary, our analysis finds that welfare is nearly as high in the two-account system with a completely liquid account and a completely illiquid account as it is in the three-account system that adds a partially illiquid account. When a third account is added, it looks and performs somewhat like a U.S. 401(k) plan: the third account has an optimized penalty of 0.13 and generates a very high rate of leakage in equilibrium. This high leakage rate is even higher than the empirical leakage rate observed in the U.S. system.

One explanation for the difference between the model-predicted leakage rate (90%) and the empirically observed leakage rate (40%) is that initial account balances in the model are generated by government fiat, whereas almost all of the dollars in real-world 401(k)/IRA accounts are voluntarily deposited, implying that they are coming from households with higher $\beta$ values and lower $\theta$ values in the first place. In this sense, one can’t directly compare the leakage rate in the model (which is the aftermath of universal forced savings in a DC system) and the leakage rate in the US economy (which is the aftermath of voluntary savings in a DC system). Accordingly, differential selection makes this an apples to oranges comparison.

Another key factor that explains the high model-predicted leakage rate is the fact that the planner endogenously (optimally) chooses to put a large fraction of each household’s resources into the completely illiquid account, thereby reducing the marginal value of re-

\footnote{The third account offers the welfare benefit of additional separation for high-$\theta$ households and low-$\theta$ households. However, the third account has two effects that jointly offset the welfare gains from separation. First, the third account enables low-$\beta$ households to increase their period 1 over-consumption. Second, withdrawals from the third account generate (socially inefficient) transfers of resources from low-$\beta$ and high-$\theta$ households to high-$\beta$ and low-$\theta$ households because of the penalties that are paid for period 1 withdrawals from the third account. These tax revenues are redistributed in the mechanism.}
Figure 5: The distribution function of the ratio $c_2/c_1$ of period-2 consumption to period-1 consumption in the population as a whole in the case in which: (i) the population has heterogeneous $\beta$; (ii) the planner is confined to a mechanism with three accounts, with penalties $\pi_1$, $\pi_2$ and $\pi_3$, respectively; (iii) $\pi_1 = 0$ (i.e., the first account is completely liquid); (iv) $\pi_3 = 1$ (i.e., the third account is completely illiquid); (v) both $\pi_2$ and the account allocations are chosen to maximize the welfare of the population as a whole. There are two atoms in the distribution: a large atom accounting for about 76% of the total mass near $c_2/c_1 = 0.94$; and a small atom accounting for about 1% of the total mass near $c_2/c_1 = 1.70$. Individuals at the second atom have withdrawn the entire balance from the first (liquid) account, but have not yet touched the second account. Individuals at the first atom have withdrawn the entire balance from both the first and the second accounts. In particular, they have paid the penalty $\pi_2$ on the entire balance of the second account.
tirement consumption, and implicitly encouraging pre-retirement withdrawals of balances in the partially illiquid account. However, we find that leakage rates from the partially illiquid account are substantially lower under a system in which contributions to the completely illiquid account (e.g., Social Security) are exogenously set at a lower level than is recommended by our normative framework. Our mechanism-design framework generates nearly complete consumption smoothing between period 1 and period 2, even if the household relies only on the resources from the completely illiquid account to finance period-2 consumption. By implication, our framework recommends far more funding for the completely illiquid account than we actually observe in the United States. If we adopt an empirically more realistic (modest) funding rule for the completely illiquid account, then we observe much lower levels of leakage from the partially illiquid account because its balances are more urgently needed to support consumption in period 2 (i.e., retirement). We return to these issues in Subsection 6.1.

6 Optimal Policy with Transfers and Heterogeneous Present Bias: Robustness

In the previous section, which studied the case in which inter-household transfers are allowed and present bias is heterogeneous in the population, three key findings emerged:

1. The constrained-efficient social optimum is approximated by a two-account system, with one account that is completely liquid and a second account that is completely illiquid. Little welfare gain is obtained by moving beyond this simple two-account system.

2. If a third account is added, its optimized early-withdrawal penalty is 13%.

3. The equilibrium leakage rate from this third account is 90%.
In the current section, we document the robustness of these three findings when the distribution of \( \beta \) is heterogeneous.\(^{30}\) With respect to the first finding, the largest incremental welfare gain that we generate in our robustness checks by extending the system of savings accounts \textit{beyond} one completely liquid and one completely illiquid account is 0.081\% of income. Hence, we continue to find that a simple system with one completely liquid account and one completely illiquid account is approximately socially optimal.

With respect to the second finding, the optimized penalty on the partially illiquid account ranges from 11\% to 14\% across our calibrated economies. Hence, we continue to find that the partially illiquid account has a penalty that is similar to the penalties on 401(k)s and IRAs.

With respect to the third finding, the equilibrium leakage rate ranges from 84\% to 99\%. Hence, we continue to find that our simulations generate very high equilibrium leakage rates.

The detailed results are reported in the three panels of Table 3, which report the welfare gain (relative to the laissez faire case) for (i) the two-account system \( \pi_1 = 0 \) and \( \pi_2 = 1 \), (ii) the three-account system with \( \pi_1 = 0 \), \( 0 < \pi_2 < 1 \), and \( \pi_3 = 1 \), and (iii) the upper bound of the general mechanism (i.e., the ‘relaxed’ non-linear case described in Appendix B). For case (ii), in addition to the welfare gain, we also report the penalty \( \pi_2 \) and the leakage rate. Note that the upper bound on the welfare gain—case (iii)—is always tight in the sense that it is economically close to the constrained \( N \)-account cases that we study.

Table 3a varies the value of the coefficient of relative risk aversion (\( \gamma \)). In our benchmark calibration, we set \( \gamma = 1 \). In Table 3a, we study the cases \( \gamma = 1/2 \), \( \gamma = 1 \) (for comparison), and \( \gamma = 2 \).

Table 3b varies the shape of the density of \( \theta \). Recall that in our benchmark calibration, the density of the multiplicative taste shocks is truncated normal. Specifically, we start with a normal distribution, with mean \( \mu = 1 \) and standard deviation \( \sigma = 0.25 \); the resulting density is truncated (and reweighted) with support \([1 - \chi, 1 + \chi]\), with \( \chi = 2/3 \). In Table

---

\(^{30}\)When \( \beta \) takes on a degenerate distribution – i.e., \( \beta \) is homogeneous in the population – these results no longer apply (see the last column in Table 3c).
Table 3: Robustness checks for welfare gains, optimal penalties and leakage rates. In each subtable: row 1 contains welfare gains for a two-account system with one completely liquid account and one completely illiquid account; row 2 contains welfare gains for a three-account system with one completely liquid account, one account with an endogenous penalty and one completely illiquid account; rows 3 and 4 contain the optimal penalty and leakage rate from the endogenous-penalty account associated with the system in row 2; and row 5 contains welfare gains for the general mechanism (relaxed), which allows for an arbitrary non-linear budget set and does not impose the monotonicity restriction. Table 3a varies the value of the coefficient of relative risk aversion $\gamma$. Table 3b varies the parameter $\sigma_\theta$ of the truncated-normal distribution of $\theta$. Table 3c varies the parameter $\sigma_\beta$ of the truncated-normal distribution of $\beta$. 

<table>
<thead>
<tr>
<th>Value of $\gamma$</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Liquid, 1 Illiquid</td>
<td>8.851</td>
<td>6.105</td>
<td>3.261</td>
</tr>
<tr>
<td>1 Liquid, 1 Flexible, 1 Illiquid</td>
<td>8.919</td>
<td>6.137</td>
<td>3.274</td>
</tr>
<tr>
<td>Penalty $\pi^*_2$</td>
<td>0.13</td>
<td>0.13</td>
<td>0.11</td>
</tr>
<tr>
<td>Leakage Rate</td>
<td>0.89</td>
<td>0.90</td>
<td>0.99</td>
</tr>
<tr>
<td>General Mechanism (Relaxed)</td>
<td>8.932</td>
<td>6.144</td>
<td>3.278</td>
</tr>
</tbody>
</table>

(a) Variation of the coefficient of relative risk aversion $\gamma$

<table>
<thead>
<tr>
<th>Value of $\sigma_\theta$</th>
<th>0.30</th>
<th>0.25</th>
<th>0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Liquid, 1 Illiquid</td>
<td>5.918</td>
<td>6.105</td>
<td>6.323</td>
</tr>
<tr>
<td>1 Liquid, 1 Flexible, 1 Illiquid</td>
<td>5.958</td>
<td>6.137</td>
<td>6.344</td>
</tr>
<tr>
<td>Penalty $\pi^*_2$</td>
<td>0.14</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td>Leakage Rate</td>
<td>0.84</td>
<td>0.90</td>
<td>0.89</td>
</tr>
<tr>
<td>General Mechanism (Relaxed)</td>
<td>5.966</td>
<td>6.144</td>
<td>6.349</td>
</tr>
</tbody>
</table>

(b) Variation of the standard deviation $\sigma_\theta$ of the taste shock

<table>
<thead>
<tr>
<th>Value of $\sigma_\beta$</th>
<th>$+\infty$</th>
<th>1.0</th>
<th>0.5</th>
<th>0.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Liquid, 1 Illiquid</td>
<td>6.105</td>
<td>6.019</td>
<td>5.772</td>
<td>2.542</td>
</tr>
<tr>
<td>1 Liquid, 1 Flexible, 1 Illiquid</td>
<td>6.137</td>
<td>6.053</td>
<td>5.810</td>
<td>2.841</td>
</tr>
<tr>
<td>Penalty $\pi^*_2$</td>
<td>0.13</td>
<td>0.13</td>
<td>0.14</td>
<td>0.36</td>
</tr>
<tr>
<td>Leakage Rate</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.73</td>
</tr>
<tr>
<td>General Mechanism (Relaxed)</td>
<td>6.144</td>
<td>6.060</td>
<td>5.819</td>
<td>2.881</td>
</tr>
</tbody>
</table>

(c) Variation of the standard deviation $\sigma_\beta$ of the present bias distribution
3b, we study the cases $\sigma = 0.30$, $\sigma = 0.25$ (for comparison), and $\sigma = 0.20$.

Table 3c varies the standard deviation of the distribution of $\beta$ values (holding the mean fixed). In our benchmark calibration, we studied the case of a uniform distribution of $\beta$ between 0.2 and 1.0. In Table 3c, we study truncated normal distributions of $\beta$, with 0.2 and 1.0 serving as the truncation points. Our original benchmark is equivalent to the (truncated) normal case with $\sigma_\beta = \infty$ and $\mu_\beta = 0.6$. We now reduce $\sigma_\beta$ to 1, $1/2$ and 0 (holding the truncation points and $\mu_\beta$ fixed). The case $\sigma_\beta = 0$ is the degenerate case in which all agents have the same value of $\beta = 0.6$. Our results do not generalize to the degenerate case (the last column of Table 3c). As we showed in Section 5, the homogeneous-$\beta$ case engenders a Pigouvian tax as the approximately optimal policy. Gathering these results together, we infer that (at least partially unobservable) heterogeneity in $\beta$ is necessary for a fully illiquid account to be optimal.

6.1 How our normative framework deviates from the U.S. retirement savings system

Our model predicts that the planner should set up a completely illiquid account and populate it with enough assets so that, in equilibrium, there is at most a small drop in consumption at retirement. For example, under the optimal policy with a three-account system—one completely liquid, one partially illiquid (with an optimized 0.13 early-withdrawal penalty), and one completely illiquid, the respective account allocations are 36.4%, 16.2%, and 47.4% of lifetime resources.

Accordingly, the completely illiquid account alone (with 47.4% of total lifetime resources) is sufficient to generate nearly equal consumption in periods 1 and 2, even if the household consumes all of its completely liquid and partially illiquid assets in period 1. The high level of completely illiquid retirement assets explains the high level of equilibrium leakage from the partially illiquid account (in period 1). The partially illiquid account is a source of retirement consumption that can be used to supplement the consumption that will be generated by
the assets in the completely illiquid account. Because the mandatory, completely illiquid retirement assets is so large (at the social optimum), households are not strongly motivated to preserve the assets in the partially illiquid account until retirement. Accordingly, the equilibrium leakage rate from the partially illiquid account is 90.2%.\textsuperscript{31} Hence, very high rates of equilibrium leakage are consistent with optimized policy in an economy populated by agents with present-bias.

In the United States, the actual allocation to completely illiquid accounts is far lower than our optimized policy implies (e.g., mandatory savings is not sufficient to generate approximate consumption smoothing on its own in the United States). Relatedly, the fully liquid account plays a far more important role in practice than it does in our model. In addition, in the United States some withdrawals from retirement accounts are not penalized (e.g., education expenses, large unreimbursed health expenses, the purchase of a first home). To account for these factors, we report an illustrative calibration of the model where we exogenously fix the account balance allocations (rather than endogenously optimizing them) to reflect the operation of the status quo system in the United States. We exogenously allocate 60\% of lifetime resources to the liquid account, 10\% of lifetime resources to the partially illiquid retirement account with a 0.10 early-withdrawal penalty to match U.S. system, 10\% of lifetime resources to another partially illiquid retirement account with a 0.01 early-withdrawal ‘penalty’ to conceptually capture the fact that some retirement assets are accessible with only small logistical costs (i.e., non-penalized withdrawals), and 20\% of assets to the completely illiquid account. With this calibration, we obtain an aggregate leakage rate (total leakage divided by total balances in the two partially illiquid retirement accounts) of 31\%, which is within the range of historical leakage rates in the United States (see Argento, Bryant and Sabelhaus, 2014).

\textsuperscript{31}The high leakage rate implies that the partially illiquid account has very little impact on almost all households (relative to a world in which the funds from the partially illiquid account were instead put in the liquid account). This explains why the partially illiquid account has such a small effect on total social welfare relative to the two-account benchmark, with a completely liquid account and a completely illiquid account.
7 Conclusions and Directions for Future Work

To summarize, we focus on the case in which agents have heterogeneous present bias and the planner can implement mechanisms with inter-household transfers. Three findings emerge from our analysis:

1. The constrained-efficient social optimum is well-approximated by a two-account system, with one account that is completely liquid and a second account that is completely illiquid. Little welfare gain is obtained by moving beyond this simple two-account system. Accordingly, the two-account system identified in AWA (in a model with homogeneous $\beta$ and no inter-household transfers) turns out to be approximately optimal in our new setting (with heterogeneous $\beta$ and inter-household transfers).

2. If a third account is added, its optimized early-withdrawal penalty is only slightly above 10%.

3. In equilibrium, the leakage rate from this (partially illiquid) third account is high. We report a range of equilibrium leakage rates, depending on the calibration. With optimal allocations to all three accounts—completely liquid, partially illiquid, and completely illiquid—equilibrium leakage rates from the partially illiquid account range from 73% to 99%. By contrast, when we calibrate the model to match actual empirical allocations to the completely illiquid account (e.g., treating Social Security as the empirical analog of the model’s completely illiquid account), the implied equilibrium leakage rate from the partially illiquid account drops to 46%.

These properties have analogs in the U.S. retirement savings system. The United States has completely liquid accounts (e.g., a standard checking account), completely illiquid accounts (e.g., Social Security), and a partially illiquid defined-contribution system with a 10% penalty for early withdrawals (e.g., an IRA or a 401(k)). This partially illiquid DC system has a leakage rate of approximately 40% (see Argento, Bryant and Sabelhaus, 2014).
Despite these similarities, it is inappropriate to conclude that our findings demonstrate the social optimality of the U.S. system. Most importantly, our theoretical model includes several key simplifications. First, we assume a particular conceptual formulation of self-defeating behavior (present bias). Second, we assume only two periods (e.g., working life and retirement). Third, we assume a particular form of multiplicative taste shifter, $\theta$. Fourth, we assume that households are naive with respect to their present bias parameter, $\beta$. Fifth, we study a limited set of distributions of $\theta$ and $\beta$ (and no correlation).

Moreover, our simulations imply that retirement consumption should not be allowed to fall far below working life consumption (recall that the illiquid account has a high funding level when we calculate the socially optimal system). In the actual data on U.S. households, consumption proxies appear to decline between working life and retirement, raising the normative possibility that mandatory savings might be underutilized in the U.S. However, there is an active debate about both the existence and normative interpretation of the observed distribution of consumption changes for households transitioning from work life into retirement.

Our normative result on underutilization of mandatory savings is closely related to a similar result reported in Moser and Olea de Souza e Silva (2017). Though many elements of the two models differ, both models assume that agents are present biased, and they both imply that optimal savings mechanisms are characterized by more mandatory savings than

---

32 In addition, the U.S. system contains some scope for tax arbitrage, which is not present in our model.


34 Infinite horizon problems introduce technical challenges with respect to multiple equilibria. However, there has been progress on this issue. For example, see Harris and Laibson (2012) and Cao and Werning (2018).

35 We assume $\theta u(c)$, but we could have instead assumed $u(c - \theta)$.

36 Research is only beginning on the distribution of present bias. For analysis of this issue, see Lockwood (2016), Moser and Olea de Souza e Silva (2017), and Cohen, et al., (2019).


38 In our model, mandatory savings are achieved through a funded system. Our model takes no position on the distinction between funded (e.g., the superannuation scheme in Australia) and unfunded (e.g., U.S. Social Security) mandatory savings systems.

39 See Beshears, et al., (2018) for a recent review of the literature on consumption dynamics at and through retirement.
currently exists in the U.S. system.

Much more robustness work is needed to interrogate the three findings that we summarized above, as well as the additional finding that more mandatory savings would be socially optimal. It is not yet clear whether these results will continue to hold as future research enriches and expands this type of analysis.
8 References


Online Appendices

A Proof of Theorem 1

A.1 Formulation of the Theorem

In the main text we assumed that the income $Y$ of a household was 1 and that the total mass $F(\bar{y})$ of households was 1. This was done in order to reduce notation. In this appendix we will work with general $Y$ and general $F(\bar{y})$, since it is easier to follow the derivations in the general case.

The first step in formulating Theorem 1 is then to define $\theta_1$ in this more general setting. Recalling that the function $\Gamma$ is given by the formula

$$\Gamma(\theta) = (1 - \beta) \theta F'(\theta) + F(\theta),$$

we define $\theta_1$ to be the minimum $\theta \in (0, \bar{y})$ such that

$$\frac{1}{\bar{y} - t} \int_t^{\bar{y}} \Gamma(s) ds \geq F(\bar{y})$$

for all $t \in [\theta, \bar{y})$. Assumptions A1-A5 are then assumed to hold exactly as stated in the main text. Finally, we restate Theorem 1 for the reader’s convenience.

**Theorem 1** (Cf. Proposition 3 of AWA (2006).) Suppose that $\beta$ is the same for all households. Suppose further that inter-household transfers are not possible. Then welfare is maximized by dividing the endowment between two accounts: a completely liquid account (that can be used in both period 1 and period 2) and a completely illiquid account (that can be used only in period 2). In particular types in the separating interval $\Theta_S$ – which consists of those $\theta \in \Theta$ such that $\theta < \theta_1$, and which will be empty if $\theta_1 \leq \underline{\theta}$ – choose $c_1$ strictly less than the balance of the liquid account; and types in the pooling interval $\Theta_P$ – which consists of those
\( \theta \in \Theta \) such that \( \theta \geq \theta_1 \), and which is never empty – set \( c_1 \) equal to the balance of the liquid account (and therefore set \( c_2 \) equal to the balance of the completely illiquid account).

Our theorem generalizes AWA’s analysis in two respects. First, AWA’s analysis covers the case \( \theta_1 > \theta \), whereas our analysis holds for all values of \( \theta_1 \). Second, AWA’s analysis shows that the two-account system is optimal in the class of continuous incentive-compatible consumption allocations, whereas our analysis shows that the two-account system is optimal in the class of all incentive-compatible consumption allocations. The first point could be expressed by saying that AWA’s analysis covers the partial-separation case, whereas our analysis covers both the pooling and the partial-separation case. The second point is important, because many incentive-compatible consumption allocations – including some of the simplest possible incentive-compatible consumption allocations – are discontinuous.

### A.2 A Candidate Utility Allocation

Our strategy of proof is to construct a candidate utility allocation and a candidate Lagrange multiplier, and then show that the utility allocation maximises the Lagrangian when violations of the resource constraint are penalized using the Lagrange multiplier.

We begin by constructing a candidate consumption allocation. This is obtained by requiring that: (i) all types \( \theta \) in the separating interval \( \Theta_S = \{ \theta \mid \theta \in \Theta, \ \theta < \theta_1 \} = [\theta, \theta_1) \) choose freely from the unconstrained budget line, namely the set of all \( (c_1, c_2) \) such that \( c_1 \geq 0 \), \( c_2 \geq 0 \) and \( c_1 + c_2 = Y \); and (ii) all types \( \theta \) in the pooling interval \( \Theta_P = \{ \theta \mid \theta \in \Theta, \ \theta \geq \theta_1 \} = [\max\{ \theta, \theta_1 \}, \infty) \) receive the allocation that the (possibly hypothetical) type \( \theta_1 \) would choose freely from the unconstrained budget line.\(^{40}\)

We transform the candidate consumption allocation \( (c_1, c_2) : \Theta \to (0, \infty) \) into a candidate utility allocation \( (r_1, r_2) : \Theta \to \mathbb{R} \) by setting \( r_1(\theta) = u_1(c_1(\theta)) \) and \( r_2(\theta) = u_2(c_2(\theta)) \). We

\(^{40}\)The consumption allocation will be interior if and only if

\[ \frac{u'(Y)}{u'(0+)} < \frac{\min\{\theta, \theta_1\}}{\beta} < \frac{\theta_1}{\beta} < \frac{u'_2(0+)}{u'_1(Y)}. \]

Assumption A2 obviously implies this condition.
would like to show that the utility allocation \((r_1, r_2)\) is optimal among all economically meaningful utility allocations \((v_1, v_2)\).

This sets up a mathematical hurdle. For, while \((r_1, r_2)\) itself is fairly regular (it is a continuously differentiable function of \(\theta\) with a kink at \(\theta_1\)), the alternative utility allocations \((v_1, v_2)\) may not even be continuous. We will get over this hurdle by using the one regularity property that incentive-compatible utility allocations do possess: they are monotonic. Hence they are functions of bounded variation.

A.3 Functions of Bounded Variation on \(\Theta\)

There are a number of competing definitions of a function of bounded variation. According to one elementary definition, a function \(f : \Theta \rightarrow \mathbb{R}\) is of bounded variation iff it is the difference of two bounded and non-decreasing functions \(f_+, f_- : \Theta \rightarrow \mathbb{R}\). The most serious drawback with this definition for our purposes is that the functions defined in this way do not form a function space. This definition cannot therefore be used in a Lagrangian analysis. A second drawback of the definition is that it does not capture the behaviour of a function of bounded variation at the endpoints of \(\Theta\). We shall therefore adopt a definition that leads directly to a usable function space, and which ties down the behaviour of a function at the endpoints of \(\Theta\).

The intuitive idea is to say that \(f\) is a function of bounded variation on \(\Theta\) iff it is the distribution function of a bounded Borel measure on \(\Theta\) plus a constant of integration. More precisely, we begin from a constant of integration, denoted suggestively by \(f_L(\emptyset)\), and a bounded Borel measure on \(\Theta\), denoted suggestively by \(f'\). We then define the left-hand limits \(f_L\) of \(f\) by

\[ f_L(\theta) = f_L(\emptyset) + f'([\theta, \theta]) \]

for all \(\theta \in \Theta\) (including \(\emptyset\)) and the right-hand limits \(f_R\) of \(f\) by

\[ f_R(\theta) = f_L(\emptyset) + f'([\theta, \emptyset]) \]

51
for all \( \theta \in \Theta \) (including \( \theta \)). And we endow the set of functions obtained in this way with the norm

\[
\|f\|_{BV} = |f_L(\theta)| + \|f'\|_{TV},
\]

where \( \|\cdot\|_{TV} \) is the total-variation norm on bounded Borel measures on \( \Theta \).

This definition has at least three advantages: it is concrete; it builds on familiar ideas like distribution functions and the total-variation norm; and it brings out the subtleties implicit in the concept of a function of bounded variation. One subtlety is the fact that a “function” of bounded variation is not a function in the narrow sense of that word: it is only well defined where \( f_L = f_R \), and there may be a countable set of points at which this is not the case. (These points are precisely the atoms of the bounded Borel measure \( f' \). As such, they may include the endpoints \( \theta \) and \( \theta' \).) A second subtlety is the fact that a function of bounded variation has limits from both the left and right at all points of \( \Theta \), including a limit from the left at \( \theta \) and a limit from the right at \( \theta' \). (This makes perfect sense if one views functions of bounded variation on \( \Theta \) as restrictions to \( \Theta \) of functions of bounded variation on \( (0, \infty) \).) In view of these subtleties, one cannot simply adopt a convention that all functions of bounded variation are (say) right continuous.

### A.4 Functions of Bounded Variation on \((0, \infty)\)

The discussion of the previous section applies \textit{mutatis mutandis} to functions of bounded variation on \((0, \infty)\). The main differences are that: (i) we do not need to consider behaviour at the endpoints of the interval \((0, \infty)\); and (ii) it is preferable to specify the constant of integration at an interior point. Rather than work through this material in general, we shall simply discuss the special case of \( F' \).

We note first that – according to Assumption A2 – the support of \( F' \) (as a function) is contained in \( \Theta \). It follows, first, that \( F'_L(\theta) = F'_R(\theta) = 0 \). It follows, second, that the support of \( F'' \) (as a measure) is contained in \( \Theta \). In other words, \(|F''|(0, \theta)) = |F''|(\theta, \infty)) =
Now, because $|F''|((0, \theta)) = 0$, we can suppress the constant of integration in the formulae for $F'$ in terms of $F''$. More explicitly, we have

$$F_L'(\theta) = F''((0, \theta)),$$

$$F_R'(\theta) = F''((0, \theta))$$

for all $\theta > 0$. It then follows that

$$0 = F_R'(\theta) = F''((0, \theta]) = F''((0, \theta]) + F''([\theta, \theta])$$

$$= F_L'(\theta) + F''([\theta, \theta]) = F''([\theta, \theta]).$$

In other words, $F''$ assigns total mass 0 to $\Theta$.

### A.5 The Lagrangian

Denote by $\mathcal{BV}(\Theta, \mathbb{R})$ the Banach space of functions of bounded variation on $\Theta$ with the norm $\|\cdot\|_{\mathcal{BV}}$, and by

$$\mathcal{O}_t = \mathcal{BV}(\Theta, (u_t(0+), u_t(\infty-)))$$

the subset of $\mathcal{BV}(\Theta, \mathbb{R})$ consisting of functions taking values in the interior of the range of $u_t$. (Recall that $u_t$ is the utility function for date $t$.) Denote by $\Omega$ the set of utility allocations

$$v = (v_1, v_2) \in \mathcal{O}_1 \times \mathcal{O}_2$$

such that

$$\theta v'_1 + \beta v'_2 = 0$$

(PLIC)

and

$$v'_1 \geq 0.$$  

(ICM)
(The idea here is to split the incentive-compatibility condition into the linear part $IC_L$ and the monotonic part $IC_M$.) In other words, let $\Omega$ be the set of incentive-compatible utility allocations. Define the objective function

$$M : BV(\Theta, \mathbb{R})^2 \to \mathbb{R}$$

by the formula

$$M(v) = \int (\theta v_1 + v_2) F' \ell(d\theta),$$

where $\ell$ is Lebesgue measure, and define the budget operator

$$N : O_1 \times O_2 \to BV(\Theta, \mathbb{R})$$

by the formula

$$(N(v))(\theta) = Y - C_1(v_1(\theta)) - C_2(v_2(\theta)).$$

Then the planner’s problem is to maximize $M$ over the the set of all utility allocations $v \in \Omega$ such that $N(v) \geq 0$.

**Remark 5** We use the notation $\ell(d\theta)$ rather than $d\theta$ in the formula for $M$ in order to be consistent with the notation for integration elsewhere in this appendix.

Since $N$ takes values in $BV(\Theta, \mathbb{R})$, a Lagrange multiplier is a continuous linear functional on $BV(\Theta, \mathbb{R})$. Denote the space of all continuous linear functionals on $BV(\Theta, \mathbb{R})$ by $BV(\Theta, \mathbb{R})^*$. Then the Lagrangian is the mapping

$$L : \Omega \times BV(\Theta, \mathbb{R})^* \to \mathbb{R}$$

given by the formula

$$L(v; \lambda) = M(v) + \langle N(v), \lambda \rangle.$$
where \( \langle N(v), \lambda \rangle \) denotes the real number obtained when the continuous linear functional \( \lambda \in BV(\Theta, \mathbb{R})^* \) is evaluated at the point \( N(v) \in BV(\Theta, \mathbb{R}) \).

**Remark 6** Notice that both \( M \) and \( N \) are defined on open sets containing \( \Omega \), and not just on \( \Omega \) itself.

**Remark 7** \( M \) is well defined since \( v_1 \) and \( v_2 \) are well defined except at a countable number of points.

**Remark 8** \( N \) is well defined since \( u_t(0+) < \min v_t \leq \max v_t < u_t(\infty-) \) and \( C_t \) is continuously differentiable on \( (u_t(0+), u_t(\infty-)) \). Hence

\[
\| (C_t \circ v_t)' \|_{TV} \leq K \| v_t' \|_{TV},
\]

where

\[
K = \max \{ C'_t(w) \mid w \in [\min v_t, \max v_t] \}.\]

**Remark 9** According the the Riesz representation theorem, the dual \( C(\Theta, \mathbb{R})^* \) of the space \( C(\Theta, \mathbb{R}) \) of continuous functions on \( \Theta \) can be represented by the space \( M(\Theta, \mathbb{R}) \) of bounded Borel measures on \( \Theta \). Unfortunately, there does not seem to be a correspondingly tractable representation for the dual \( BV(\Theta, \mathbb{R})^* \) of the space \( BV(\Theta, \mathbb{R}) \) of functions of bounded variation on \( \Theta \). This might be an obstacle to analyzing necessary conditions, where we would not have any control over the Lagrange multiplier. It is less of a problem when it comes to analyzing sufficiency conditions, where we are free to choose the Lagrange multiplier.

### A.6 A Space of Lagrange Multipliers

One can associate continuous linear functionals in \( BV(\Theta, \mathbb{R})^* \) with bounded Borel measures in \( M(\Theta, \mathbb{R}) \) as follows. Suppose that we are given \( \Lambda \in M(\Theta, \mathbb{R}) \). Then we can construct
\[ \lambda_R \in \mathcal{BV}(\Theta, \mathbb{R})^* \] by means of the formula

\[ \langle f, \lambda_R \rangle = \int f_R(\theta) \Lambda(d\theta), \]

where \( f_R \) is the right-continuous version of \( f \). In this way we obtain a closed linear subspace of \( \mathcal{BV}(\Theta, \mathbb{R})^* \). It turns out that this subspace is big enough for our purposes.

**Remark 10** By the same token, we can construct \( \lambda_L \in \mathcal{BV}(\Theta, \mathbb{R})^* \) by means of the formula

\[ \langle f, \lambda_L \rangle = \int f_L(\theta) \Lambda(d\theta), \]

where \( f_L \) is the left-continuous version of \( f \).

**Remark 11** Notice that \( \lambda_L \neq \lambda_R \) and, while both \( \lambda_L \) and \( \lambda_R \) seem quite natural, neither seems to have a claim to being canonical.

**Remark 12** We use the notation \( \Lambda(d\theta) \) in the definition of \( \langle f, \lambda_R \rangle \) and \( \langle f, \lambda_L \rangle \) in order to emphasize that the integral in question is the Lebesgue integral of a measurable function with respect to the measure \( \Lambda \). (The notation \( d\Lambda(\theta) \) might be taken to suggest that the integral in question was the Riemann-Stieltjes integral of a continuous function with respect to the function of bounded variation \( \Lambda \).

### A.7 The Directional Derivative of the Lagrangian

Let us fix \( \Lambda \in \mathcal{M}(\Theta, \mathbb{R}) \) and consider \( L(\cdot; \lambda_R) \). If our candidate allocation \( r \in \Omega \) maximizes \( L(\cdot; \lambda_R) \) then, for all \( v \in \Omega \), the directional derivative \( \nabla_s L(r; \lambda_R) \) of \( L(\cdot; \lambda_R) \) at \( r \) in the direction \( s = v - r \) must be non-positive. Conversely if, for all \( v \in \Omega \), \( \nabla_s L(r; \lambda_R) \) is non-positive, then \( r \in \Omega \) maximizes \( L(\cdot; \lambda_R) \). The purpose of the present section is to derive a formula for \( \nabla_s L(r; \lambda_R) \). This formula will then be used to guide our eventual choice of \( \Lambda \).
In view of our choice of $\lambda_R$, we have

$$L(v; \lambda_R) = \int (\theta v_1 + v_2) F'(d\theta) + \int (Y - C_1(v_1) - C_2(v_2)) \Lambda(d\theta).$$

Hence

$$\nabla_s L(r; \lambda_R) = \int (\theta s_1 + s_2) F'(d\theta) - \int (C'_1(r_1) s_1 + C'_2(r_2) s_2) \Lambda(d\theta).$$

Now, because $F$ is continuous, the standard formula for integration by parts shows that

$$\int s_2 F' \ell(d\theta) = [s_2 F]_{\theta^+}^{\theta^-} - \int F s'_2(d\theta),$$

where:

- $[s_2 F]_{\theta^+}^{\theta^-}$ denotes the difference between the right-hand limit of $s_2 F$ at $\theta$ and the left-hand limit of $s_2 F$ at $\theta$;

- $\int F s'_2(d\theta)$ denotes the integral of $F$ with respect to the measure $s'_2$.

Furthermore, it follows from incentive compatibility that $\theta s'_1 + \beta s'_2 = 0$. Hence

$$\int F s'_2(d\theta) = -\int F \frac{\theta}{\beta} s'_1(d\theta) = -\frac{1}{\beta} [s_1(\theta F)]_{\theta^+}^{\theta^-} + \frac{1}{\beta} \int s_1(\theta F)' \ell(d\theta),$$

(integrating by parts again, and using the fact that $F$ is continuous). Hence the first integral in the directional derivative

$$\int (\theta s_1 + s_2) F'(d\theta) = \int \theta s_1 F'(d\theta) + \int s_2 F'(d\theta)$$

$$= \int \theta s_1 F'(d\theta) + [s_2 F]_{\theta^+}^{\theta^-} + \frac{1}{\beta} [s_1(\theta F)]_{\theta^+}^{\theta^-} - \frac{1}{\beta} \int s_1(\theta F)' \ell(d\theta)$$

$$= (\theta s_1(\theta) + s_2(\theta)) F(\theta) - \frac{1}{\beta} \int s_1((1 - \beta) \theta F' + F) \ell(d\theta)$$

(where we have used the fact that $F(\theta) = 0$).
Next, $G$ be the distribution function of the measure $C'_2(r_{2R}) \Lambda$. I.e. let $G$ be the unique element of $\mathcal{B} \mathcal{V}(\Theta, \mathbb{R})$ such that $G' = C'_2(r_{2R}) \Lambda$ and $G'_L(\theta) = 0$. Then

$$
\int C'_2(r_{2R}) s_{2R} \Lambda(d\theta) = \int s_{2R} G'(d\theta)
$$

$$
= \left[ s_2 G \right]_{\theta=0}^{\theta=\infty} - \int G' s_2(d\theta) + \sum_{\theta \in [\underline{\theta}, \overline{\theta}]} \Delta s_2 \Delta G,
$$

where $\Delta s_2$ and $\Delta G$ denote the jumps in $s_2$ and $G$ at $\theta$ (if any). Furthermore, it follows from incentive compatibility that $\theta' s'_1 + \beta s'_2 = 0$. In particular, $\theta \Delta s_1 + \beta \Delta s_2 = 0$. Hence

$$
\int G' s'_2(d\theta) = - \int G' s'_1(d\theta)
$$

$$
= -\frac{1}{\beta} \left[ s_1 (\theta G) \right]_{\theta=0}^{\theta=\infty} + \frac{1}{\beta} \int s_{1R} (\theta G)'(d\theta) - \frac{1}{\beta} \sum_{\theta \in [\underline{\theta}, \overline{\theta}]} \Delta s_1 \Delta (\theta G)
$$

$$
= -\frac{1}{\beta} \left[ s_1 (\theta G) \right]_{\theta=0}^{\theta=\infty} + \frac{1}{\beta} \int s_{1R} (\theta G)'(d\theta) - \frac{1}{\beta} \sum_{\theta \in [\underline{\theta}, \overline{\theta}]} \Delta s_1 \theta \Delta G
$$

(integrating by parts again and using the fact that $\Delta (\theta G) = \theta \Delta (G)$), and

$$
\sum_{\theta \in [\underline{\theta}, \overline{\theta}]} \Delta s_2 \Delta G = -\frac{1}{\beta} \sum_{\theta \in [\underline{\theta}, \overline{\theta}]} \theta \Delta s_1 \Delta G.
$$

Overall,

$$
\int C'_1(r_{1R}) s_{1R} \Lambda(d\theta) = \int \frac{C'_1(r_{1R})}{C'_2(r_{2R})} s_{1R} G'(d\theta)
$$

and

$$
\int C'_2(r_{2R}) s_{2R} \Lambda(d\theta) = \left[ s_2 G \right]_{\theta=0}^{\theta=\infty} + \frac{1}{\beta} \left[ s_1 (\theta G) \right]_{\theta=0}^{\theta=\infty} - \frac{1}{\beta} \int s_{1R} (\theta G)'(d\theta)
$$

$$
= s_{2R} (\overline{\theta}) G_R(\overline{\theta}) + \frac{1}{\beta} s_{1R}(\overline{\theta}) \overline{\theta} G_R(\overline{\theta}) - \frac{1}{\beta} \int s_{1R} (\theta G'(d\theta) + G \ell(d\theta))
$$

$$
= \left( \frac{\overline{\theta}}{\beta} s_{1R}(\overline{\theta}) + s_{2R}(\overline{\theta}) \right) G_R(\overline{\theta}) - \frac{1}{\beta} \int s_{1R} (\theta G'(d\theta) + G \ell(d\theta))
$$

(where we have used the facts that $G'_L(\theta) = 0$ and $(\theta G)'(d\theta) = \theta G'(d\theta) + G \ell(d\theta)$).
Finally, putting all of this information together, we have

$$
\nabla \mathcal{L}(r; \lambda_R) = \left( \frac{\pi}{3} s_{1R}(\overline{\theta}) + s_{2R}(\overline{\theta}) \right) F(\overline{\theta}) - \frac{1}{3} \int s_{1R} ((1 - \beta) \theta F' + G) \ell(d\theta)
$$

$$
- \int \frac{C'_1(r_{1R})}{C'_2(r_{2R})} s_{1R} G'(d\theta)
$$

$$
- \left( \frac{\pi}{3} s_{1R}(\overline{\theta}) + s_{2R}(\overline{\theta}) \right) G_R(\overline{\theta}) + \frac{1}{3} \int s_{1R} (\theta G'(d\theta) + G \ell(d\theta))
$$

$$
= \left( \frac{\pi}{3} s_{1R}(\overline{\theta}) + s_{2R}(\overline{\theta}) \right) (F(\overline{\theta}) - G_R(\overline{\theta}))
$$

$$
+ \frac{1}{3} \int s_{1R} (G - (1 - \beta) \theta F' - F) \ell(d\theta)
$$

$$
+ \frac{1}{3} \int s_{1R} \left( \theta - \beta \frac{C'_1(r_{1R})}{C'_2(r_{2R})} \right) G'(d\theta).
$$

A.8 A Candidate Lagrange Multiplier

We are now in a position to motivate our choice of Lagrange multiplier \( \Lambda \). We shall do this in two steps. First, we motivate our choice of \( \Gamma \). Second, we show how to translate our choice of \( G \) into a choice of \( \Lambda \).

In choosing \( G \), the broad aim is to ensure that \( \nabla \mathcal{L}(r; \lambda_R) \leq 0 \). However, given that we have only limited control over \( s \), it will be helpful to make as many of the terms in the formula for \( \nabla \mathcal{L}(r; \lambda_R) \) vanish as possible.

Recall that

$$
\nabla \mathcal{L}(r; \lambda_R) = \left( \frac{\pi}{3} s_{1R}(\overline{\theta}) + s_{2R}(\overline{\theta}) \right) (F(\overline{\theta}) - G_R(\overline{\theta}))
$$

$$
+ \frac{1}{3} \int s_{1R} (G - \Gamma) \ell(d\theta)
$$

$$
+ \frac{1}{3} \int s_{1R} \left( \theta - \beta \frac{C'_1(r_{1R})}{C'_2(r_{2R})} \right) G'(d\theta),
$$

where \( \Gamma = (1 - \beta) \theta F' - F \). We can therefore make a start by requiring that

$$
G_R(\overline{\theta}) = F(\overline{\theta}).
$$
This will ensure that the first term vanishes.

Remark 13 At this point we have specified both $G_L(\theta)$ and $G_R(\bar{\theta})$. It remains to specify $G$ in the interior of $\Theta$.

Next, suppose that $\theta_1 > \theta_2$. Then

$$\frac{C_1'(r_{1R})}{C_2'(r_{2R})} = \begin{cases} \frac{\theta_1}{\theta_2} & \text{for } \theta \in \Theta_S = [\theta_1, \theta_2) \\ \frac{\theta_2}{\theta_1} & \text{for } \theta \in \Theta_P = [\theta_1, \theta_2] \end{cases}.$$ 

Hence the expression for $\nabla_s L(r; \lambda_R)$ simplifies to

$$\frac{1}{\beta} \int_{\Theta_S \cup \Theta_P} s_{1R} (G - \Gamma) \ell(d\theta) + \frac{1}{\beta} \int_{\Theta_P} s_{1R} (\theta - \theta_1) G'(d\theta).$$

Suppose further that we follow the suggestion of AWA (2006), and put $G = \Gamma$ on $\Theta_S$, where $\Gamma$ is the function defined in Section A.1 above. Then the contribution to $\nabla_s L(r; \lambda_R)$ from the separating interval $\Theta_S$ vanishes altogether, and all that is left is the contribution

$$\frac{1}{\beta} \int_{\Theta_P} s_{1R} (G - \Gamma) \ell(d\theta) + \frac{1}{\beta} \int_{\Theta_P} s_{1R} (\theta - \theta_1) G'(d\theta)$$

to $\nabla_s L(r; \lambda_R)$ from the pooling interval $\Theta_P$. Suppose finally that we follow the suggestion of AWA (2006), and put $G = F(\bar{\theta})$ on $(\theta_1, \bar{\theta})$. Then the measure $G'$ will have an atom of size $F'(\bar{\theta}) - \Gamma_L(\theta_1)$ at $\theta_1$, and it will vanish on $(\theta_1, \bar{\theta}]$. Since the term $\theta - \theta_1$ multiplying $G'(d\theta)$ vanishes at $\theta_1$, the second integral itself vanishes, and the first integral reduces to

$$\frac{1}{\beta} \int_{\Theta_P} s_{1R} (F'(\bar{\theta}) - \Gamma) \ell(d\theta).$$

Next, suppose that $\theta_1 \leq \theta_2$. In this case, the expression for $\nabla_s L(r; \lambda_R)$ simplifies to

$$\frac{1}{\beta} \int_{\Theta_P} s_{1R} (G - \Gamma) \ell(d\theta) + \frac{1}{\beta} \int_{\Theta_P} s_{1R} (\theta - \theta_1) G'(d\theta).$$
Suppose further that we follow the suggestion of AWA (2006), and put $G = F(\bar{\theta})$ on the whole of $(\underline{\theta}, \bar{\theta})$. Then the measure $G'$ will have an atom of size $F(\bar{\theta})$ at $\underline{\theta}$, and it will vanish on $(\underline{\theta}, \bar{\theta})$. Hence the expression for $\nabla_s L(r; \lambda_R)$ becomes

$$\frac{1}{\beta} \int s_{1R} \left( F(\bar{\theta}) - \Gamma \right) \ell(d\theta) + \frac{1}{\beta} s_{1R}(\underline{\theta})(\underline{\theta} - \theta_1) F(\bar{\theta}).$$

In other words, compared with the case $\theta_1 > \underline{\theta}$, there is an extra term arising from the atom of $G'$ at $\underline{\theta}$.

Finally, we obtain the Lagrange multiplier $\Lambda$ itself from the formula

$$\Lambda = \frac{1}{C_2'(r_{2R})} G'.$$

### A.9 Non-Negativity of the Lagrange Multiplier

Since $C_2''(r_{2R}) > 0$, $\Lambda \geq 0$ iff $G' \geq 0$. We will show that $G' \geq 0$. Suppose first that $\theta_1 > \underline{\theta}$. Then we have

$$G_L(\underline{\theta}) = 0$$

$$G = \Gamma \text{ on } (\underline{\theta}, \theta_1)$$

$$G = F(\bar{\theta}) \text{ on } (\theta_1, \bar{\theta})$$

$$G_R(\bar{\theta}) = F(\bar{\theta})$$

Now, it follows from the formula for $\Gamma$ that

$$G_R(\underline{\theta}) = \Gamma_R(\underline{\theta}) = (1 - \beta) \theta F'_R(\underline{\theta}) + F(\underline{\theta}) = (1 - \beta) \theta F'_R(\underline{\theta}) \geq 0.$$

And $G_L(\underline{\theta}) = 0$ by construction. Hence

$$\Delta G(\underline{\theta}) = G_R(\underline{\theta}) - G_L(\underline{\theta}) \geq 0.$$
Next, it follows from Assumption A4 that $\Gamma$ is non-decreasing on $(\underline{\theta}, \theta_1)$. Hence $G' = \Gamma' \geq 0$ there. Third, we have

$$\Delta G(\theta_1) = F(\overline{\theta}) - \Gamma_L(\theta_1).$$

But if it were the case that $\Gamma_L(\theta_1) > F(\overline{\theta})$ then there would be an open interval $(\theta_1 - \varepsilon, \theta_1)$ on which $\Gamma > F(\overline{\theta})$. This would contradict the definition of $\theta_1$ as the minimum $\theta \in (0, \overline{\theta})$ such that $\frac{1}{\theta - t} \int_t^\theta \Gamma(s) ds \geq F(\overline{\theta})$ for all $t \in [\theta, \overline{\theta})$. Hence $\Gamma_L(\theta_1) \leq F(\overline{\theta})$ and $\Delta G(\theta_1) \geq 0$.

Fourth, we have $G' = 0$ on $(\theta_1, \overline{\theta})$. Finally, we obviously have $\Delta G(\overline{\theta}) = 0$.

Suppose now that $\theta_1 \leq \underline{\theta}$. Then we have

$$G_L(\underline{\theta}) = 0$$

$$G = F(\overline{\theta}) \text{ on } (\underline{\theta}, \overline{\theta})$$

$$G_R(\overline{\theta}) = F(\overline{\theta}).$$

So it is obvious that $G' \geq 0$ on the whole of $[\underline{\theta}, \overline{\theta}]$.

A.10 Non-Positivity of the Directional Derivative

Suppose that $\theta_1 > \underline{\theta}$. Then, in the light of the discussion in Section A.8, we have

$$\nabla_s L(r; \lambda_R) = \frac{1}{\theta - \tau} \int_{\Theta_p} s_{1R} \left( F(\overline{\theta}) - \Gamma \right) \ell(d\theta).$$

Define $H : (0, \infty) \to \mathbb{R}$ by the formula

$$H(\theta) = \int_{\theta}^{\overline{\theta}} (\Gamma - F(\overline{\theta})) \ell(d\theta).$$
Then

\[
\int_{\Theta_P} s_{1R} \left( F(\bar{\theta}) - \Gamma \right) \ell(d\theta) = \int_{[\theta_1, \bar{\theta}]} s_{1R} \left( F(\bar{\theta}) - \Gamma \right) \ell(d\theta) = \int_{[\theta_1, \bar{\theta}]} s_{1R} H' \ell(d\theta) = \left[ s_{1H} \bar{\theta}^+ - \int_{[\theta_1, \bar{\theta}]} H s'_1(d\theta) \right]
\]

(integrating by parts and using the fact that \( H \) is continuous). Moreover

\[
\left[ s_{1H} \bar{\theta}^+ \right]_{\theta_1} = s_{1R}(\bar{\theta}) H(\bar{\theta}) - s_{1L}(\theta_1) H(\theta_1)
\]

and

\[
\int_{[\theta_1, \bar{\theta}]} H s'_1(d\theta) = H(\theta_1) \Delta s_1(\theta_1) + \int_{(\theta_1, \bar{\theta})} H s'_1(d\theta) + H(\bar{\theta}) \Delta s_1(\bar{\theta}).
\]

Hence, overall,

\[
\nabla_s L(r; \lambda_R) = -H(\theta_1) s_{1R}(\theta_1) - \int_{(\theta_1, \bar{\theta})} H s'_1(d\theta) + H(\bar{\theta}) s_{1L}(\bar{\theta}) = -\int_{(\theta_1, \bar{\theta})} H s'_1(d\theta)
\]

(since \( H(\bar{\theta}) = 0 \) by construction and \( H(\theta_1) = 0 \) by definition of \( \theta_1 \)). Now \( v'_1 \geq 0 \) on the whole of \( \Theta \), since \( v_1 \) is non-decreasing, and \( r'_1 = 0 \) on \((\theta_1, \bar{\theta})\), since \( r_1 \) is constant there. Hence \( s'_1 \geq 0 \) on \((\theta_1, \bar{\theta})\). On the other hand, for all \( \theta \in [\theta_1, \bar{\theta}] \), we have

\[
H(\theta) = \int_{\theta}^{\bar{\theta}} (\Gamma - F(\bar{\theta})) \ell(d\theta) = (\bar{\theta} - \theta) \left( \frac{1}{\bar{\theta} - \theta} \int_{\theta}^{\bar{\theta}} \Gamma \ell(d\theta) - F(\bar{\theta}) \right) \geq 0,
\]

by definition of \( \theta_1 \). Hence \( \nabla_s L(r; \lambda_R) \leq 0 \), as required.

**Remark 14** Notice that \( s_1 \) is the difference of the two non-decreasing functions \( v_1 \) and \( r_1 \). Hence there is no general reason why \( s_1 \) should be non-decreasing. The situation is saved by the fact that \( r_1 \) is constant on \((\theta_1, \bar{\theta})\).
Suppose now that \( \theta_1 \leq \bar{\theta} \). Then, in the light of the discussion in Section A.8, we have

\[
\nabla_s L(r; \lambda_R) = \frac{1}{\beta} s_1 R(\bar{\theta}) (\bar{\theta} - \theta_1) F(\bar{\theta}) + \frac{1}{\beta} \int s_1 R(F(\bar{\theta}) - \Gamma) \ell(d\theta).
\]

Now, arguing as in the case \( \theta_1 > \bar{\theta} \), we have

\[
\int s_1 R(F(\bar{\theta}) - \Gamma) \ell(d\theta) = \int_{[\theta_1, \bar{\theta}]} s_1 R(F(\bar{\theta}) - \Gamma) \ell(d\theta)
= \int_{[\theta_1, \bar{\theta}]} s_1 R H' \ell(d\theta)
= [ s_1 H]_{\theta_1}^{\bar{\theta}} + \int_{[\theta_1, \bar{\theta}]} H s'_1(d\theta)
= -H(\bar{\theta}) s_1 R(\bar{\theta}) - \int_{[\theta_1, \bar{\theta}]} H s'_1(d\theta) + H(\bar{\theta}) s_1 L(\bar{\theta})
= -H(\bar{\theta}) s_1 R(\bar{\theta}) - \int_{[\theta_1, \bar{\theta}]} H s'_1(d\theta)
\]

(since \( H(\bar{\theta}) = 0 \) by construction). Hence, overall, we have

\[
\beta \nabla_s L(r; \lambda_R) = ((\bar{\theta} - \theta_1) F(\bar{\theta}) - H(\bar{\theta})) s_1 R(\bar{\theta}) - \int_{[\theta_1, \bar{\theta}]} H s'_1(d\theta).
\]

But

\[
(\bar{\theta} - \theta_1) F(\bar{\theta}) - H(\bar{\theta}) = \int_{\theta_1}^{\bar{\theta}} F(\bar{\theta}) \ell(d\theta) - \int_{\theta_1}^{\bar{\theta}} (\Gamma - F(\bar{\theta})) \ell(d\theta)
\]

(by definition of \( H \))

\[
= -\int_{\theta_1}^{\bar{\theta}} (\Gamma - F(\bar{\theta})) \ell(d\theta) - \int_{\theta_1}^{\bar{\theta}} (\Gamma - F(\bar{\theta})) \ell(d\theta)
\]

(since \( \Gamma = 0 \) on \([\theta_1, \bar{\theta})\))

\[
= -H(\theta_1)
\]

(by definition of \( H \) again)

\[
= 0.
\]
(by definition of \( \theta_1 \)). Hence

\[
\beta \nabla_s L(r; \lambda_R) = - \int_{(\theta, \theta)} H s'(d\theta).
\]

Hence, arguing as in the case \( \theta_1 > \theta \), \( \nabla_s L(r; \lambda_R) \leq 0 \).

## B Proof of Theorem 2

As explained in the text, we continue to make Assumptions A1-A5 and add an additional assumption, namely Assumption A6.

**Theorem 2** Suppose that inter-household transfers are possible. Then a two-account system with one completely liquid account and one completely illiquid account does not maximize welfare.

### B.1 The Optimization Problem of the Planner

If self 1 is presented with two accounts, a perfectly liquid account containing the amount \( x_{\text{liquid}} > 0 \) and a perfectly illiquid account containing the amount \( x_{\text{illiquid}} \geq 0 \), then the outcome will depend on her type \( \theta \). There will exist \( \theta_2 \in (0, \infty) \) such that: if \( \theta < \theta_2 \), then she consumes less than the balance \( x_{\text{liquid}} \) in her liquid account; and, if \( \theta \geq \theta_2 \), then she consumes the whole of \( x_{\text{liquid}} \). The cutoff \( \theta_2 \) need not lie in \( [\theta, \theta] \). It could be that \( \theta_2 < \theta \), in which case there will be perfect pooling: all types will consume the whole of \( x_{\text{liquid}} \) and both \( c_1 \) and \( c_2 \) will be constant. Or it could be that \( \theta_2 > \theta \), in which case there will be perfect separation: all types will consume less than \( x_{\text{liquid}} \); \( c_1 \) will be strictly increasing in \( \theta \) and \( c_2 \) will be strictly decreasing in \( \theta \).

More generally, we will obtain consumption allocations \( c_1, c_2 : \Theta \to (0, \infty) \) and associated utility allocations \( r_1, r_2 : \Theta \to \mathbb{R} \), where the latter are given by the formulae \( r_1(\theta) = u_1(c_1(\theta)) \)
and \( r_2(\theta) = v_2(c_2(\theta)) \). The overall utility allocation \( r = (r_1, r_2) \) will be a smooth function of \( \theta \) for \( \theta < \theta_2 \), have a kink at \( \theta_2 \), and be constant for \( \theta > \theta_2 \). The idea behind the proof is to find necessary conditions for utility allocations of this type to be optimal, and to use these necessary conditions to derive a contradiction.

The first step is to formulate the optimization problem of the planner. We do this in terms of general utility allocations \( v_1, v_2 : \Theta \to \mathbb{R} \), reserving the notation \( r_1, r_2 \) for the specific allocations arising from two-account systems with one completely liquid account and one completely illiquid account. Accordingly, the planner seeks to maximize social welfare

\[
\int (\theta v_1(\theta) + v_2(\theta)) \, dF(\theta)
\]

over utility allocations

\[
(v_1, v_2) : [\overline{\overline{\theta}}, \overline{\overline{\theta}}] \to (u_1(0+), u_1(\infty-)) \times (u_2(0+), u_2(\infty-))
\]

subject to aggregate budget balance and incentive compatibility. Aggregate budget balance can be expressed in the form

\[
\int (Y - C_1(v_1(\theta)) - C_2(v_2(\theta))) \, dF(\theta) \geq 0,
\]

(BC)

where \( C_t = u_t^{-1} \) for \( t \in \{1, 2\} \). Incentive compatibility breaks down into two parts, a linear part

\[
\theta v'_1 + \beta v'_2 = 0
\]

(ICL)

and a monotonic part

\[
v'_2 \leq 0.
\]

(ICM)

**Remark 15** The two conditions (ICL) and (ICM) are simply the differential counterpart of the usual integral representation of incentive compatibility in a mechanism-design problem.
B.2 The Case $\theta_2 \in (\underline{\theta}, \overline{\theta})$

Consider first the case in which $x_{\text{liquid}}$ and $x_{\text{illiquid}}$ are such that $\theta_2 \in (\underline{\theta}, \overline{\theta})$. In this case, the second step is to parameterize candidate solutions $v = (v_1, v_2)$ to the planner’s problem in terms of boundary values $v_1(\overline{\theta})$, $v_2(\overline{\theta})$ and continuous functions $v'_{1L} : [\underline{\theta}, \theta_2] \to \mathbb{R}$, $v'_{1R} : [\theta_2, \overline{\theta}] \to \mathbb{R}$. More precisely, we can put:

1. $v_1(\theta) = v_1(\overline{\theta}) - \int_{\theta}^{\overline{\theta}} v'_{1R}(t) \, dt$ for $\theta \in [\theta_2, \overline{\theta}]$;
2. $v_1(\theta) = v_1(\theta_2) - \int_{\theta}^{\theta_2} v'_{1L}(t) \, dt$ for $\theta \in [\underline{\theta}, \theta_2]$;
3. $v'_{2R}(\theta) = -\frac{\partial}{\partial \theta} v'_{1R}(\theta)$ for $\theta \in [\theta_2, \overline{\theta}]$;
4. $v'_{2L}(\theta) = -\frac{\partial}{\partial \theta} v'_{1L}(\theta)$ for $\theta \in [\underline{\theta}, \theta_2]$;
5. $v_2(\theta) = v_2(\overline{\theta}) - \int_{\theta}^{\overline{\theta}} v'_{2R}(t) \, dt$ for $\theta \in [\theta_2, \overline{\theta}]$;
6. $v_2(\theta) = v_2(\theta_2) - \int_{\theta}^{\theta_2} v'_{2L}(t) \, dt$ for $\theta \in [\underline{\theta}, \theta_2]$.

In other words: $v_1$ is the continuous function with continuous derivative $v'_{1L}$ on $[\underline{\theta}, \theta_2)$, continuous derivative $v'_{1R}$ on $(\theta_2, \overline{\theta}]$ and value $v_1(\overline{\theta})$ at $\overline{\theta}$; and $v_2$ is the continuous function with continuous derivative $v'_{2L}$ on $[\underline{\theta}, \theta_2)$, continuous derivative $v'_{2R}$ on $(\theta_2, \overline{\theta}]$ and value $v_2(\overline{\theta})$ at $\overline{\theta}$.

Remark 16 Notice that the two-account system described in Theorem 2 gives rise to a utility allocation $r = (r_1, r_2)$ satisfying conditions 1-6. Moreover – as we shall see below – in order to show that $r$ is not optimal, it suffices to consider variations in this same class. We simply do not need to consider variations in which (say) $\theta_2$ changes or $v = (v_1, v_2)$ can be discontinuous.
The third step is to formulate the Langrangian. This can be written

\[
L(v_1(\theta), v_2(\theta), v'_1L, v'_1R, \lambda, \zeta_L, \zeta_R) = \int (\theta v_1(\theta) + v_2(\theta)) \, dF(\theta) + \lambda \int (Y - C_1(v_1(\theta)) - C_2(v_2(\theta))) \, dF(\theta) - \int_{[\theta_1, \theta_2]} v'_{2L}(\theta) \, d\zeta_L(\theta) - \int_{[\theta_2, \theta]} v'_{2R}(\theta) \, d\zeta_R(\theta),
\]

where:

1. the arguments of \( L \) are the parameters \( v_1(\theta), v_2(\theta), v'_1L \) and \( v'_1R \), and the multipliers \( \lambda, \zeta_L \) and \( \zeta_R \);
2. \( \lambda \) is a scalar (namely the multiplier on the aggregate budget constraint);
3. \( \zeta_L \) is a finite non-negative Borel measure on \([\theta_1, \theta_2]\) (namely the multiplier associated with the non-positivity constraint on \( v'_{2L} \));
4. \( \zeta_R \) is a finite non-negative Borel measure on \([\theta_2, \theta]\) (namely the multiplier associated with the non-positivity constraint on \( v'_{2R} \));
5. the variables \( v_1, v_2, v'_2L \) and \( v'_2R \) on the right-hand side are determined by the parameters \( v_1(\theta), v_2(\theta), v'_1L \) and \( v'_1R \) as explained above.

**Remark 17** The Langrangian does not include a term corresponding to (ICL). This is because we have used (ICL) to solve for \( v'_{2L} \) and \( v'_{2R} \) in terms of \( v'_1L \) and \( v'_1R \).

The fourth step is to note that we can associate parameters \( (r_1(\theta), r_2(\theta), r'_1L, r'_1R) \) with the reference utility allocation \( (r_1, r_2) \) and parameters \( (v_1(\theta), v_2(\theta), v'_1L, v'_1R) \) with the alternative utility allocation \( (v_1, v_2) \) in the obvious way, and take the derivative of the Langrangian at the parameter values \( (r_1(\theta), r_2(\theta), r'_1L, r'_1R) \) in the direction \( (s_1(\theta), s_2(\theta), s'_1L, s'_1R) \), where \( s = v - r \). Furthermore, this calculation can be simplified by noting that the variables
$(v_1, v_2, v'_2L, v'_2R)$ in the RHS of the equation for the Langrangian are linear in the underlying parameters $(v_1(\overline{\theta}), v_2(\overline{\theta}), v'_1L, v'_1R)$. Hence we can simply take the derivative of the RHS at the point $(r_1, r_2, r'_2L, r'_2R)$ in the direction $(s_1, s_2, s'_2L, s'_2R)$ and only then substitute for $(s_1, s_2, s'_2L, s'_2R)$ in terms of $(s_1(\overline{\theta}), s_2(\overline{\theta}), s'_1L, s'_1R)$.

Taking the derivative of the RHS at the point $(r_1, r_2, r'_2L, r'_2R)$ in the direction $(s_1, s_2, s'_2L, s'_2R)$, we obtain

$$0 = \int (\theta s_1 + s_2) dF - \lambda \int (C'_1(r_1) s_1 + C'_2(r_2) s_2) dF$$

$$- \int_{[\theta, \theta_2]} s'_2L(\theta) d\zeta_L(\theta) - \int_{[\theta_2, \theta]} s'_2R(\theta) d\zeta_R(\theta) \quad (9)$$

for all feasible $(s_1, s_2, s'_2L, s'_2R)$. Moreover, the constraints must all be satisfied. That is,

$$0 = \int (Y - C_1(r_1(\theta)) - C_2(r_2(\theta))) dF(\theta),$$

$$0 \geq r'_2L,$$

$$0 \geq r'_2R.$$

Finally, constraint qualification must hold. That is,

$$0 = \int_{[\theta, \theta_2]} r'_2L(\theta) d\zeta_L(\theta), \quad (10)$$

$$0 = \int_{[\theta_2, \theta]} r'_2R(\theta) d\zeta_R(\theta). \quad (11)$$

Furthermore, a variation $(s_1, s_2, s'_2L, s'_2R)$ is feasible iff it can be expressed in terms of the underlying parameters $(s_1(\overline{\theta}), s_2(\overline{\theta}), s'_1L, s'_1R)$. We therefore substitute for the variation $(s_1, s_2, s'_2L, s'_2R)$ in terms of the underlying parameters $(s_1(\overline{\theta}), s_2(\overline{\theta}), s'_1L, s'_1R)$ and manipulate the RHS in such a way as to expose the linear dependence of the RHS on $s_1(\overline{\theta}), s_2(\overline{\theta}), s'_1L$ and $s'_1R$. 
The first contribution to the RHS is \( \int \theta s_1 dF(\theta) \). Putting \( \overline{F}(\theta) = \int_{\theta} F(t) \, dt \), and noting that \( \theta F - \overline{F} \) and \( s_1 \) are both continuous, we can integrate this contribution by parts to obtain

\[
\int \theta s_1 dF(\theta) = \left[ (\theta F - \overline{F}) s_1 \right]_{\theta_0}^{\theta} - \int (\theta F - \overline{F}) s'_1 \, d\theta \\
= (\overline{F}(\overline{\theta}) - \overline{F}(\underbar{\theta})) s_1(\overline{\theta}) - \int (\theta F - \overline{F}) s'_1 \, d\theta \\
= (\overline{F}(\overline{\theta}) - \overline{F}(\underbar{\theta})) s_1(\overline{\theta}) \\
- \int_{[\theta_0, \theta_1]} (\theta F - \overline{F}) s'_{1L} \, d\theta - \int_{[\theta_2, \theta_\infty]} (\theta F - \overline{F}) s'_{1R} \, d\theta,
\]

The second contribution to the RHS is \( \int s_2 dF(\theta) \). For this contribution, we have

\[
\int s_2 dF(\theta) = [F s_2]_{\overline{\theta}}^{\theta} - \int F s'_2 \, d\theta \\
= F(\overline{\theta}) s_2(\overline{\theta}) - \int F s'_2 \, d\theta \\
= F(\overline{\theta}) s_2(\overline{\theta}) - \int_{[\theta_0, \theta_1]} F s'_{2L} \, d\theta - \int_{[\theta_2, \theta_\infty]} F s'_{2R} \, d\theta \\
= F(\overline{\theta}) s_2(\overline{\theta}) + \int_{[\theta_0, \theta_1]} F \frac{\theta}{\overline{\theta}} s'_{1L} \, d\theta + \int_{[\theta_2, \theta_\infty]} F \frac{\theta}{\overline{\theta}} s'_{1R} \, d\theta.
\]

Next, putting \( \Lambda_1(\theta) = \int_{[\theta, \theta]} C'_1(r_1(t)) \, dF(t) \), we have

\[
-\lambda \int C'_1(r_1) s_1 \, dF = -\int s_1 \lambda \Lambda'_1 \, d\theta \\
= -[s_1 \lambda \Lambda_1]_{\overline{\theta}}^{\theta} - \int \lambda \Lambda_1 s'_1 \, d\theta \\
= -s_1(\overline{\theta}) \lambda \Lambda_1(\overline{\theta}) + \int \lambda \Lambda_1 s'_1 \, d\theta \\
= -s_1(\overline{\theta}) \lambda \Lambda_1(\overline{\theta}) \\
+ \int_{[\theta_0, \theta_1]} \lambda \Lambda_1 s'_{1L} \, d\theta + \int_{[\theta_2, \theta_\infty]} \lambda \Lambda_1 s'_{1R} \, d\theta.
\]
Similarly, putting $\Lambda_2(\theta) = \int_{\theta}^{\bar{\theta}} C_2'(r_2(t)) \, dF(t)$,

$$-\lambda \int C_2'(r_2) \, s_2 \, dF = - \int s_2 \, \lambda \, \Lambda_2' \, d\theta$$

$$= - s_2(\theta) \, \lambda \, \Lambda_2(\theta) + \int \lambda \, \Lambda_2 \, s_2' \, d\theta$$

$$= - s_2(\theta) \, \lambda \, \Lambda_2(\theta)$$

$$+ \int_{[\theta, \theta_2]} \lambda \, \Lambda_2 \, s_{2L}' \, d\theta + \int_{[\theta_2, \bar{\theta}]} \lambda \, \Lambda_2 \, s_{2R}' \, d\theta$$

$$= - s_2(\theta) \, \lambda \, \Lambda_2(\theta)$$

$$- \int_{[\theta, \theta_2]} \lambda \, \Lambda_2 \, \frac{\theta}{\pi} \, s_{1L}' \, d\theta - \int_{[\theta_2, \bar{\theta}]} \lambda \, \Lambda_2 \, \frac{\theta}{\pi} \, s_{1R}' \, d\theta.$$  

Finally, we have

$$- \int_{[\theta, \theta_2]} s_{2L}'(\theta) \, d\zeta_L(\theta) = \int_{[\theta, \theta_2]} \frac{\theta}{\pi} \, s_{1L}'(\theta) \, d\zeta_L(\theta)$$

and

$$- \int_{[\theta_2, \bar{\theta}]} s_{2R}'(\theta) \, d\zeta_R(\theta) = \int_{[\theta_2, \bar{\theta}]} \frac{\theta}{\pi} \, s_{1R}'(\theta) \, d\zeta_R(\theta).$$

The fifth step is to equate the coefficients of $s_1(\bar{\theta})$, $s_2(\bar{\theta})$, $s_{1L}'$ and $s_{1R}'$ to 0. Doing so yields:

$$0 = \bar{\theta} \, F(\bar{\theta}) - F(\bar{\theta}) - \lambda \, \Lambda_1(\bar{\theta}), \quad (12)$$

$$0 = F(\bar{\theta}) - \lambda \, \Lambda_2(\bar{\theta}), \quad (13)$$

$$0 = -(\theta \, F - \bar{\theta} \, F) \, d\theta + \frac{\theta}{\pi} \, F \, d\theta + \lambda \, \Lambda_1 \, d\theta - \frac{\theta}{\pi} \, \lambda \, \Lambda_2 \, d\theta + \frac{\theta}{\pi} \, d\zeta_L, \quad (14)$$

$$0 = -(\theta \, F - \bar{\theta} \, F) \, d\theta + \frac{\theta}{\pi} \, F \, d\theta + \lambda \, \Lambda_1 \, d\theta - \frac{\theta}{\pi} \, \lambda \, \Lambda_2 \, d\theta + \frac{\theta}{\pi} \, d\zeta_R. \quad (15)$$

Now, we certainly have $r_{2L}' < 0$ on $[\theta, \theta_2]$. (This is because, if $\theta < \theta_2$, then self 1 consumes less than $x_{\text{liquid}}$. Hence $r_{1L}' > 0$ and $r_{2L}' < 0$.) It therefore follows from constraint qualification
(namely (10)) that \( \zeta_L = 0 \). Equation (14) therefore implies that

\[
\lambda (\theta \Lambda_2 - \beta \Lambda_1) = \theta F - \beta (\theta F - F) = (1 - \beta) \theta F + \beta F = \Gamma
\]  

(16)

almost everywhere on \([\underline{\theta}, \theta_2]\), where \( \Gamma = (1 - \beta) \theta F' + F \) and \( \Gamma(\theta) = \int_{[\underline{\theta}, \theta]} \Gamma(t) \, dt \). Furthermore, since \( F' \) is of bounded variation,

\[
\begin{align*}
\frac{\theta \Lambda_2(\theta)}{\theta - \bar{\theta}} &\to \theta C'_2(r_2(\underline{\theta})) F'(\underline{\theta}^+), \\
\frac{\beta \Lambda_1(\theta)}{\theta - \bar{\theta}} &\to \beta C'_1(r_1(\underline{\theta})) F'(\underline{\theta}^+), \\
\frac{\Gamma}{\theta - \bar{\theta}} &\to \Gamma(\underline{\theta}^+) = (1 - \beta) \theta F'(\underline{\theta}^+)
\end{align*}
\]

as \( \theta \downarrow \underline{\theta} \). But, since \((r_1(\underline{\theta}), r_2(\underline{\theta}))\) is chosen freely from the ambient budget line by the \( \underline{\theta} \) type, we must have

\[
\frac{C'_1(r_1(\underline{\theta}))}{\theta} = \frac{C'_2(r_2(\underline{\theta}))}{\beta}.
\]

We therefore have

\[
\frac{\theta \Lambda_2(\theta) - \beta \Lambda_1(\theta)}{\theta - \bar{\theta}} \to 0
\]

as \( \theta \downarrow \underline{\theta} \). On the other hand,

\[
\frac{\Gamma}{\theta - \bar{\theta}} \to (1 - \beta) \theta F'(\underline{\theta}^+) > 0
\]

as \( \theta \downarrow \underline{\theta} \). Passing to the limit in equation (16), we therefore obtain

\[
0 = (1 - \beta) \theta F'(\underline{\theta}^+).
\]

But all three terms on the RHS are strictly positive. Indeed: \( \beta < 1; \underline{\theta} > 0 \); and \( F' \) is bounded away from 0 on \((\underline{\theta}, \bar{\theta})\). We have therefore reached a contradiction. This establishes that we cannot have \( \theta_2 \in (\underline{\theta}, \bar{\theta}) \).
B.3 The Case $\theta_2 \in [\overline{\theta}, \infty)$

Consider now the case in which $x_{\text{liquid}}$ and $x_{\text{iliquid}}$ are such that $\theta_2 \in [\overline{\theta}, \infty)$. In this case, we can derive equations (12, 13 and 14) exactly as in Section B.2 above. In particular, we can still derive equation (14). We can therefore derive a contradiction by essentially the same argument.

B.4 The Case $\theta_2 \in (0, \overline{\theta}]$

Consider now the case in which $x_{\text{liquid}}$ and $x_{\text{iliquid}}$ are such that $\theta_2 \in (0, \overline{\theta}]$. In this case, we can still derive equations (12, 13 and 15). However, we can no longer derive equation (14). We therefore need new arguments. The first point to note is that, since $\theta_2 \leq \overline{\theta}$ all types $\theta \in [\theta, \overline{\theta}]$ choose the point that a hypothetical $\theta_2$ type would choose from the ambient budget set. We therefore have

\begin{align}
    \Lambda_1(\overline{\theta}) &= \int_{[0, \overline{\theta}]} C'_1(r_1(t)) dF(t) = F(\overline{\theta}) C'_1(r_1(\theta_2)), \quad (17) \\
    \Lambda_2(\overline{\theta}) &= \int_{[0, \overline{\theta}]} C'_2(r_2(t)) dF(t) = F(\overline{\theta}) C'_2(r_2(\theta_2)). \quad (18)
\end{align}

Furthermore, since the $\theta_2$ type chooses freely from the ambient budget set, we have

\[
\frac{C'_1(r_1(\theta_2))}{\theta_2} = \frac{C'_2(r_2(\theta_2))}{\beta}.
\]

Using (12) and (13), we therefore obtain

\[
\frac{\overline{\theta} F(\overline{\theta}) - F'(\overline{\theta})}{F(\overline{\theta})} = \frac{\Lambda_1(\overline{\theta})}{\Lambda_2(\overline{\theta})} = \frac{C'_1(r_1(\theta_2))}{C'_2(r_2(\theta_2))} = \frac{\theta_2}{\beta}. \quad (19)
\]
Hence

\[(\bar{\theta} - \theta_2) F(\bar{\theta}) = \bar{\theta} F(\bar{\theta}) - \beta \left( \bar{\theta} F(\bar{\theta}) - F(\bar{\theta}) \right)\]

\[= (1 - \beta) \bar{\theta} F(\bar{\theta}) + \beta F(\bar{\theta})\]

\[= \Gamma(\bar{\theta}), \quad (20)\]

where \(\Gamma\) and \(\bar{\Gamma}\) are as above.

**Remark 18** Bearing in mind that \(\theta_2 \leq \underline{\theta}\), so that \(\Gamma(\theta_2) = 0\), this equation can also be written

\[(\bar{\theta} - \theta_2) F(\bar{\theta}) = \Gamma(\bar{\theta}) - \Gamma(\theta_2)\]

or

\[\frac{1}{\bar{\theta} - \theta_2} \int_{[\theta_2, \bar{\theta}]} \Gamma(t) \, dt = F(\bar{\theta}). \quad (21)\]

The significance of this observation is that \(\theta_1\) satisfies equation (21) too. So, while the necessary conditions that we have used here do not quite imply that \(\theta_2 = \theta_1\), they do highlight a close relationship between the two. The intuitive reason for this relationship is clear.

If \(\theta_2 \leq \underline{\theta}\) then all types make the same choice. In particular, there are no interpersonal transfers. Since this outcome is — by hypothesis — the optimum in the class of outcomes with or without transfers, then a fortiori it is the optimum in the class of outcomes without transfers.

However, we have not yet used equation (15). It follows from this equation that

\[d\zeta_R = \frac{\partial}{\partial \theta} \left( \theta F - \bar{F}' \right) d\theta - F \, d\theta + \lambda \left( \Lambda_2 - \frac{\partial}{\partial \theta} \Lambda_1 \right) \, d\theta.\]

In other words, \(\zeta_R\) is absolutely continuous w.r.t. Lebesgue measure, with density

\[\zeta'_R = \frac{\partial}{\partial \theta} \left( \theta F - \bar{F}' \right) - F + \lambda \left( \Lambda_2 - \frac{\partial}{\partial \theta} \Lambda_1 \right).\]
Furthermore:

\[
\Lambda_1(\theta) = \int_{[\theta, \bar{\theta}]} C_1'(r_1(t)) \, dF(t) = F(\theta) \, C_1'(r_1(\theta_2)) = \frac{F(\theta)}{F(\bar{\theta})} \Lambda_1(\bar{\theta}) \\
= \frac{F(\theta)}{F(\bar{\theta})} \frac{\theta_2}{\beta} \Lambda_2(\bar{\theta}) = \frac{F(\theta)}{F(\bar{\theta})} \frac{\theta_2}{\beta} \frac{F(\bar{\theta})}{\lambda} = \frac{\theta_2}{\beta} \frac{F(\theta)}{\lambda}
\]

(where the last line follows from (19) and (13)); and

\[
\Lambda_2(\theta) = \int_{[\theta, \bar{\theta}]} C_2'(r_2(t)) \, dF(t) = F(\theta) \, C_2'(r_2(\theta_2)) = \frac{F(\theta)}{F(\bar{\theta})} \Lambda_2(\bar{\theta}) \\
= \frac{F(\theta)}{F(\bar{\theta})} \frac{F(\bar{\theta})}{\lambda} = \frac{F(\theta)}{\lambda}
\]

(where the last line follows from (13)). Hence

\[
\lambda (\theta \Lambda_2 - \beta \Lambda_1) = (\theta - \theta_2) \, F(\theta)
\]

and

\[
\theta \zeta'_R = \beta (\theta F - \bar{\Gamma}) - \theta F + (\theta - \theta_2) F \\
= (\theta - \theta_2) F(\theta) - \bar{\Gamma}.
\]

Now, \( F(\bar{\theta}) = \Gamma(\bar{\theta}) = 0 \). Hence \( \theta \zeta'_R(\bar{\theta}) = 0 \). Furthermore, we must have \( \theta \zeta'_R \geq 0 \) on \((\bar{\theta}, \theta)\). Hence

\[
\frac{\theta \zeta'_R(\theta) - \theta \zeta'_R(\bar{\theta})}{\theta - \bar{\theta}} \geq 0.
\]

Letting \( \theta \to \bar{\theta}^+ \), we therefore obtain

\[
(\theta \zeta'_R)'(\bar{\theta}^+) = (\beta \theta - \theta_2) \, F'(\bar{\theta}^+) \geq 0.
\]
Since $F'(\theta^+)>0$, it follows that
\[ \theta_2 \leq \beta \theta. \quad (22) \]

Similarly, (20) implies that $(\bar{\vartheta} - \theta_2) F' (\bar{\vartheta}) - \Gamma (\bar{\vartheta}) = 0$. Hence $\bar{\vartheta} \zeta'_R (\bar{\vartheta}) = 0$. Hence
\[ \frac{\bar{\vartheta} \zeta'_R (\bar{\vartheta}) - \theta \zeta'_R (\theta)}{\bar{\vartheta} - \theta} \leq 0. \]

Letting $\theta \to \bar{\vartheta}^-$, we therefore obtain
\[ (\theta \zeta'_R)' (\bar{\vartheta}^-) = (\beta \bar{\vartheta} - \theta_2) F' (\bar{\vartheta}^-) \leq 0. \]

Since $F'(\bar{\vartheta}^-)>0$, it follows that
\[ \theta_2 \geq \beta \bar{\vartheta}. \quad (23) \]

But inequalities (22) and (23) are inconsistent with one another, so we have a contradiction.

**Remark 19** We can use the preceding analysis to obtain some perspective on why a pooling mechanism in which all resources are placed in the illiquid account is never optimal. Suppose that we replace the inequality constraint $0 \geq r'_{2R}$ with an equality constraint and choose the multiplier $\zeta_R$ in such a way that this constraint is respected. Then, proceeding almost exactly as above, we will obtain
\[ (\theta \zeta'_R)' = (\theta - \theta_2) F' + F - \Gamma \]
\[ = (\theta - \theta_2) F' - (1 - \beta) \theta F' \]
\[ = (\beta \theta - \theta_2) F'. \]

Moreover we will have the boundary conditions $\bar{\vartheta} \zeta'_R (\bar{\vartheta}) = 0$ and $\bar{\vartheta} \zeta'_R (\bar{\vartheta}) = 0$. It follows that $\theta_2 \in (\beta \bar{\vartheta}, \bar{\vartheta})$ and $\theta \zeta'_R < 0$ on $(\theta, \bar{\vartheta})$. Hence a small change in the direction of any incentive-compatible and fully separating mechanism is desirable. (This would have the effect of reducing $r'_2$ from 0 – and increasing $r'_1$ from 0 – at all points in the range $(\theta, \bar{\vartheta})$. ) In other
words, it is always desirable to allow some flexibility to the decision maker to respond to the information contained in $\theta$.

C Analysis of a General (Non-Linear) Mechanism

C.1 The Mechanism-Design Problem

In the general mechanism-design problem, the planner chooses a budget set

$$C \subset (0, \infty)^2$$

and consumption allocations $c_1, c_2 : \Theta \times B \rightarrow (0, \infty)$ to maximize welfare

$$\int \int (\theta u_1(c_1(\theta, \beta)) + u_2(c_2(\theta, \beta))) f(\theta) g(\beta) d\theta d\beta$$

subject to the resource constraint

$$\int \int (Y - c_1(\theta, \beta) - \frac{1}{R} c_2(\theta, \beta)) f(\theta) g(\beta) d\theta d\beta \geq 0$$

and the incentive-compatibility constraint

$$(c_1(\theta, \beta), c_2(\theta, \beta)) \in \arg\max_{(c_1, c_2) \in C} \left\{ \theta u_1(\tilde{c}_1) + \beta u_2(\tilde{c}_2) \right\}.$$ 

Here, $f$ is the density of $\theta$ (associated with distribution function $F$ in the main text); $g$ is the density of $\beta$ (associated with distribution function $G$ in the main text); $Y$ is the per capita endowment; and $R$ is the gross rate of return. Furthermore, we assume that: $\Theta = [\underline{\theta}, \bar{\theta}]$; $B = [\underline{\beta}, \bar{\beta}]$; $0 < \underline{\theta} < \bar{\theta} < \infty$; $0 < \underline{\beta} < \bar{\beta} < \infty$; $f$ is continuous and bounded away from 0 on $\Theta$; $g$ is continuous and bounded away from 0 on $B$. 

77
Remark 20 For example: $f$ might take the form
\[
f(\theta) = \frac{\exp\left(-\frac{1}{2} \left(\frac{\theta - \mu}{\sigma}\right)^2\right)}{\int_{\theta}^{\overline{\theta}} \exp\left(-\frac{1}{2} \left(\frac{\theta - \mu}{\sigma}\right)^2\right) d\theta}
\]
for $\theta \in [\underline{\theta}, \overline{\theta}]$ and $f(\theta) = 0$ otherwise, i.e., $f$ might be the density of the univariate normal distribution with mean $\mu$ and variance $\sigma^2$ truncated to the interval $[\underline{\theta}, \overline{\theta}]$; and $g$ might take the form
\[
g(\beta) = \frac{1}{\beta - \underline{\beta}} \quad \text{for} \quad \beta \in [\underline{\beta}, \overline{\beta}]
\]
and $g(\beta) = 0$ otherwise, i.e., $g$ might be the density of the uniform distribution on the interval $[\underline{\beta}, \overline{\beta}]$.

C.2 Transforming the Problem

The first step in solving this problem is to note that
\[
(c_1, c_2) \in \arg\max \{\theta u_1(\tilde{c}_1) + \beta u_2(\tilde{c}_2)\}
\]
iff
\[
(c_1, c_2) \in \arg\max \left\{\frac{\phi}{\beta} u_1(\tilde{c}_1) + u_2(\tilde{c}_2)\right\}.
\]
The set of optimal choices of the individual therefore depends only on $\phi = \theta / \beta$. Combining this fact with the assumed continuity of the distribution functions $F$ and $G$ of $\theta$ and $\beta$ implies that, if we put $\Phi = [\underline{\phi}, \overline{\phi}]$ where $\underline{\phi} = \theta / \overline{\beta}$ and $\overline{\phi} = \theta / \underline{\beta}$, then the planner can work with consumption allocations $c_1, c_2 : \Phi \to (0, \infty)$ instead of with consumption allocations $c_1, c_2 : \Theta \times B \to (0, \infty)$.

The second step is to note that we can work with utility allocations $v_1, v_2 : \Phi \to \mathbb{R}$ instead of with consumption allocations $c_1, c_2 : \Phi \to (0, \infty)$. The former are related to the latter via the formulae $v_1(\phi) = u_1(c_1(\phi))$ and $v_2(\phi) = u_2(c_2(\phi))$. We can also invert these formulae to
get $c_1(\phi) = C_1(v_1(\phi))$ and $c_2(\phi) = C_2(v_2(\phi))$.

The third step is to note that we can change variables in the integral defining welfare and in the integral giving the resource constraint, replacing $(\theta, \beta)$ with $(\phi, \beta)$.

At this point, the planner’s problem can be expressed as that of choosing $v_1, v_2 : \Phi \rightarrow \mathbb{R}$ to maximize welfare

$$\int \int (\beta \phi v_1(\phi) + v_2(\phi)) \beta f(\beta \phi) g(\beta) d\phi d\beta$$

subject to the resource constraint

$$\int \int \left( Y - C_1(v_1(\phi)) - \frac{1}{R} C_2(v_2(\phi)) \right) \beta f(\beta \phi) g(\beta) d\phi d\beta \geq 0$$

and the incentive-compatibility constraint, which now has two parts, namely a linear part,

$$0 = \phi v_1'(\phi) + v_2'(\phi)$$

(ICL)

and a monotonic part,

$$0 \leq -v_2'(\phi).$$

(ICM)

**Remark 21** Notice that, whenever $c_1$ and $c_2$ are chosen from a budget set $C$, $v_1$ will be non-decreasing and $v_2$ will be non-increasing. However, neither function need be differentiable (or even continuous). Hence the derivatives $v_1'$ and $v_2'$ might in principle be a non-negative and a non-positive measure respectively. This does not invalidate (IC1) or (IC2), both of which make sense for measures. However, in what follows, we will sometimes reason as if $v_1'$ and $v_2'$ exist in the usual sense.

The fourth step is to introduce the marginal density $h$ of $\phi$ and the conditional density $j$ of $\beta$ given $\phi$, namely

$$h(\phi) = \int \beta f(\beta \phi) g(\beta) d\beta$$

(24)
and
\[
j(\beta \mid \phi) = \frac{\beta f(\beta \phi) g(\beta)}{h(\phi)}.
\] (25)

We can also introduce the conditional expectation of \(\beta\), namely
\[
b(\phi) = \int \beta j(\beta \mid \phi) \, d\beta.
\] (26)

**Remark 22** The limits of integration in the definition of \(h\) (namely (24)) are implicit in the definitions of \(f\) and \(g\). Since the integrand will only be non-zero if both \(f(\beta \phi)\) and \(g(\beta)\) are non-zero, these limits are \(\max\{\underline{\beta}, \theta / \phi\}\) and \(\min\{\overline{\beta}, \overline{\theta} / \phi\}\). In particular, the support of the conditional distribution of \(\beta\) varies with \(\phi\):

1. For \(\phi \in [\underline{\phi}, \min\{\theta / \underline{\beta}, \theta / \overline{\beta}\}]\), the support of \(\beta\) is \([\theta / \phi, \overline{\beta}]\). In other words: the range of \(\beta\) types that is consistent with \(\phi\) is increasing in \(\phi\), and this range always includes \(\overline{\beta}\). By the same token, the range of \(\theta\) types that is consistent with \(\phi\) is increasing in \(\phi\), and this range always includes \(\underline{\theta}\).

2. For \(\phi \in [\max\{\theta / \underline{\beta}, \theta / \overline{\beta}\}, \overline{\phi}]\), the support of \(\beta\) is \([\underline{\beta}, \theta / \phi]\). In other words: the range of \(\beta\) types that is consistent with \(\phi\) is decreasing in \(\phi\), and this range always includes \(\underline{\beta}\).

3. If \(\theta / \underline{\beta} < \overline{\theta} / \overline{\beta}\) then, for \(\phi \in [\min\{\theta / \underline{\beta}, \theta / \overline{\beta}\}, \max\{\theta / \underline{\beta}, \theta / \overline{\beta}\}]\), the support of \(\beta\) is \([\underline{\beta}, \overline{\beta}]\). In other words, if the range of \(\theta\) types is large relative to the range of \(\beta\) types, then all \(\beta\) types are consistent with intermediate values of \(\phi\).

4. If \(\theta / \overline{\beta} > \overline{\theta} / \overline{\beta}\) then, for \(\phi \in [\min\{\theta / \underline{\beta}, \theta / \overline{\beta}\}, \max\{\theta / \underline{\beta}, \theta / \overline{\beta}\}]\), the support of \(\beta\) is \([\theta / \phi, \theta / \phi]\). In other words, if the range of \(\theta\) types is small relative to the range of \(\beta\) types, then there is no value of \(\phi\) for which all \(\beta\) types are consistent with that value.

Armed with \(b\) and \(h\), the integral defining welfare and the integral giving the resource
constraint can be expressed

\[
\int \left( b(\phi) \phi v_1(\phi) + v_2(\phi) \right) h(\phi) \, d\phi
\]

(W)

and

\[
\int \left( Y - C_1(v_1(\phi)) - \frac{1}{R} C_2(v_2(\phi)) \right) h(\phi) \, d\phi \geq 0.
\]

(R)

We have therefore completed the transformation of our initial two-dimensional problem into a purely one-dimensional problem.

The Langrangian for the one-dimensional problem can be written

\[
\int \left( b(\phi) \phi v_1(\phi) + v_2(\phi) \right) h(\phi) \, d\phi \\
+ \lambda \int \left( Y - C_1(v_1(\phi)) - \frac{1}{R} C_2(v_2(\phi)) \right) h(\phi) \, d\phi \\
- \int (\phi v'_1(\phi) + v'_2(\phi)) \mu(\phi) h(\phi) \, d\phi \\
- \int v'_2(\phi) \nu(\phi) h(\phi) \, d\phi,
\]

where the Lagrange multipliers on the resource constraint, the incentive-compatibility constraint (ICL) and the incentive-compatibility constraint (ICM) take the form \( \lambda \in \mathbb{R}, \mu : \Phi \to \mathbb{R} \) and \( \nu : \Phi \to \mathbb{R} \).

### C.3 The First-Order Conditions

In order to derive first-order conditions from this Langrangian, we must first eliminate \( v'_1 \) and \( v'_2 \). We can do this by integrating by parts. Taking the third term of the Langrangian, we obtain

\[
- \int (\phi v'_1 + v'_2) \mu h \, d\phi = - \int (\phi v_1)' - v_1 + v'_2 \mu h \, d\phi \\
= \int v_1 \mu h \, d\phi - \int (\phi v_1)' + v'_2 \mu h \, d\phi,
\]

81
where we have dropped the dependence of \( v_1, v_2, \mu \) and \( h \) on \( \phi \). Moreover

\[
-\int ((\phi v_1') + v_2') \mu \, h \, d\phi \quad = \quad - \left[ (\phi v_1 + v_2) \mu \, h \frac{\partial \phi}{\partial \phi} \right] + \int ((\phi v_1) + v_2) (\mu h)' \, d\phi \\
= \int ((\phi v_1) + v_2) (\mu h)' \, d\phi
\]

(since \( h(\phi) = h(\phi) = 0 \)). Similarly, taking the fourth term,

\[
-\int v_2' \nu \, h \, d\phi \quad = \quad - [v_2 \nu h] \frac{\partial \phi}{\partial \phi} + \int v_2 (\nu h)' \, d\phi \\
= \int v_2 (\nu h)' \, d\phi.
\]

The Langrangian can therefore be written

\[
\int \left( \left( (b \phi + \mu) v_1 + v_2 + \lambda \left( Y - C_1(v_1) - \frac{1}{R} C_2(v_2) \right) \right) h \\
+ (\phi v_1 + v_2) (\mu h)' + v_2 (\nu h)' \right) \, d\phi.
\]

Differentiating the latter Langrangian with respect to \( v_1 \) and \( v_2 \), we obtain the first-order conditions

\[
0 = (b \phi + \mu - \lambda C_1'(v_1)) \quad h + \phi (\mu h)' \\
\]

and

\[
0 = \left( 1 - \frac{\lambda}{R} C_2'(v_2) \right) \quad h + (\mu h)' + (\nu h)'.
\]

We also have: (IC1), namely

\[
0 = \phi v_1' + v_2'
\]

the complementary slackness condition associated with the resource constraint, namely

\[
0 \leq \int \left( Y - C_1(v_1) - \frac{1}{R} C_2(v_2) \right) \, h \, d\phi \\
0 \leq \lambda
\]
and the complementary slackness condition associated with (IC2), namely

\[
\begin{align*}
0 &\leq -\nu'
\end{align*}
\]

\[
\begin{align*}
0 &\leq \nu
\end{align*}
\]

### C.4 The Relaxed Problem

We focus on the relaxed version of the problem, in which we do not impose (IC2). Furthermore, we look for a solution of the relaxed problem in which the resource constraint holds as an equality. We therefore drop \(\nu\) from the equations and tackle the three differential equations

\[
\begin{align*}
0 &= \left( b \phi + \mu - \lambda C_1'(v_1) \right) h + \phi (\mu h)', \\
0 &= \left( 1 - \lambda \frac{1}{R} C_2'(v_2) \right) h + (\mu h)', \\
0 &= \phi v_1' + v_2'
\end{align*}
\]

and the integral equation

\[
0 = \int \left( Y - C_1(v_1) - \frac{1}{R} C_2(v_2) \right) h \, d\phi.
\]

The first step is to make \(v_1\) and \(v_2\) the subjects of equations (27) and (28). Putting \(U_1 = (C_1')^{-1}\) and \(U_2 = (C_2')^{-1}\), we obtain

\[
\begin{align*}
v_1 &= U_1 \left( \frac{a_1}{\lambda} \right), \\
v_2 &= U_2 \left( \frac{a_2}{\lambda} \right),
\end{align*}
\]
where

\[
\begin{align*}
a_1 &= b \phi + \mu + \frac{\phi (\mu h)'}{h}, \\
a_2 &= R \left(1 + \frac{(\mu h)'}{h}\right).
\end{align*}
\]

(33)  
(34)

C.5 Solving (27-29) where \( b \) and \( h \) are Smooth

Consider the equations (27-29) in the open region \( \Phi = \Phi \setminus \{ \phi, \theta / \beta, \bar{\theta} / \bar{\beta}, \bar{\varphi} \} \). In this region, both \( b \) and \( h \) are smooth. Hence we may differentiate (31,32) to obtain

\[
\begin{align*}
v'_1 &= U'_1 \left( \frac{a_1}{\lambda} \right) \frac{a'_1}{\lambda}, \\
v'_2 &= U'_2 \left( \frac{a_2}{\lambda} \right) \frac{a'_2}{\lambda}
\end{align*}
\]

(35)  
(36)

and, substituting (35,36) in (29),

\[
0 = \phi U'_1 \left( \frac{a_1}{\lambda} \right) \frac{a'_1}{\lambda} + U'_2 \left( \frac{a_2}{\lambda} \right) \frac{a'_2}{\lambda}.
\]

Next, provided that \( u_1 \) and \( u_2 \) have the same coefficient of relative risk aversion \( \gamma \), the latter equation is homogeneous in \( \lambda \). It therefore simplifies further to

\[
0 = \phi U'_1(a_1) a'_1 + U'_2(a_2) a'_2.
\]

(If \( u_1 \) and \( u_2 \) have coefficient of relative risk aversion \( \gamma \), then \( U'_1(x) = U'_2(x) = \frac{1}{\gamma} x^{1-\gamma} \).)
Next, substituting for \( a'_1 \) and \( a'_2 \) and collecting terms in \( \mu'' \), \( \mu' \) and \( \mu \), we obtain

\[
0 = \left( \phi^2 U'_1 (a_1) + R U'_2 (a_2) \right) h^2 \mu'' \\
+ (\phi (\phi h' + 2 h) U'_1 (a_1) + h' R U'_2 (a_2)) h \mu' \\
+ (\phi (h (\phi h'' + h') - \phi h'^2) U'_1 (a_1) + (h h'' - h'^2) R U'_2 (a_2)) \mu \\
+ \phi (\phi b' + b) U'_1 (a_1) h^2. 
\] (37)

In other words, in the region \( \Phi \), equations (27-29) reduce to a second-order ordinary differential equation for \( \mu \).

### C.6 Solving (27-29) where \( b \) and \( h \) have Kinks

Now consider the equations (27-29) at the points \( \phi_1 = \frac{\theta}{\beta} \) and \( \phi_2 = \frac{T}{\beta} \), where both \( b \) and \( h \) have kinks. We cannot differentiate (31,32) at these points. However, we do have

\[
\Delta v_1(\phi_i) = U_1\left(\frac{a_1(\phi_i+)}{\lambda}\right) - U_1\left(\frac{a_1(\phi_i-)}{\lambda}\right),
\]

\[
\Delta v_2(\phi_i) = U_2\left(\frac{a_2(\phi_i+)}{\lambda}\right) - U_2\left(\frac{a_2(\phi_i-)}{\lambda}\right),
\]

where

\[
a_1(\phi_i+) = b \phi_i + \mu(\phi_i+) + \frac{\phi (\mu h)'(\phi_i+)}{h(\phi_i)},
\]

\[
a_1(\phi_i-) = b \phi_i + \mu(\phi_i-) + \frac{\phi (\mu h)'(\phi_i-)}{h(\phi_i)},
\]

\[
a_2(\phi_i+) = R \left( 1 + \frac{(\mu h)'(\phi_i+)}{h(\phi_i)} \right),
\]

\[
a_2(\phi_i-) = R \left( 1 + \frac{(\mu h)'(\phi_i-)}{h(\phi_i)} \right).
\]

85
Hence, at $\phi$, we can impose the value-matching condition

$$0 = \Delta \mu(\phi) = \mu(\phi+) - \mu(\phi-)$$

(38)

and the incentive condition

$$0 = \phi_i (U_1(a_1(\phi_i+)) - U_1(a_1(\phi_i-))) + (U_2(a_2(\phi_i+)) - U_2(a_2(\phi_i-))) .$$

(39)

C.7 Solving (27-29) at the Endpoints

Assuming for concreteness that $\phi_1 < \phi_2$, we now have the second-order ordinary differential equation (37) in the three open intervals $(\phi, \phi_1)$, $(\phi_1, \phi_2)$ and $(\phi_2, \phi)$. Moreover, we have two boundary conditions at each of $\phi_1$ and $\phi_2$. (Cf. (38) and (39).) The obvious way of completing the equation would therefore be to require that $\mu$ take on appropriate values at the boundaries $\phi$ and $\bar{\phi}$. However, $h$ decays linearly to 0 at both $\phi$ and $\bar{\phi}$. Moreover, inspection of (37) shows that:

1. the coefficient of $\mu''$ is positive and of order $h^2$ near $\phi$ and $\bar{\phi}$;
2. the coefficient of $\mu'$ is positive and of order $h$ near $\phi$, and negative and of order $h$ near $\bar{\phi}$;
3. the coefficient of $\mu$ is negative and of order 1 near $\phi$ and $\bar{\phi}$.

Hence $\mu$ will not take on boundary values at $\phi$ and $\bar{\phi}$ in the usual way.\(^{41}\) On the other hand, the inhomogeneous term, namely

$$\phi (\phi b' + b) U_1'(a_1) h^2,$$

\(^{41}\)Intuitively speaking, the dynamics of $\phi$ move away from the endpoints $\phi$ and $\bar{\phi}$.\)
is of order $h^2$ near $\phi$ and $\phi$. In particular, it is bounded. Hence the relevant solution of the equation is the one that is bounded near $\phi$ and $\phi$.\textsuperscript{42}

C.8 Solving for $\lambda$

As we have seen, we can find $\mu$ by solving the second-order o.d.e. (37) with the required boundary conditions at the internal boundaries $\phi_1$ and $\phi_2$ and the required boundedness properties at the endpoints $\phi$ and $\phi$. Like $b$ and $h$, $\mu$ can be expected to have kinks at $\phi_1$ and $\phi_2$. The next step is to solve for $\lambda$. This can be done using the resource equation (30).

Indeed, if $u_1$ and $u_2$ have the same coefficient of relative risk aversion $\gamma$, then we have

$$C_i(v_i) = C_i\left(U_i\left(\frac{a_i}{\lambda}\right)\right) = \left(\frac{a_i}{\lambda}\right)^{\frac{1}{\gamma}}.$$ 

Hence, substituting in (30),

$$0 = \int \left( Y - \left(\frac{a_1}{\lambda}\right)^{\frac{1}{\gamma}} - \frac{1}{R} \left(\frac{a_2}{\lambda}\right)^{\frac{1}{\gamma}} \right) h d\phi = \lambda^{-\frac{1}{\gamma}} \int \left( \lambda^{\frac{1}{\gamma}} Y - \frac{1}{R} \left(\frac{a_1}{\lambda}\right)^{\frac{1}{\gamma}} - \frac{1}{R} \left(\frac{a_2}{\lambda}\right)^{\frac{1}{\gamma}} \right) h d\phi,$$

or

$$\lambda^{\frac{1}{\gamma}} = \frac{\int \left( a_1^{\frac{1}{\gamma}} + \frac{1}{R} a_2^{\frac{1}{\gamma}} \right) h d\phi}{\int Y h d\phi}.$$

Bearing in mind that $a_1$ and $a_2$ are given in terms of $\mu$ by equations (33) and (34), this gives us a formula for $\lambda$ in terms of $\mu$.

\textsuperscript{42}Since the inhomogeneous term is of order $h^2$ near $\phi$ and $\phi$, the solution can in fact be expected to decay quadratically to 0 at both $\phi$ and $\phi$. In particular, we would expect that it would satisfy $\mu(\phi) = \mu'(\phi) = 0$ and $\mu(\phi) = \mu'(\phi) = 0$. These equations cannot, however, be used as boundary conditions. For one thing, there are too many of them! (There are 4 instead of 2.) They are simply additional properties that we would expect the unique bounded solution to possess.
C.9 Completing the Solution

It is then a straightforward matter to find the remaining unknowns in the model: \( v_1 \) and \( v_2 \) are given in terms of \( \lambda \) and \( \mu \) by (31) and (32); and \( c_1 \) and \( c_2 \) are given in terms of \( v_1 \) and \( v_2 \) by the formulae \( c_1 = C_1(v_1) \) and \( c_2 = C_2(v_2) \).

C.10 Numerical implementation

We generate a numerical solution (using Matlab’s bvp4c function\(^{43}\)) for the second-order differential equation for \( \mu \) (equation 37) with the boundary conditions described in section C.7 of this appendix.

D Analysis of the Quasi-Linear Limit Case for a Population of Agents with Heterogeneous \( \beta \)

D.1 Introduction

In Subsections 4.1 and 5.1, we discuss the quasi-linear limit case of our model: i.e., the case in which the utility function in the second period is linear (i.e., \( u_2(c_2) = c_2 \)). In this case, the planner’s problem can be written

\[
\max \int \left( \theta u_1(c_1) + u_2(c_2) \right) dF(\theta) dG(\beta) = \max \int \left( \theta u_1(c_1) + c_2 \right) dF(\theta) dG(\beta),
\]

subject to

\[
\int \left( c_1 + c_2 \right) dF(\theta) dG(\beta) = Y,
\]

\[
\phi \in \arg \max_{\phi' \in \Phi} \{ \phi u_1(c_1(\phi')) + u_2(c_2(\phi')) \} \quad \text{(IC)}
\]

for \( \phi \equiv \theta / \beta \).

\(^{43}\)See https://www.mathworks.com/help/matlab/ref/bvp4c.html
We study equilibria that satisfy the revelation principle, and, following the literature, refer to these as direct mechanisms. When we talk about $\phi$, we refer to the true value of $\phi$ elicited from each agent in an equilibrium that satisfies the revelation principle.

We now turn to proving Theorem 4.

**D.2 Proof of Theorem 4**

**D.2.1 Implementability**

Given the representation of the problem in the space of $\phi$, we now effectively have a single-type mechanism-design problem. We begin by transforming the problem into the promised utility space, $v_1(\phi) = u_1(c_1(\phi))$ and $v_2(\phi) = u_2(c_2(\phi)) = c_2(\phi)$. We invoke the standard equivalence between global incentive compatibility and the combination of integral incentive compatibility and monotonicity. Monotonicity implies $v'_1(\phi) \geq 0$, and in the standard way we solve the relaxed mechanism (not subject to monotonicity) and verify that the solution satisfies monotonicity.

Integral incentive compatibility is the standard condition, derived from the Envelope Theorem. In particular, the Envelope Theorem implies $\frac{d}{d\phi} (\phi v_1(\phi) + v_2(\phi)) = v_1(\phi)$, and we obtain integral incentive compatibility by integrating:

$$\phi v_1(\phi) + v_2(\phi) = \frac{\phi}{\phi} v_1(\phi) + v_2(\phi) + \int_{\phi} \phi v_1(\zeta) d\zeta.$$

We then use integral incentive compatibility to define the function $v_2$ in terms of the function $v_1$ and the constant $v_2(\phi)$, which gives us the implementing function $v_2$ that guarantees integral incentive compatibility:

$$v_2(\phi) = \phi v_1(\phi) + v_2(\phi) + \int_{\phi} v_1(\zeta) d\zeta - \phi v_1(\phi).$$

We then characterize $v_2(\phi)$ from $v_1$ using the resource constraint. Rewriting the resource
constraint over promised utility in the \( \phi \) space:

\[
\int \left( u_1^{-1}(v_1(\phi)) + v_2(\phi) \right) dH(\phi) = Y.
\]

Rearranging:

\[
\int v_2(\phi) dH(\phi) = Y - \int u_1^{-1}(v_1(\phi)) dH(\phi).
\]

Or, in other words, given a specification of a function \( v_1 \), we can use this condition plus the implementability condition to pin down \( v_2 \). In other words, if we substitute in the implementability condition for \( v_2 \), we get an equation for \( v_2(\phi) \) in terms of \( v_1 \):

\[
v_2(\phi) = Y - \int u_1^{-1}(v_1(\phi)) dH(\phi) - \bar{\phi} v_1(\phi) - \int \left( \int_\phi v_1(\zeta) d\zeta - \phi v_1(\phi) \right) dH(\phi).
\]

D.2.2 Completing the Model

Lastly, let us rewrite the objective function in terms of \( \phi \) and \( v_1 \). The contribution of type-\( \phi \) agents to social welfare is \( E[\theta | \phi] v_1(\phi) + v_2(\phi) \). Therefore, the planner objective function is:

\[
\int \left( E[\theta | \phi] v_1(\phi) + v_2(\phi) \right) dH(\phi).
\]

Substituting in the characterization of \( v_2 \) above, we get:

\[
\max_{v_1} \left\{ \int \left( E[\theta | \phi] v_1(\phi) - u_1^{-1}(v_1(\phi)) \right) dH(\phi) + Y \right\} \quad \text{s.t. (Monotonicity)}.
\]

That is, the planner chooses a non-decreasing function \( v_1 \), with the implementability conditions above defining the function \( v_2 \) that implements this outcome.

From here, we solve the relaxed problem, not subject to monotonicity. The relaxed problem is simply given by

\[
\max_{v_1} \left\{ \int \left( E[\theta | \phi] v_1(\phi) - u_1^{-1}(v_1(\phi)) \right) dH(\phi) + Y \right\}
\]
and so has a solution given by the first order condition for optimal allocation

\[ E[\theta | \phi] u'_1(c_1(\phi)) = 1. \]

From here, all that remains is to verify that this allocation satisfies monotonicity. Monotonicity arises provided that \( E[\theta | \phi] \) is non-decreasing. Hence, provided \( E[\theta | \phi] \) is non-decreasing, we have characterized the optimal allocation.

D.2.3 The Optimal Penalty

Consider the implied marginal penalty \( \pi(\phi) \) that implements the above allocation rule. The marginal trade-off of a private agent is then:

\[ (1 - \pi(\phi)) \phi u'_1(c_1(\phi)) = 1. \]

Therefore, the marginal penalty is:

\[ 1 - \pi(\phi) = \frac{E[\theta | \phi]}{\phi} = E \left[ \frac{\theta}{\phi} | \phi \right] = E[\beta | \phi]. \]

D.2.4 Homogeneous \( \beta \)

If \( \beta \) is homogeneous, then \( E[\beta | \phi] = \beta \), and we have:

\[ \pi(\phi) = 1 - \beta. \]

That is, we simply have a Pigouvian tax. This gives another proof of Theorem 3.
D.2.5 Heterogeneous $\beta$

If $\beta$ is heterogeneous and the regularity condition of Theorem 1 is satisfied, then as mentioned before we have:

$$\pi(\phi) = 1 - \mathbb{E}[\beta | \phi].$$

That is, we have an “average Pigouvian tax”: the optimal tax rate on the margin for a type-$\phi$ agent is the average tax rate in that population.

We know that $\pi(\phi)$ must be close to $1 - \beta$ near $\phi$, where the highest $\beta$ types are the only ones with that $\phi$ type. Similarly, we know that $\pi(\phi) \approx 1 - \beta$ near $\overline{\phi}$. This suggests a large degree of flexibility over initial withdrawals, and much tighter restrictions on flexibility for households withdrawing a lot.