

# Commitment-Flexibility Trade-off and Withdrawal Penalties

Attila Ambrus  
Duke University

Georgy Egorov  
Northwestern University

March 2012

ERID Working Paper Number 130

This paper can be downloaded without charge from the  
Social Science Research Network Electronic Paper Collection:

<http://ssrn.com/abstract=2070094>

**Economic Research Initiatives at Duke  
WORKING PAPERS SERIES**



# Commitment-Flexibility Trade-off and Withdrawal Penalties\*

Attila Ambrus<sup>†</sup>

Duke University, Department of Economics

Georgy Egorov<sup>‡</sup>

Northwestern University, Kellogg-MEDS

March 2012

## Abstract

Withdrawal penalties are common features of time deposit contracts offered by commercial banks, as well as individual retirement accounts and employer-sponsored plans. Moreover, there is a significant amount of early withdrawals from these accounts, despite the associated penalties, and empirical evidence shows that liquidity shocks of depositors are a major driving force of this. Using the consumption-savings model proposed by Amador, Werning and Angeletos in their 2006 *Econometrica* paper (henceforth AWA), in which individuals face the trade-off between flexibility and commitment, we show that withdrawal penalties can be part of the optimal contract, despite involving money-burning from an ex ante perspective. For the case of two states (which we interpret as “normal times” and a “negative liquidity shock”), we provide a full characterization of the optimal contract, and show that within the parameter region where the first best is unattainable, the likelihood that withdrawal penalties are part of the optimal contract is decreasing in the probability of a negative liquidity shock, increasing in the severity of the shock, and it is nonmonotonic in the magnitude of present bias. We also show that contracts with the same qualitative feature (withdrawal penalties for high types) arise in continuous state spaces, too. Our conclusions differ from AWA because the analysis in the latter implicitly assumes that the optimal contract is interior (the amount withdrawn from the savings account is strictly positive in each period in every state). We show that for any utility function consistent with their framework there is an open set of parameter values for which the optimal contract is a corner solution, inducing money burning in some states.

**Keywords:** Commitment, flexibility, self-control, money-burning

**JEL Classification:** D23, D82, D86.

---

\*We would like to thank Manuel Amador for useful suggestions, and Alexander Groves for valuable research assistance.

<sup>†</sup>E-mail: aa231@duke.edu

<sup>‡</sup>E-mail: g-egorov@kellogg.northwestern.edu

# 1 Introduction

Retail deposits constitute a large fraction of the assets of commercial banks (see Gilkeson et al. (1999)). Their distinctive feature is that a withdrawal option is bundled with each account. In particular, time deposits allow early withdrawal prior to stated maturity, but assess an early withdrawal penalty. Similarly, tax-deferred accounts, such as individual retirement accounts and employer-sponsored plans, which became important components of individual retirement planning, are subject to a 10% early withdrawal penalty. Empirical work on the subject shows that early withdrawal is at an economically significant level, even though it cannot be justified by profitable reinvestment motives (Gilkeson et al. (1999)), and finds direct evidence that it is connected to liquidity shocks of depositors (Amromin and Smith (2003)). Negative liquidity shocks such as job loss, divorce, and home purchases increase the likelihood of early withdrawals from employer-sponsored plans by an average of 3-10 percentage points each, with significantly stronger increases among the poorest households.

While early withdrawal penalties are prevalent in practice, it is not easy to justify them in a contract theory setting.<sup>1</sup> Even in the case of time-inconsistent consumers, early withdrawal penalties are essentially money burning schemes from the point of view of a consumer, making them a costly incentive device to discipline savings behavior. Along these lines, in an important paper AWA studied the optimal savings rule in a model where people are tempted to consume earlier, along the line of Strotz (1956), Phelps and Pollack (1968) and Laibson (1997, 1998),<sup>2</sup> but full commitment is undesirable as it does not allow for incorporation of new information, such as taste shocks and income shocks. They provided an optimal rule in two broad situations: if the shock variable can only take two values, and if the shock variable is continuous but a simple regularity condition on the density holds (essentially, the density of the distribution of taste shocks—the desire to consume early—does not decrease too fast). In the latter continuous case, AWA find that the optimal rule takes a simple form: it either allows a type to choose her optimal allocation, or imposes a minimal savings rule. In particular, in the continuous setting the optimal contract bunches the high liquidity shock types at a level of savings that is less than what they would find optimal. An important feature of the optimum in the above characterization results is that there is no money burning (withdrawal penalty).<sup>3</sup>

---

<sup>1</sup>For a theoretical model in which withdrawal penalties are assumed, instead of derived as part of an optimal contract, see Gilkeson and Ruff (1996).

<sup>2</sup>See also Gul and Pesendorfer (2001) and Dekel, Lipman and Rustichini (2001) for axiomatic foundations for preferences that imply temptation by present consumption and relatedly demand for commitment.

<sup>3</sup>Analogously, Athey et al. (2004) and Athey et al. (2005) show in various contract theory settings that money burning is not part of the optimal contract. Ambrus and Egorov (2009), in a principal-agent setting different from the one in the current paper, characterize cases when money burning can be part of an optimal delegation

We revisit the model of AWA and first analyze the case of two possible taste shocks. We show that money-burning may be used in equilibrium, imposed on the impatient type, in order to provide incentives for the more patient type not to imitate the impatient type. This is in contrast with Proposition 1 in AWA. The reason is that in the arguments in AWA implicitly assume that the optimal contract involves allocating strictly positive amounts of the good to be consumed at both time periods, in every state. However, we show that for every utility function in the framework they propose, there is an open set of parameter values for which the optimal contract involves 0 consumption in the second time period in case of a negative liquidity shock in the first time period. This can even be the case when the marginal utility of consumption in the second period is infinity at consumption level 0, that is when every consumer type would prefer to choose an interior consumption plan. This is because pushing consumption in the second period in a high liquidity shock state can relax the incentive constraints for the consumption allocation in the low state, and the resulting ex ante utility gain in the low state can exceed the utility loss in the high state.

There is a natural way to modify the framework in AWA that allows for utility functions that indeed guarantee that the optimal contract always specifies an interior consumption plan, and hence the analysis in AWA is valid: adding utility functions that at consumption level 0 take a value of  $-\infty$ . Such utility functions would indeed be realistic in macroeconomic applications that motivated AWA, that is when a representative consumer is contracting with a social planner, as long as the whole income of the consumer is contractible so she does not have any other sources of consumption than withdrawals from the savings account. In micro finance applications such utility functions would not be realistic, as presumably no consumer would ever want to hold literally all her present and future wealth in one savings account. In any case, we append the AWA framework with utility functions like above, and conduct our analysis in the extended framework.

For utility functions in the original AWA framework, we show that money burning becomes part of the optimal contract when the probability of the impatient type is not too large, and when the negative liquidity shock is severe enough. The intuition behind the first result is that money burning becomes a realized loss in the high liquidity shock state, hence when the probability of this state is large it becomes too costly to use as an incentive device. The result holds because if the possible liquidity shock is severe then allowing the consumer to withdraw everything from the account, at the price of a withdrawal penalty, becomes a more efficient incentive device to deter the low type consumer to imitate the high type than specifying a consumption plan for the

---

scheme. See also Amador and Bagwell (2011).

high type with no money burning. In contrast to the above clear comparative statics results, we find that the presence of money burning is not necessarily monotonic in the magnitude of present bias. This is because increasing the degree of present bias has two opposite effects: it increases the conflict of interest between the ex ante and the period-1 ex post selves, which makes money burning more likely. On the other hand, more present bias implies that the ex post self is more difficult to discipline through money burning, which makes it less likely to be used in the optimal contract. In general there is an intermediate region of present bias parameters for which money burning is used.

To summarize, the scenarios in which withdrawal penalties are most likely to be optimal are those in which individuals face the possibility of a rare but severe negative liquidity shock.

We also show that money-burning by impatient types may be part of the optimal contract if there are more than two types, in particular if there is a continuum of types. This is not consistent with Proposition 2 from AWA, for exactly the same reasons as our results differ from theirs above. However, Proposition 2 becomes true when an additional assumption, that AWA introduces later, is imposed on the distribution function.

Our interpretation of these results is that early withdrawal penalties can be part of an optimal savings contract in a much broader set of environments than previously thought. Moreover, our results are consistent with the empirical finding that it is liquidity constrained consumers with more severe negative income shocks who exercise early withdrawal despite the associated penalty. On the other hand we do not think that our results diminish the importance of minimum savings rules, which AWA highlighted. Both of these simple types of savings contracts are common in practice, and which one is likely to be used depends on the given situation (the uncertainty and the degree of present bias that the consumer faces). In particular, in macroeconomic applications it can be realistic to restrict attention to specifications of the model that justify the implicit assumption in AWA that the optimal contract specifies an interior consumption plan.

We would also like to point out that the main results of AWA regarding minimum savings rules, Propositions 3 and 4, are not affected by our analysis, and are valid in their original framework, without imposing any extra assumptions.

## 2 The model

The setup reintroduces the model from AWA, and we preserve the notation. There are two periods and a single good. A consumer has a budget  $y$  and chooses his consumption in periods 1 and 2,  $c$  and  $k$ , respectively, so his budget set  $B$  is defined by  $c \geq 0$ ,  $k \geq 0$ ,  $c + k \leq y$  (the interest rate is normalized to 0). The utility of *self-0* (the individual before the consumption

periods) is given by:

$$\theta U(c) + W(k),$$

where  $U, W : \mathbb{R}^+ \rightarrow \mathbb{R} \cup \{-\infty\}$  are two strictly increasing, strictly<sup>4</sup> concave and continuously differentiable functions, and  $\theta \in \Theta$  is a taste shock which is realized in period 1. AWA assume that  $U, W$  map  $\mathbb{R}^+$  to  $\mathbb{R}$ , thus ruling out the possibility of, say,  $U(c) = \log c$ . We extend their framework as allowing for this does not complicate the analysis, and as we show leads to some new insights regarding the possibility of money-burning in optimum.

We assume that  $\Theta$  is bounded and normalized so that  $\mathbb{E}\theta = 1$ . Denote the c.d.f. of  $\theta$  by  $F(\cdot)$  and the p.d.f. of  $\theta$  by  $f(\cdot)$ . The utility of *self-1* is given by

$$\theta U(c) + \beta W(k),$$

where  $0 < \beta \leq 1$  captures the degree of agreement between self-0 and self-1 (and  $1 - \beta$  captures the strength of temptation towards earlier consumption). The goal is to characterize the optimal contract with self-0 as the principal and self-1 as the agent, i.e., the consumption scheme that self-0 would choose from behind the veil of ignorance about the realization of the taste shock  $\theta$ .

The optimization problem of the principal is:

$$\max_{(c(\theta), k(\theta))_{\theta \in \Theta}} \int_{\theta \in \Theta} (\theta U(c(\theta)) + W(k(\theta))) dF(\theta) \quad (1)$$

subject to

$$(c(\theta), k(\theta)) \in B \text{ for every } \theta \in \Theta, \quad (2)$$

$$\theta U(c(\theta)) + \beta W(k(\theta)) \geq \theta U(c(\theta')) + \beta W(k(\theta')) \text{ for every } \theta, \theta' \in \Theta. \quad (3)$$

Hereinafter, we find it convenient to characterize contracts in terms of utilities rather than allocations (each is a monotone transformation of the other). We let  $C(u)$  and  $K(w)$  be the inverse functions of  $U(c)$  and  $W(k)$ , respectively, and we let set  $A$  be given by

$$A = \{(u, w) \in \mathbb{R}^2 : u \geq U(0), w \geq W(0), C(u) + K(w) \leq y\}.$$

Since  $C(u)$  and  $K(w)$  are strictly convex functions, the set  $A$  is convex. Let us also introduce function  $z(\cdot)$  by

$$z(x) = W(y - C(x));$$

then  $z(\cdot)$  is decreasing and strictly concave. The set  $\{u, w : w = z(u)\}$  is the frontier of the set  $A$  where there is no money-burning:  $C(u) + K(w) = y$ .

---

<sup>4</sup>Assuming strict concavity rules out linear utility functions, but simplifies characterization a lot. Clearly, any linear function may be approximated by strictly concave functions, so the results may be applied to characterize the properties of optimal contracts with linear utility functions as well.

In the terms  $u$  and  $w$ , the first-best utilities of the ex ante self,  $(u^{fb}(\theta), w^{fb}(\theta)) = (U(c^{fb}(\theta)), W(k^{fb}(\theta)))$ , solve

$$(u^{fb}(\theta), w^{fb}(\theta)) = \arg \max_{(u,w) \in A} (\theta u + w),$$

and utilities for the full flexibility benchmark (where the period-1 self is free to make the choice),  $(u^f(\theta), w^f(\theta)) = (U(c^f(\theta)), W(k^f(\theta)))$  solve

$$(u^f(\theta), w^f(\theta)) = \arg \max_{(u,w) \in A} (\theta u + \beta w).$$

Finally, the rewritten optimization problem of the principal is:

$$\max_{(u(\theta), w(\theta))_{\theta \in \Theta}} \int_{\theta \in \Theta} (\theta u(\theta) + w(\theta)) dF(\theta) \quad (4)$$

subject to

$$(u(\theta), w(\theta)) \in A \text{ for every } \theta \in \Theta, \quad (5)$$

$$\theta u(\theta) + \beta w(\theta) \geq \theta u(\theta') + \beta w(\theta') \text{ for every } \theta, \theta' \in \Theta. \quad (6)$$

### 3 Two types

Here we consider the case of two types, so that  $\Theta = \{\theta_l, \theta_h\}$  with  $0 < \theta_l < \theta_h$  (and given the normalization  $E\theta = 1$ ,  $\theta_l < 1 < \theta_h$ ). This setup can be interpreted such that state  $\theta_l$  represents “normal times”, while state  $\theta_h$  represents a negative liquidity shock, such as a job loss.

If we denote the probability that  $\theta = \theta_l$  by  $\mu$ , we must have

$$\mu\theta_l + (1 - \mu)\theta_h = 1. \quad (7)$$

We are thus solving the problem

$$\max_{(u_l, w_l), (u_h, w_h) \in A} (\mu(\theta_l u_l + w_l) + (1 - \mu)(\theta_h u_h + w_h)) \quad (8)$$

subject to

$$\theta_l u_l + \beta w_l \geq \theta_l u_h + \beta w_h, \quad (9)$$

$$\theta_h u_h + \beta w_h \geq \theta_h u_l + \beta w_l. \quad (10)$$

Throughout this section, we use subscripts  $l$  and  $h$  to denote the values at  $\theta_l$  and  $\theta_h$ , respectively, e.g.,  $u_l \equiv u(\theta_l)$ , etc.

AWA, as part of Proposition 1 in this paper, characterizes the parameter regions in which (i) the optimal contract achieves the first best; (ii) does not achieve the first best but implies

separation of the two types; and (iii) implies pooling of the two types. Parts of this proof relied on an argument in the proof that there is no money-burning in the optimal contract. Since below we show that this does not hold without additional assumptions, we provide the complete proof of this result in the Appendix, even though this part of Proposition 1 of AWA is correct as stated.

**Proposition 1** *Suppose  $\Theta = \{\theta_l, \theta_h\}$  with  $\theta_l < \theta_h$ . Suppose that  $\theta_l < \left| \frac{dz}{du} \Big|_{u=U(y)} \right|$  and  $\theta_h > \left| \frac{dz}{du} \Big|_{u=U(0)} \right|$ .<sup>5</sup> Then there exists  $\beta^* \in (\theta_l/\theta_h, 1)$  such that for  $\beta \in [\beta^*, 1]$  the first-best allocation is implementable.*

*If  $\beta \leq \theta_l/\theta_h$ , then pooling is optimal, i.e.,  $u_h = u_l$  and  $w_h = w_l$ ; moreover, there is no money-burning in this case:  $w_l = z(u_l)$ .*

*If, however,  $\beta \in \left( \frac{\theta_l}{\theta_h}, \beta^* \right)$ , then separation is optimal, i.e.,  $u_h > u_l$  and  $w_h < w_l$ . In this last case,  $w_l = z(u_l)$ , but both  $w_h = z(u_h)$  and  $w_h < z(u_h)$  is possible. In either case, the IC constraint of the low type (9) is binding and the IC constraint of the high type (10) is not.*

Proposition 1 of AWA also claims that money burning is never part of the optimal contract, which, as we find, does not have to hold in general. Our next result below gives a necessary and sufficient condition for money burning to be part of the optimal contract. The proof of Proposition 1 in AWA is invalid without further assumptions at the point where the authors write “Then an increase in  $c(\theta_h)$  and a decrease in  $k(\theta_h)$  that holds  $(\theta_l/\beta)U(c(\theta_h)) + U(k(\theta_h))$  unchanged...”, which implicitly assumes that a decrease in  $k_h = k(\theta_h)$  is possible. If  $k_h = 0$ , so type  $\theta_h$  consumes only in period 1, then such a decrease is clearly impossible, and no contradiction is obtained. We prove that this is the only possible case consistent with money burning (i.e., money-burning implies  $c_h < y$ ,  $k_h = 0$ ), and it is only possible if  $W(0) \neq -\infty$  (Proposition 2). In fact, if  $k_h > 0$  in the optimal contract then the argument in AWA goes through, ruling out the possibility of money burning.

As a prelude to the next result, the following figures illustrate the two types of separating contracts that are possible in optimum. Note that if the IC constraint is binding for the low type then the line connecting  $(u_l, w_l)$  and  $(u_h, w_h)$  has to have a slope of  $-\theta_l/\beta_l$ . Below we refer to this line as the  $IC_l$  line. Figure 1 represents a possibility such that at the optimum the  $IC_l$  line intersects set  $A$  twice at the Pareto frontier. This corresponds to a separating equilibrium with no money burning, as in AWA. Figure 2 represents a different possibility, when at the optimum

<sup>5</sup>This requirement ensures that the first best contract is not pooling, and should have been included in Proposition 1 of AWA as well. If  $\theta_l \geq \left| \frac{dz}{du} \Big|_{u=U(y)} \right|$ , then the optimal contract is  $c_l^{fb} = c_h^{fb} = y$ ,  $k_l^{fb} = k_h^{fb} = 0$ , and if  $\theta_h \geq \left| \frac{dz}{du} \Big|_{u=U(0)} \right|$ , then the optimal contract is  $c_l^{fb} = c_h^{fb} = 0$ ,  $k_l^{fb} = k_h^{fb} = y$ . In either of these cases, the first best is implementable for all  $\beta$ . (If  $z(u)$  does not have a left derivative at  $u = U(0)$ , or  $U(0) = -\infty$ , then  $\left| \frac{dz}{du} \Big|_{u=U(0)} \right|$  is  $-\infty$ .)



the  $IC_l$  line crosses the horizontal boundary of set  $A$  (on the  $w = W(0)$  line), implying that there is money burning in equilibrium. Below we show that both of these cases can indeed occur at the optimum.

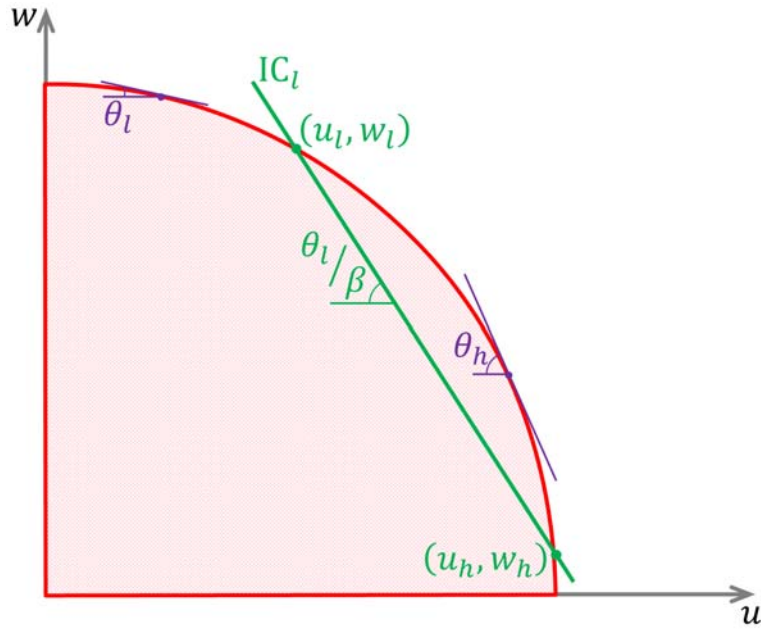


Figure 1: Optimal contract with no money burning.

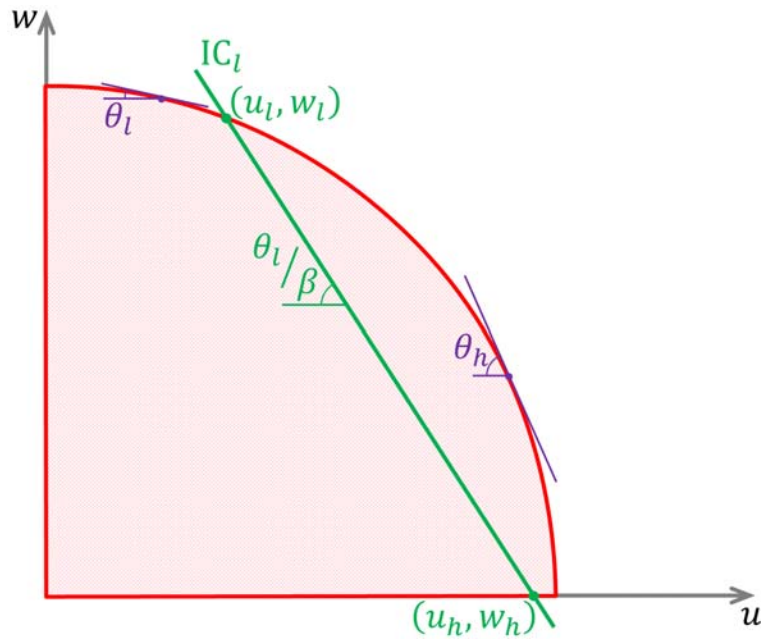


Figure 2: Optimal contract with money burning.

In order to give a precise characterization of when money-burning is part of a separating optimal contract, we need to introduce some further notation. Proposition 1 implies that the IC constraint of type  $\theta_l$  is binding; let us denote, for any  $K \in \mathbb{R}$ ,

$$\lambda^K = \left\{ (u, w) \in A : u + \frac{\beta}{\theta_l} w = K \right\}. \quad (11)$$

For any  $K$ , the above set of points is either a line segment, a point, or the empty set, although for simplicity we just refer to it as the  $IC_l$  line. Whenever  $\lambda^K \neq \emptyset$ , let  $\lambda_l^K = (u_l^K, w_l^K)$  and  $\lambda_h^K = (u_h^K, w_h^K)$  be the points of  $\lambda^K$  that minimize and maximize  $u$ , respectively. Fixing  $K = u_l + \frac{\beta}{\theta_l} w_l = u_h + \frac{\beta}{\theta_l} w_h$ , we observe that  $(u_l, w_l) = \lambda_l^K$  and  $(u_h, w_h) = \lambda_h^K$  (if it were not the case, then moving  $(u_l, w_l)$  north-west along the  $IC_l$  line would not violate (9) or (10) and would increase (8), as  $\theta_l < \frac{\theta_l}{\beta}$ , and moving  $(u_h, w_h)$  along the same line would have the same effect as  $\theta_h > \frac{\theta_l}{\beta}$ ). Let us now take a particular value of  $K$ ,

$$K_0 = U(y) + \frac{\beta}{\theta_l} W(0); \quad (12)$$

then  $K_0$  is finite if  $W(0) \neq -\infty$  and  $K_0 = -\infty$  otherwise. In the case  $K_0$  is finite, notice that  $\lambda_h^{K_0} = (U(y), W(0))$  by definition. The leftmost point of intersection of  $\lambda^{K_0}$  with  $A$ ,  $\lambda_l^{K_0}$ , plays a critical role in the following formulation, and we let  $u_0 \equiv u_l^{K_0}$ .

**Proposition 2** *Suppose  $\frac{\theta_l}{\theta_h} < \beta < \beta^*$ , so the optimal contract is separating. Money-burning will be used as part of the optimal contract if and only if (i)  $W(0) \neq -\infty$ , (ii)  $u_0 > U(0)$ , where  $u_0$  is defined as  $u_l^{K_0}$  for  $K_0 = U(y) + \frac{\beta}{\theta_l} W(0)$ , and (iii) for the following inequality holds:*

$$\mu \frac{1 - \beta}{\left| \frac{dz}{du} \Big|_{u=u_0} \right| - \frac{\beta}{\theta_l}} > 1. \quad (13)$$

While the formal proof is in the Appendix, we provide a summary of the reasoning behind the proof below. First, let us reformulate the problem (8) in terms of  $K$ , which remains the only degree of freedom:

$$\max_{K: \lambda^K \neq \emptyset} \left( \mu (\theta_l u_l^K + w_l^K) + (1 - \mu) (\theta_h u_h^K + w_h^K) \right); \quad (14)$$

the constraints (9) and (10) would then hold automatically.<sup>6</sup>

Next, we show that the maximand in (14) is strictly concave in  $K$ . The idea of the proof for this fact is illustrated in Figure 3. If we take two contracts  $(u_l^{K_1}, w_l^{K_1}, u_h^{K_1}, w_h^{K_1})$  and

---

<sup>6</sup>The IC constraint of type  $\theta_l$  (9) would hold as equality because  $(u_l^K, w_l^K)$  and  $(u_h^K, w_h^K)$  lie on the same  $\lambda^K$ , and the IC constraint of type  $\theta_h$  (10) would follow from the fact that (9) holds with equality and  $u_h^K \geq u_l^K$ .

$(u_l^{K_2}, w_l^{K_2}, u_h^{K_2}, w_h^{K_2})$  with ex-ante expected payoffs  $V^{K_1}$  and  $V^{K_2}$ , then a linear combination of these contracts with weights  $\delta \in (0, 1)$  and  $1 - \delta$ ,  $(u'_l, w'_l, u'_h, w'_h)$ , would yield the ex-ante payoff  $V' = \delta V^{K_1} + (1 - \delta) V^{K_2}$  due to linearity of (14). However, for  $K = \delta K_1 + (1 - \delta) K_2$ ,  $(u_l^K, w_l^K, u_h^K, w_h^K)$  yields an even higher utility  $V^K > V' = \delta V^{K_1} + (1 - \delta) V^{K_2}$ , which establishes the strict concavity.

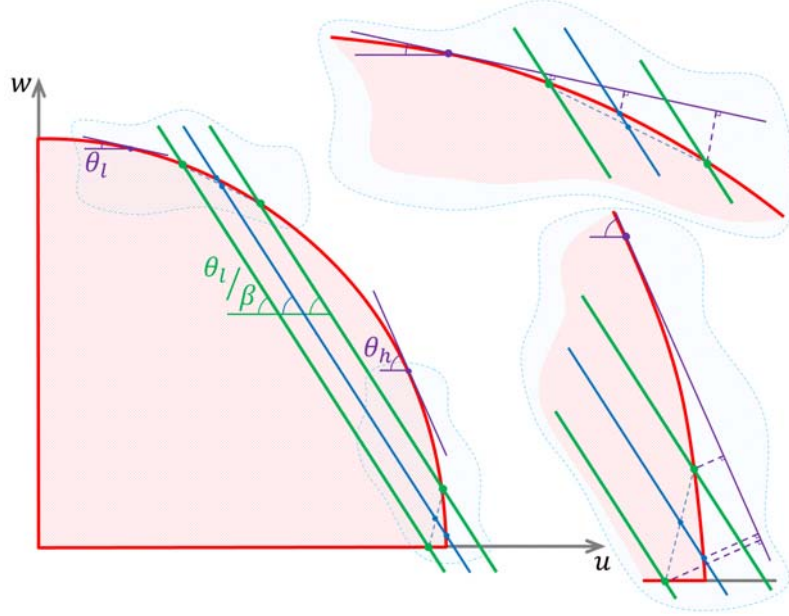


Figure 3: Argument for concavity of the maximand in (14) as a function of  $K$ .

Because the maximand in (14) is a single-peaked function of  $K$ , money-burning is optimal if and only if (14) increases as  $K$  decreases from  $K_0 = U(y) + \frac{\beta}{\theta_l} W(0)$ , which corresponds to the corner point on Figure 4. This gives us the derivation for the formula (13). We take  $\varepsilon$  small and let  $K_\varepsilon = K_0 - \varepsilon$ ; then  $u_h^{K_\varepsilon} = u_h^{K_0} - \varepsilon = U(y) - \varepsilon$  and  $w_h^{K_\varepsilon} = w_h^{K_0} = 0$ . Linearizing the function  $z(u)$  near the point  $u_0$ , we find that  $u_l^{K_\varepsilon} = u_l^{K_0} - \frac{\varepsilon}{\frac{\theta_l}{\beta} - |\frac{dz}{du}(\tilde{u}(\varepsilon))|}$  and  $w_l^{K_\varepsilon} = w_l^{K_0} + \frac{\varepsilon |\frac{dz}{du}(\tilde{u}(\varepsilon))|}{\frac{\theta_l}{\beta} - |\frac{dz}{du}(\tilde{u}(\varepsilon))|}$  for some  $\tilde{u}(\varepsilon) \in [u_l^{K_\varepsilon}, u_0]$ . Consequently, the total change of the expected utility of self-0 from this manipulation is  $\mu \left( -\theta_l \frac{\varepsilon}{1 - \frac{\beta}{\theta_l} |\frac{dz}{du}(\tilde{u}(\varepsilon))|} + \frac{\varepsilon |\frac{dz}{du}(\tilde{u}(\varepsilon))|}{1 - \frac{\beta}{\theta_l} |\frac{dz}{du}(\tilde{u}(\varepsilon))|} \right) + (1 - \mu) \left( \theta_h \left( -\varepsilon \frac{\beta}{\theta_l} \right) \right)$ . Rearranging, using (7), and observing that  $\frac{dz}{du}(\tilde{u}(\varepsilon)) \rightarrow \frac{dz}{du}(u_0)$  as  $\varepsilon \rightarrow 0$ , we get (13). These arguments are illustrated in Figure 4.

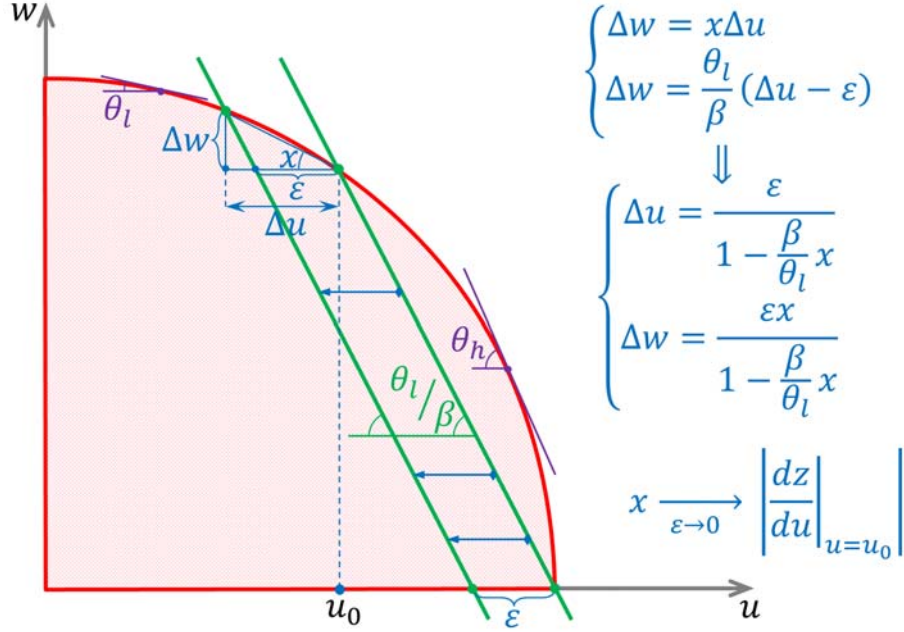


Figure 4: Derivation of the condition for money-burning (13).

The next corollary, which follows directly from the proof described above, further clarifies the set of cases where money-burning may be used.

**Corollary 1** *If the optimal contract requires money-burning, then self-1 with type  $\theta_l$  is impatient enough to prefer allocation  $(y, 0)$  to  $(c^{fb}(\theta_l), k^{fb}(\theta_l))$ , i.e.,*

$$\theta_l U(y) + \beta W(0) > \theta_l u_l^{fb} + \beta w_l^{fb}. \quad (15)$$

*In particular,  $W(0)$  must be finite, so  $W(k)$  must be bounded away from  $-\infty$ . Moreover, whenever the optimal contract requires money-burning, we must have  $k_h = 0$ .<sup>7</sup>*

Note that the case  $W(0) = -\infty$  is only realistic if the consumer literally keeps all her resources in the savings account, and has no other source of consumption, and this is unlikely to hold for savings accounts in practice. Putting it differently, it is reasonable in real life savings situations that self-1 is impatient enough ( $\beta$  is low) to want to withdraw all money from the account in period 1, if this option is feasible.

The next corollary studies what happens if money-burning is optimal, but the contract is exogenously constrained not to involve money-burning.

<sup>7</sup>If  $\beta = \frac{\theta_l}{\theta_h}$ , then the optimal contract is not uniquely defined, and among them there may be contracts with money-burning and  $k_h > 0$ . But in this case we can find an optimal contract without money burning.

**Corollary 2** Suppose  $\frac{\theta_l}{\theta_h} < \beta < \beta^*$  and money-burning is part of the optimal contract, i.e., (13) holds. Then the optimal contract under the additional constraint that money-burning is impossible ( $w_l = z(u_l)$ ,  $w_h = z(u_h)$ ) is given by  $u_h = U(y)$ ,  $w_h = W(0)$  (i.e.,  $c_h = y$ ,  $k_h = 0$ ),  $u_l = u_0$ ,  $w_l = z(u_0)$  (where  $u_0$  is as defined in Proposition 2).

This result is straightforward. As part of the proof of Proposition 2, we established that the maximand in (14) is strictly concave in  $K$ . If the optimum satisfied  $K < K_0$  (for  $K_0$  given by (12)) and we effectively add the constraint  $K \geq K_0$ , then the optimum would be at  $K_0$ . Consequently, allowing type  $\theta_h$  to consume everything in period 1 while keeping type  $\theta_l$  indifferent would be optimal under the constraint of no money-burning. Corollary 2 makes it easy to compute gains from allowing money-burning.

The following is an example in which money burning is part of the optimal contract.

**Example 1** Suppose  $U(c) = c^{0.7}$ ,  $W(k) = k^{0.7}$ ,  $y = 1$  (then  $z(u) = (1 - u^{0.7})^{0.7}$ ),  $\theta_l = 0.2$ ,  $\theta_h = 8.2$ ,  $\mu = 0.9$ ,  $\beta = 0.1$ . In this case,  $\beta^* = 0.1973$ , and  $\frac{\theta_l}{\theta_h} = 0.0244$ , so  $\beta \in \left(\frac{\theta_l}{\theta_h}, \beta^*\right)$  and a separation contract is optimal. Then  $u_0 = 0.7955$ , so  $\left|\frac{dz}{du}\right|_{u=u_0} = 1.33$ . Therefore,

$$\mu \frac{1 - \beta}{\left|\frac{dz}{du}(u_0)\right| - \frac{\beta}{\theta_l}} = 0.9 \frac{1 - 0.1}{1.33 - \frac{0.1}{8.2}} = 3.2158 > 1$$

The optimal contract is  $c_l = 0.2892$ ,  $k_l = 0.7108$ ,  $c_h = 0.7444$ ,  $k_h = 0$ , and indeed involves money burning. (The optimal contract with the constraint that money-burning is not allowed would be  $c_l = 0.7212$ ,  $k_l = 0.2788$ ,  $c_h = 1$ ,  $k_h = 0$ , and the ex-ante expected utilities in the two contracts are 1.4512 and 1.3312, respectively, with a difference of 0.1199.) For these parameter values the maximal  $\beta$  that would still result in an optimal contract with money burning is  $\beta_{mb} = 0.1597 > \beta$ .

Let us now take  $\beta = 0.16$ , which also satisfies  $\beta \in \left(\frac{\theta_l}{\theta_h}, \beta^*\right)$ , so separation is optimal. We then have  $u'_0 = 0.3065$  and  $\left|\frac{dz}{du}\right|_{u=u'_0} = 0.6405$ . Therefore

$$\mu \frac{1 - \beta}{\left|\frac{dz}{du}(u_0)\right| - \frac{\beta}{\theta_l}} = 0.9 \frac{1 - 0.16}{0.6405 - \frac{0.16}{8.2}} = -15.7106 < 1$$

Note that  $\beta' > \beta_{mb}$ . In this case money-burning is not part of the optimal contract:  $c_l = 0.1847$ ,  $k_l = 0.8153$ ,  $c_h = 1$ ,  $k_h = 0$ .<sup>8</sup>

Example 1, which shows that money-burning is possible, is not atypical. In particular, this has nothing to do with the choice of utility functions (except that we must have  $W(0) \neq -\infty$

<sup>8</sup>It is not a coincidence that the allocation for type  $\theta_h$  is exactly  $c_h = 1$ ,  $k_h = 0$ , i.e., at the corner. Clearly, at  $K_0$ , the maximand in (14) has a kink. Consequently, for a positive measure of parameter values this will be the optimal point.

to get money-burning). As a matter of fact, as long as the utility functions in both periods are the same, one can find an open set of parameter values (relative to the possible set of parameter values defined in the model) for which having money burning is optimal, i.e., (13) is satisfied. In other words, such situations are not knife-edge cases.

**Proposition 3** *Take any convex functions  $U(\cdot)$  and  $W(\cdot)$  such that the function  $z(u)$  has at least one point  $u_0 \in (0, y)$  with  $|\frac{dz}{du}|_{u=u_0}| \geq 1$  (this would be the case, for example, if  $W = U$ , or if  $W'(0) = \infty$  and  $W(0) \neq -\infty$ ). Then there exists an open set of parameter values  $\mu, \theta_l, \beta$  (with  $\theta_h$  found from (7)) such that the optimal contract necessarily includes money-burning.*

We can strengthen this result further if we put some constraints on the set  $A$ .

**Proposition 4** *If  $\beta \in (\frac{\theta_l}{\theta_h}, \beta^*)$  (so that the optimal contract is separating but not the first-best), and  $|\frac{dz}{du}|_{u=u_0}| \geq 1$ , then the optimal contract involves money-burning. In particular, if  $z(u)$  is such that  $|\frac{dz}{du}|_{u=U(0)}| \geq 1$  and  $|\frac{dz}{du}|_{u=U(y)}| \leq \theta_h$  (i.e.,  $|\frac{dz}{du}(u)| \in [1, \theta_h]$  for all  $u \in [U(0), U(y)]$ ), then for **every**  $\beta \in (\frac{\theta_l}{\theta_h}, \beta^*)$  the optimal contract involves money-burning.*

This proposition gives a sufficient condition for money-burning in the optimal contract: if the slope of the boundary of set  $A$  at the leftmost intersection with  $\lambda^{K_0}$  is at least 1, then there must be money-burning. The condition is not necessary: e.g., in Example 1,  $|\frac{dz}{du}|_{u=u_0}| = \frac{3}{4} < 1$ , yet there is money-burning. Intuitively, this condition corresponds to the cases where  $\beta$  is small (yet still greater than  $\frac{\theta_l}{\theta_h}$ ), so the preferences of self-0 and self-1 are sufficiently different. However, we do not get unambiguous comparative statics with respect to  $\beta$ . Indeed, there are two effects. On the one hand, (13) is increasing in  $\beta$  in the relevant region. This happens because the slope of the indifference line decreases, so every unit of money-burning for type  $\theta_h$  means a larger gain for type  $\theta_l$ , so money-burning is more profitable on the margin. However, as the slope decreases,  $u_0$  also decreases, and  $|\frac{dz}{du}|_{u=u_0}|$  decreases and becomes closer to  $\theta_l$ , which means that self-0 of type  $\theta_l$  becomes less sensitive to moving across the boundary of set  $A$ . Intuitively, on the one hand, a larger  $\beta$  makes self-1 more concerned about money-burning, and thus easier to discipline with just a little money-burning; on the other hand, it also makes preferences of self-0 and self-1 better aligned, so money-burning is not so important from the start. Interestingly, the second effect disappears as the boundary of set  $A$  becomes close (in  $C^1$ -metrics) to linear (i.e., if  $U(\cdot)$  and  $W(\cdot)$  become close to linear).

The second part of Proposition 4 deals with a special case where the slope of the boundary of set  $A$  lies between 1 and  $\theta_h$ . (For example, if  $U(\cdot)$  and  $W(\cdot)$  are identical linear functions, the slope would be 1.) In this situation, self-0 of type  $\theta_h$  would prefer to consume everything

in period 1 while self-0 of type  $\theta_l$  would prefer to consume everything in period 2. As we show above, such strong conflict of interest makes money-burning part of the optimal contract.

Next we examine comparative statics for the region where money burning is part of the optimal contract.

**Proposition 5** *Suppose that  $\mu, \theta_l, \theta_h, \beta$  are such that (7) holds and  $\beta \in \left(\frac{\theta_l}{\theta_h}, \beta^*\right)$  (so the optimal contract is separating but not the first best) and it involves no money burning. Then, for a fixed  $\theta_l$  and  $\beta$ , a decrease in  $\mu$  involves no money-burning either. For a fixed  $\frac{\theta_l}{\beta}$  and  $\mu$ , a higher  $\beta$  implies no money-burning.*

This means that within the parameter region which imply separation, but not the first-best, money-burning is more likely if the high state is sufficiently rare. Intuitively, if  $\mu$  is high enough, then committing to money burning in the state  $\theta_h$  does not affect the expected utility of self-0 too negatively. The second part of the result states that increasing  $\beta$  and  $\theta_l$  in a way that keeps their ratio constant makes money burning less likely to be part of the optimal contract. Note that increasing  $\theta_l$  for a fixed  $\mu$  implies decreasing  $\theta_h$ , hence the statement can be reworded such that decreasing the severity of the liquidity shock in the high state (keeping  $\frac{\theta_l}{\beta}$  constant) makes money burning less likely.

Figures 5-7 illustrate these comparative statics results, for utility functions  $U(\cdot) = W(\cdot) = sqr(\cdot)$ , by depicting how in the  $\theta_l$ - $\beta$  parameter space, for different levels of  $\mu$ , the region where the optimal contract is separating but not first best is divided into a region with money-burning and a region with no money burning. For  $\mu = 0.7$  the money-burning region is small, and it is concentrated on very low values of  $\theta_l$  (equivalently, very high levels of  $\theta_h$ ). As  $\mu$  increases (the probability of high state decreases), the money-burning subregion takes over a larger and larger part of the separating region.

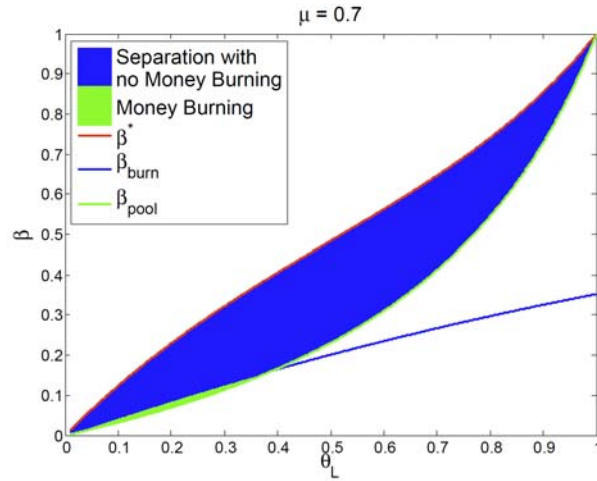


Figure 5: Money-burning and no money-burning separating regions for  $\mu = 0.7$ .

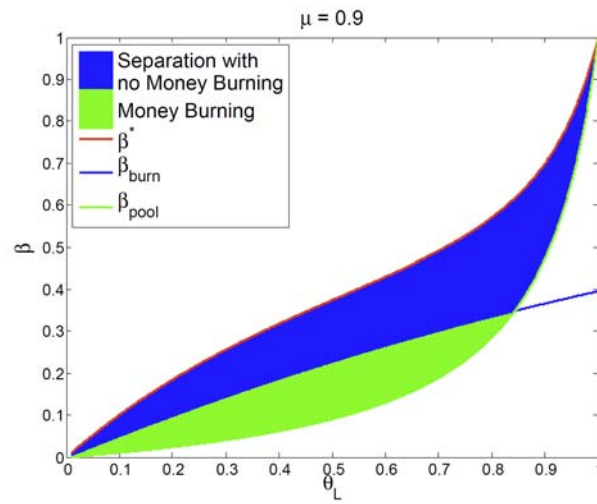


Figure 6: Money-burning and no money-burning separating regions for  $\mu = 0.9$ .



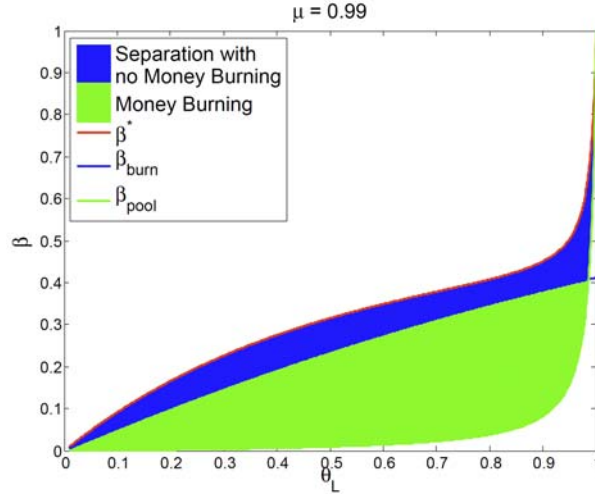


Figure 7: Money-burning and no money-burning separating regions for  $\mu = 0.99$ .

We finish the section by pointing out that clearly, with two types, contracts that involve money-burning have a “withdrawal fee” interpretation. Indeed, they may be implemented as follows. The agent can withdraw up to  $c_l$  in period 1 free of charge. Withdrawal of any larger amount is possible, but requires paying a fee of  $y - c_h$ . In equilibrium then, type  $\theta_l$  will withdraw  $c_l$ , and type  $\theta_h$  will withdraw the full amount but consume only  $c_h < y$ .

## 4 Continuum of types

We now consider the case with an arbitrary number of types. As AWA notice, money-burning is possible. However, they suggest in Proposition 2 that the highest types never engage in money-burning. We saw that this is not true with two types, and here we show that it does not actually hold for more types either.

Let us restrict attention to the case where the support of  $\theta$  is a compact segment  $\Theta = [\theta_0, \bar{\theta}]$ , and that  $f(\theta)$  is positive on  $\Theta$ . Denote

$$G(\theta) = F(\theta) + \theta(1 - \beta)f(\theta), \quad (16)$$

and let  $\theta_p$  be the lowest  $\theta \in \Theta$  such that

$$\int_{\hat{\theta}}^{\bar{\theta}} (1 - G(\tilde{\theta})) d\tilde{\theta} \leq 0 \text{ for all } \hat{\theta} \geq \theta_p.$$

Since  $F(\bar{\theta}) = 1$  and  $f(\bar{\theta}) > 0$ , we must have  $\theta_p < \bar{\theta}$ . The following proposition proves that there is “bunching at the top”, i.e., all types  $\theta > \theta_p$  get the same allocation.

**Proposition 6** *An optimal allocation  $\{(u(\theta), w(\theta))\}_{\theta \in \Theta}$  satisfies  $u(\theta) = u(\theta_p)$  and  $w(\theta) = w(\theta_p)$  for  $\theta \geq \theta_p$ . Both  $w(\theta) = z(u(\theta))$  and  $w(\theta) < z(u(\theta))$  are possible for  $\theta \geq \theta_p$ .*

This proposition corrects Proposition 2 in AWA. Like AWA, we claim that the types  $[\theta_p, \bar{\theta}]$  are pooled. Unlike AWA, we do not claim that the budget constraint holds with equality for these types and there is no money-burning at the top. On the contrary, we show that it is possible that types  $[\theta_p, \bar{\theta}]$  will have to burn money. The difference in the conclusions again arises because of the possibility that the optimal contract does not specify an interior consumption plan. In particular, in the proof of Proposition 2 AWA suggest that if  $\theta_p$  is interior (i.e.,  $\theta_p \in (\theta_0, \bar{\theta})$ ), then  $u(\theta_p)$  can be increased in a way that the IC constraint is preserved and the objective function does not decrease. However, preserving the IC constraint for type  $\theta_p$  necessarily implies that  $w(\theta_p)$  must be decreased, which is impossible if  $w(\theta_p) = 0$ . As in the case with two types, therefore, we only can have money-burning at the top if  $w(\theta) = 0$  for high types, namely, for  $\theta \geq \theta_p$ . Earlier, AWA recognizes that in the case with a continuum of types, money-burning is possible for intermediate types; the reasoning above provides the intuition for why money-burning is possible for the high types, not just intermediate ones.

Constructing an example where there is money-burning at the top can be done by a continuous approximation of the two-type Example 1. We know that type  $\theta_h$  in that example must use money-burning in equilibrium. Let us now approximate the binary distribution on  $\{\theta_l, \theta_h\}$  with a continuous distribution with full support on  $\Theta_\varepsilon = [\theta_l - \varepsilon, \theta_h + \varepsilon]$ . Proposition 6 would then suggest that on a large part of  $\Theta$  the contract should be the same. We can now use the optimal contract for the two types (which involves money-burning) to construct a contract for the continuous distribution, which will feature money-burning for high types and will yield approximately the same ex-ante payoff. Suppose that the optimal contract does not involve money-burning for the high types, and performs even better. We can then use this optimal contract to construct a contract for the case with two types, which does not involve money-burning for  $\theta_h$  and which yields approximately the same utility. But this would violate the fact that we started with the optimal contract.

The next example provides the details.

**Example 2** *As in Example 1, let  $U(c) = \sqrt{c}$ ,  $W(k) = \sqrt{k}$ ,  $y = 1$  (then  $z(u) = \sqrt{1 - u^2}$ ),  $\beta = \frac{1}{20}$ . Take  $\varepsilon \in (0, \frac{1}{10})$ , and let  $F_\varepsilon$  be the atomless distribution with finite support be given by*

the following p.d.f.:

$$f_\varepsilon(\theta) = \begin{cases} 0 & \text{if } \theta < \frac{1}{10} - \varepsilon \\ \frac{\frac{1}{10} - \varepsilon}{\frac{1}{10} - \frac{1}{10} - \varepsilon} & \text{if } \frac{1}{10} - \varepsilon \leq \theta < \frac{1}{10} \\ \frac{\varepsilon}{\frac{1}{10} - \frac{1}{10} - \varepsilon} & \text{if } \frac{1}{10} \leq \theta < 10 \\ \frac{\frac{1}{10} - \varepsilon}{\varepsilon} & \text{if } 10 \leq \theta < 10 + \varepsilon \\ 0 & \text{if } 10 + \varepsilon \leq \theta \end{cases}.$$

We can prove that for every  $\varepsilon \in (0, \frac{1}{10})$ ,  $\theta_p < \frac{1}{2}$  (it monotonically increases and has  $\frac{1}{2}$  as the limit as  $\varepsilon$  decreases from  $\frac{1}{10}$  to 0), so for all these  $\varepsilon$ , the types from  $\frac{1}{2}$  and above receive the same allocation in the optimal contract. In the proof of Proposition 6 in the Appendix, we prove that if  $\varepsilon$  is sufficiently small, then this allocation must involve money-burning. This provides a counterexample to Proposition 2 in AWA. Interestingly, even some types  $\theta < 1$  have to burn money in the optimal contract, which shows that money-burning is not necessarily a feature of types with high  $\theta$  only.

The reason why in Example 2 money-burning for high types is part of the optimal contract is intuitive: the utility of high types is sacrificed to relax the incentive compatibility constraint of the self-1 low type so that the self-0 low type can achieve a more favorable allocation in the low state. At the same time, if high types are sufficiently numerous, money-burning may be too costly for the ex-ante self, and should not be used.

These last considerations suggest a natural condition for when money-burning will not be part of the optimal contract (neither for  $\theta < \theta_p$  nor for the pooled types  $\theta \geq \theta_p$ ), which is formulated as Proposition 3 in AWA. Our point above, establishing that no money-burning at the top does not hold in general, makes this result even more powerful.

**Proposition 7** *Suppose  $G(\theta)$  is nondecreasing in  $\theta$  for  $\theta \leq \theta_p$ . Then the optimal contract takes the form of a minimal savings rule: The types  $\theta \leq \theta_p$  choose the point on the budget set that their selves-1 prefer ( $\arg \max_{(u,w \in A)} \theta u + \beta w$ ) and the types  $\theta > \theta_p$  choose the same point as  $\theta_p$ . In particular, there is no money-burning:  $w(\theta) = z(u(\theta))$  for each  $\theta \in \Theta$ .*

In other words, if  $G(\theta)$  is nondecreasing for  $\theta \leq \theta_p$ , then not only is there no money-burning, but the contract takes a simple form of minimum savings requirement. This of course implies that under the same assumption, Proposition 2 in AWA would be correct as well. Furthermore, AWA show that if this assumption on  $G(\cdot)$  does not hold, then the optimal contract cannot take the minimum savings requirement form.

Notice that as in the case with two types, withdrawal penalties may only be used if it is very likely that  $\theta$  is going to be low, so  $G(\theta)$  cannot be increasing. This corresponds to the case

where the likely need for early withdrawal is low, and only in exceptional circumstances should a person be given the flexibility to withdraw early. We showed in the previous section that this is precisely the scenario where early withdrawal penalties may be part of the optimal contract; we see here that the intuition extends to the case with a continuous types as well.

Lastly, we note that Proposition 6 in AWA, which generalizes Proposition 2 there, is also incorrect in claiming the absence money-burning, for the same reason as Proposition 2. However, under the additional assumption that guarantees that the optimal contract takes the form of minimal savings requirement, the result goes through and there is no money-burning (Proposition 7 in AWA). We omit the details here.

## 5 Conclusion

In this paper we demonstrate that withdrawal penalties, despite implying money burning from the point of view of a consumer, can be part of the optimal savings contract when the consumer faces self-control problems with current consumption. With two states we provide a full characterization of the optimal contract, and identify the region in which the optimal contract involves withdrawal penalties. We also show that withdrawal penalties for high types can also be part of the optimal contract when the state space is continuous. A full characterization of the optimal contract in the continuous setting is beyond the reach of the current paper, and is left to future research. Another possible direction of future investigation is characterizing the optimal savings contract in the model framework of Fudenberg and Levine (2006), which is a major alternative to the quasi-hyperbolic approach used in AWA and the current paper. The specification of the above model with linear costs of self-control is particularly tractable, potentially facilitating the analysis of an infinite-horizon contracting problem.

## 6 Appendix

**Proof of Proposition 1.** First note that if  $\beta \geq \beta^*$ , where

$$\beta^* = \theta_l \frac{u_h^{fb} - u_l^{fb}}{w_l^{fb} - w_h^{fb}},$$

then first-best allocation is implementable, and moreover  $\beta^* > \frac{\theta_l}{\theta_h}$ . This is correctly proven in AWA.

From now on, consider the case  $\beta < \beta^*$ . Adding the incentive constraints (9) and (10) implies  $\theta_h(u_h - u_l) \geq \theta_l(u_h - u_l)$ , which implies  $u_h \geq u_l$ . Trivially, if (9) holds with equality, then (10) holds as well. Let us prove that (9) binds (so we could forget about (10)), and also  $(u_l, w_l) \in \partial A$ ,  $(u_h, w_h) \in \partial A$ .

To see that  $(u_l, w_l) \in \partial A$ , assume the contrary. Indeed, if  $(u_l, w_l) \notin \partial A$ , then we can use the reasoning analogous to AWA: we can lower  $u_l$  and raise  $w_l$  slightly while holding  $\theta_l u_l + \beta w_l$  unchanged, so that the modified contract is still in  $A$ ; this would not change (9), will relax (10), and will increase the objective function (8), which contradict optimality of the initial contract.

To prove that  $(u_h, w_h) \in \partial A$ , suppose that  $(u_h, w_h) \notin \partial A$ , and consider the following three cases separately. If  $\beta > \frac{\theta_l}{\theta_h}$ , then a slight increase in  $u_h$  and a corresponding decrease in  $w_h$  that holds  $\theta_l u_h + \beta w_h$  unchanged will not change (9), will relax (10), and will increase the objective function (8). If  $\beta < \frac{\theta_l}{\theta_h}$ , then a slight decrease in  $u_h$  and a corresponding increase in  $w_h$  will do the same. Finally, if  $\beta = \frac{\theta_l}{\theta_h}$ , then moving  $(u_h, w_h)$  to  $\partial A$  while preserving  $\theta_l u_h + \beta w_h$  will not violate any constraint and will preserve the objective function, so without loss of generality we may assume that  $(u_h, w_h) \in \partial A$  in the optimal contract in this case as well.

Let us now prove that (9) holds with equality in the optimal contract. Denote  $\partial_f A = \{(u, w) \in \partial A : C(u) + K(w) = y\}$  (i.e., the frontier, where there is no money-burning),  $\partial_c A = \{(u, w) \in \partial A : C(u) = 0\}$  (there is no consumption in period 1) and  $\partial_k A = \{(u, w) \in \partial A : K(u) = 0\}$  (no consumption in period 2). The latter two may be empty if  $U(0) = -\infty$  or  $W(0) = -\infty$ , respectively, but in any case  $\partial A = \partial_f A \cup \partial_c A \cup \partial_k A$ .

Suppose, to obtain a contradiction, that (9) is not binding; this already implies that the optimal contract is separating. We must have  $(u_h, w_h) \in \partial_f A$ , for otherwise we would be able to increase  $u_h$  slightly without violating either of the constraints and increasing the objective function). Second, we must have  $(u_l, w_l) \in \partial_f A$ . Indeed, suppose not, then either  $(u_l, w_l) \in \partial_c A$  or  $(u_l, w_l) \in \partial_k A$ . Notice that (10) must bind, for if (10) did not bind, we could increase  $c_l$  to increase the objective function. Now, if  $(u_l, w_l) \in \partial_k A$ , then we must have  $w_l \leq w_h$  ( $w_l$  is the lowest possible), we also have  $u_l \leq u_h$  and if the contract is separating, one of the inequalities

is strict, but then (10) cannot be binding. The remaining case is  $(u_l, w_l) \in \partial_c A \setminus \partial_f A$ . Since (10) binds, we must have  $|\frac{dz}{du}|_{u=u_h}| > \frac{\theta_h}{\beta}$ . But then slightly increasing  $w_l$  coupled with moving  $(u_h, w_h)$  along  $\partial_f A$  so as to preserve (10) would unambiguously increase the objective function. This means that if (9) is not binding, then  $(u_l, w_l) \in \partial_f A$ ,  $(u_h, w_h) \in \partial_f A$ , and also  $u_l < u_h$  (otherwise the contract would be pooling, not separating). Again, suppose first that (10) binds; then  $u_l < u_h$  means that  $(u_h, w_h)$  is the rightmost point of intersection of the line corresponding to (10) and  $\partial_f A$ , and so  $|\frac{dz}{du}|_{u=u_h}| > \frac{\theta_h}{\beta}$ ; in this case, moving  $(u_h, w_h)$  slightly in the direction of  $(u_h^{fb}, w_h^{fb})$  would relax (10) and increase the objective function. The last possibility is that (10) does not bind. Then we could move either  $(u_h, w_h)$  slightly in the direction of  $(u_h^{fb}, w_h^{fb})$  or  $(u_l, w_l)$  slightly in the direction of  $(u_l^{fb}, w_l^{fb})$  so as to increase the objective function without violating any of the non-binding constraints. The only case where such deviation would not be possible is where  $(u_h, w_h) = (u_h^{fb}, w_h^{fb})$  and  $(u_l, w_l) = (u_l^{fb}, w_l^{fb})$ . But this is not an incentive compatible contract if  $\beta < \beta^*$  by the definition of  $\beta^*$ . This contradiction proves that (9) must bind.

Consider the case  $\frac{\theta_l}{\theta_h} < \beta < \beta^*$ . Let us prove that the contract is separating. Indeed, if it were pooling, then, first of all,  $(u_l, w_l) = (u_h, w_h) \in \partial_f A$ . If this contract is  $\lambda_r^K$  (but not  $\lambda_l^K$ ) for  $K = u_l + \frac{\beta}{\theta_l} w_l$ , then we can lower  $u_l$  and raise  $w_l$  slightly while holding  $\theta_l u_l + \beta w_l$  unchanged; this would not change (9), will relax (10), and will increase the objective function (8). If this contract corresponds to  $\lambda_l^K$  (but not  $\lambda_r^K$ ), then we can raise  $u_h$  and lower  $w_h$  slightly while holding  $\theta_l u_h + \beta w_h$  unchanged with similar effects. The remaining case is where  $\lambda_l^K = \lambda_r^K$ ; this means that  $|\frac{dz}{du}|_{u=u_h}| = \frac{\theta_l}{\beta}$ , and then moving  $(u_l, w_l)$  in the direction of  $(u_l^{fb}, w_l^{fb})$  and moving  $(u_h, w_h)$  in the direction of  $(u_h^{fb}, w_h^{fb})$  in a way that (9) continues to bind will relax (10) and will increase the objective function. Consequently, the optimal contract is separating. This implies  $u_l < u_h$ , and thus (10) does not bind. From this one can easily prove that  $(u_l, w_l) \in \partial_f A$  (otherwise slightly increasing  $\theta_l u_l + \beta w_l$  would create an incentive compatible contract which yields a higher ex-ante payoff) and, moreover,  $|\frac{dz}{du}|_{u=u_l}| \in [\theta_l, \frac{\theta_l}{\beta}]$  (in particular,  $u_l \in [u_l^{fb}, u_h^{fb}]$ ). Indeed, if  $|\frac{dz}{du}|_{u=u_h}| < \theta_l$ , then moving  $(u_l, w_l)$  in the direction of  $(u_l^{fb}, w_l^{fb})$  would increase the ex-ante payoff, and  $|\frac{dz}{du}|_{u=u_l}| > \frac{\theta_l}{\beta}$  makes  $(u_l, w_l) \in \partial_f A$  and (9) binding incompatible with  $u_h > u_l$ . As for  $(u_h, w_h)$ , we can rule out  $(u_h, w_h) \in \partial_c A$  (as then  $u_l < u_h$  is impossible), but as we show, both  $(u_h, w_h) \in \partial_f A$  and  $(u_h, w_h) \in \partial_k A$  is possible.

Now consider the case  $\beta < \frac{\theta_l}{\theta_h}$ . Let us prove that the contract is pooling. If it were separating, then we can lower  $u_h$  and raise  $w_h$  slightly while holding  $\theta_l u_h + \beta w_h$  unchanged (the fact that  $(u_l, w_l) \in A$  ensures that such deviation results in a contract within  $A$ , but it also preserves (9), (10) and increases the ex-ante payoff (8). Hence, the contract is pooling. This means that

$(u_l, w_l) = (u_h, w_h) \in \partial_f A$  and also  $u_l \in [u_l^{fb}, u_h^{fb}]$ , for otherwise moving the pooled contract along  $\partial_f A$  in the direction of the first-best contract would increase the ex-ante payoff.

We thus showed that the contract is separating if  $\frac{\theta_l}{\theta_h} < \beta < \beta^*$ , pooling if  $\beta < \frac{\theta_l}{\theta_h}$ , and money-burning is possible only in the separating case and for type  $\theta_h$  only. The possibility of money-burning for type  $\theta_h$  is established by Example 1; the construction of an example without money-burning at optimum is trivial. This completes the proof. ■

**Proof of Proposition 2.** Take  $\beta \in (\frac{\theta_l}{\theta_h}, \beta^*)$ . From the proof of Proposition 1, if money-burning is part of the optimal contract, then  $(u_h, w_h) \in \partial_k A \setminus \partial_f A$ , so  $k_h = 0$ ,  $c_h < y$ . This already implies  $W(0) > -\infty$ .

Also, by Proposition 1 we know that (9) is binding. Consequently, if  $(u_l, w_l, u_h, w_h)$  is the optimal contract, then  $u_h + \frac{\beta}{\theta_l} w_h = u_l + \frac{\beta}{\theta_l} w_l$ , which we denote by  $K$ . This means that  $(u_l, w_l), (u_h, w_h) \in \lambda^K$ . Moreover, from the proof of Proposition 1 we know that  $(u_l, w_l) = \lambda_l^K$ ,  $(u_h, w_h) = \lambda_h^K$ . This proves that the optimal contract solves problem (14). Moreover, again from the proof of Proposition 1, we have  $(u_l, w_l) \in \partial_f A$ , and also  $u_l \geq u_l^{fb}$ , so it suffices to optimize over  $K \geq u_l^{fb} + \frac{\beta}{\theta_l} w_l^{fb}$  only.

Let us first establish that (14) is strictly concave in  $K$ . Take two values of  $K$ ,  $K_1$  and  $K_2$ , and denote the value of the maximand in (14) by  $v(K_1)$  and  $v(K_2)$ , respectively. Now take any  $\delta \in (0, 1)$ . Given the linearity of the objective function (14) and the constraints (9) and (10), the contract given by  $u'_l = \delta u_l^{K_1} + (1 - \delta) u_l^{K_2}$ ,  $u'_h = \delta u_h^{K_1} + (1 - \delta) u_h^{K_2}$ ,  $w'_l = \delta w_l^{K_1} + (1 - \delta) w_l^{K_2}$ ,  $w'_h = \delta w_h^{K_1} + (1 - \delta) w_h^{K_2}$  satisfies the constraints and yields the value of (14)  $v'$  equal to  $\delta v(K_1) + (1 - \delta) v(K_2)$ ; moreover, it lies in  $A$  due to convexity of  $A$ . Since we proved that we can only improve by moving  $(u_l, w_l)$  to the upper-left and  $(u_h, w_h)$  to the lower-right, we get that  $v(\delta K_1 + (1 - \delta) K_2) > \delta v(K_1) + (1 - \delta) v(K_2)$  (to see that the inequality is strict, notice that at least  $(u'_l, w'_l)$  necessarily lies in the interior of  $A$ ). Hence we established that (14) is strictly concave in  $K$ .

We now see that money-burning is optimal if and only if (14) increases if we decrease  $K$  a little bit from the value  $K_0 \geq U(y) + \frac{\beta}{\theta_l} W(0)$ . If  $u(0) = U(0)$ , then doing so decreases the value of the objective function, because both the low type and the high type will get a smaller payoff. Now consider two cases. Suppose first that  $K_0 > u_l^{fb} + \frac{\beta}{\theta_l} w_l^{fb}$ ; then the formula (13) is derived in the main text. If  $K_0 \leq u_l^{fb} + \frac{\beta}{\theta_l} w_l^{fb}$ , then  $|\frac{dz}{du}|_{u=u_0} \leq \theta_l$  as  $u_0 \leq u_l^{fb}$ . But in this case the right-hand side of (13) does not exceed  $\mu\theta_l < 1$ , so the formula is correct in this case as well. ■

**Proof of Corollary 1.** From Proposition 2, we have  $W(0) > -\infty$ , so  $W(\cdot)$  is bounded

away from  $-\infty$ . Now, we have a combination of (in)equalities:

$$\begin{aligned}\theta_l U(y) + \beta W(0) &> \theta_l u_h + \beta w_h; \\ \theta_l u_h + \beta w_h &= \theta_l u_l + \beta w_l; \\ \theta_l u_l + \beta w_l &\geq \theta_l u_l^{fb} + \beta w_l^{fb}\end{aligned}$$

The first follows as  $u_h < U(0)$  and  $w_h = W(0)$ ; the second follows because (9) is binding (as shown in the proof of Proposition 1), and the third holds because  $|\frac{dz}{du}|_{u=u_l} \in [\theta_l, \frac{\theta_l}{\beta}]$  (again from the proof of Proposition 1), so  $|\frac{dz}{du}|_{u=u_l} \leq \frac{\theta_l}{\beta}$ , and now  $u_l^{fb} < u_l$  and  $(u_l^{fb}, w_l^{fb}) \in \partial_f A$  imply the required condition. This implies  $\theta_l U(y) + \beta W(0) > \theta_l u_l^{fb} + \beta w_l^{fb}$ , which completes the proof. ■

**Proof of Corollary 2.** Consider the additional constraints  $(u_l, w_l), (u_h, w_h) \in \partial_f A$ . As in the proof of Proposition 1, we can establish that (9) is binding and the optimal contract is separating. Thus,  $(u_l, w_l) = \lambda_l^K$ ,  $(u_h, w_h) = \lambda_h^K$ , and moreover,  $K \geq K_0$ . In the proof of Proposition 2 we established that the problem (14) is strictly concave in  $K$ . If money-burning takes place in the unconstrained optimum, then the unique maximum of (14) is achieved at  $K < K_0$ . Hence, after adding the constraint  $K \geq K_0$ , we get the constrained maximum at  $K = K_0$ . ■

**Proof of Proposition 3.** Given  $U(\cdot)$  and  $W(\cdot)$ , the set  $A$  is fixed. Let  $w = z(u)$  be the equation that determines the upper boundary of this set and let  $k = |\frac{dz}{du}(u_0)| \geq 1$ . By assumption that  $W(0) \neq -\infty$  and convexity of  $A$ , the number  $s = \frac{z(u_0) - W(0)}{U(y) - u_0} \in (k, \infty)$ . For any  $\beta \in (0, \frac{1}{s}) \subset (0, 1)$ , let  $\theta_l(\beta) = \beta s$ . In this case,  $u_0$  will be the  $u_0$  from formulation of Claim 2. We have  $\mu \frac{1-\beta}{|\frac{dz}{dx}(u_0)|^{-\theta_l(\beta)}} = \mu \frac{1-\beta}{\frac{1}{k} - \frac{1}{s}}$ . But  $s \in (k, \infty)$  and  $k \geq 1$  implies  $\frac{1}{k} - \frac{1}{s} \in (0, 1)$ , which means that inequality (13) must hold for  $\beta$  sufficiently close to 0 and  $\mu$  sufficiently close to 1 (and  $\theta_l, \theta_h$  derived by  $\theta_l = \beta s$  and  $\theta_h = \frac{1-\mu\theta_l}{1-\mu}$ ). Moreover, for  $\mu$  close to 1 we will have  $\theta_h$  arbitrarily high, in particular,  $\theta_h > s = \frac{\theta_l(\beta)}{\beta}$ . The latter implies  $\beta > \frac{\theta_l}{\theta_h}$ , and we have  $\beta < \beta^*$  by construction, so in this case, indeed, a separating contract is optimal by Proposition 1. Finally, since varying  $u_0$  would not change the inequalities above, then the set of parameters  $\beta, \mu, \theta_l$  for which money-burning is optimal contains an open set. ■

**Proof of Proposition 4.** Fix  $u_0$  and thus  $|\frac{dz}{du}|_{u=u_0} = x > 1$ . Let us take  $\beta_0 = \frac{\theta_l}{\theta_h} = \frac{\theta_l(1-\mu)}{1-\mu\theta_l}$



and plug it into (13). We get:

$$\begin{aligned} \mu \frac{1 - \beta_0}{\left| \frac{dz}{du} \Big|_{u=u_0} \right| - \frac{\beta_0}{\theta_l}} - 1 &= \mu \frac{1 - \frac{\theta_l(1-\mu)}{1-\mu\theta_l}}{\frac{1}{x} - \frac{1-\mu}{1-\mu\theta_l}} - 1 \\ &= \frac{x-1}{1 - x \frac{1-\mu}{1-\mu\theta_l}} = \frac{x-1}{1 - \frac{x}{\theta_h}} \geq 0, \end{aligned}$$

because  $x \leq \frac{\theta_l}{\beta} < \frac{\theta_l}{\beta_0} = \theta_h$  and  $x > 1$ . Notice that the left-hand side of (13) is increasing in  $\beta$  (again for a fixed  $\left| \frac{dz}{du} \Big|_{u=u_0} \right| = x$ ): indeed, we have

$$\frac{d}{d\beta} \left( \mu \frac{1 - \beta}{\frac{1}{x} - \frac{\beta}{\theta_l}} \right) = x\mu\theta_l \frac{x - \theta_l}{(\theta_l - x\beta)^2} > 0,$$

as  $x > 1 > \theta_l$ . Consequently, for  $\beta > \beta_0 = \frac{\theta_l}{\theta_h}$ , condition (13) holds with strict inequality, and money-burning is optimal.

To prove the second part, it now suffices to prove that for all  $\beta \in \left( \frac{\theta_l}{\theta_h}, \beta^* \right)$ ,  $u_0 > U(0)$  (then we would have  $\left| \frac{dz}{du} \Big|_{u=u_0} \right| > 1$  and the first part would apply). But now the first-best points are  $(U(y), W(0))$  for  $\theta_h$  and  $(U(0), W(y))$  for  $\theta_l$ . Consequently, the leftmost point of the green line corresponding to  $\beta < \beta^*$  that lies in  $A$  satisfies  $u_0 = U(0)$ , and thus the previous result is applicable. This completes the proof. ■

**Proof of Proposition 5.** For fixed  $\beta$  and  $\theta_l$ ,  $u_0$  is also fixed. Then the left-hand side of (13) is clearly increasing in  $\mu$ , so if (13) did not hold for a given  $\mu$ , then it would not hold for a lower  $\mu$ . Hence, if the new optimal contract is separating, there will be no money-burning because (13) would not hold, and if the new optimal contract is not separating, then such optimal contract never involves money-burning. Finally, a decrease in  $\theta_h$ , for a fixed  $\theta_l$ , implies a lower  $\mu$ , and we can use the previous reasoning. This completes the proof. ■

**Proof of Proposition 6.** The proof that  $u(\theta) = u(\theta_p)$  for  $\theta \geq \theta_p$  in AWA is correct, and is omitted here. Trivially, we must have  $w(\theta) = w(\theta_p)$  for  $\theta \geq \theta_p$  as well (otherwise, only the contracts with the highest  $w$  will be chosen). This proves the first part of the Proposition.

Let us show that  $w(\theta_p) < z(u(\theta_p))$  is possible, so money-burning for high types is possible. Consider Example 2. We have  $G_\varepsilon(\theta) = F_\varepsilon(\theta) + \theta(1 - \beta)f_\varepsilon(\theta)$  equal to

$$G_\varepsilon(\theta) = \begin{cases} 0 & \text{if } \theta < \frac{1}{10} - \varepsilon \\ \left( \theta - \frac{1}{10} + \varepsilon \right) \frac{\frac{10}{11} - \frac{\varepsilon}{2}}{\varepsilon} + \theta \left( 1 - \frac{1}{20} \right) \frac{\frac{10}{11} - \frac{\varepsilon}{2}}{\varepsilon} & \text{if } \frac{1}{10} - \varepsilon \leq \theta < \frac{1}{10} \\ \frac{10}{11} - \frac{\varepsilon}{2} + \left( \theta - \frac{1}{10} \right) \frac{\frac{\varepsilon}{10 - \frac{1}{10}}}{\varepsilon} + \theta \left( 1 - \frac{1}{20} \right) \frac{\frac{\varepsilon}{10 - \frac{1}{10}}}{\varepsilon} & \text{if } \frac{1}{10} \leq \theta < 10 \\ \frac{10}{11} + \frac{\varepsilon}{2} + \left( \theta - 10 \right) \frac{\frac{1}{11} - \frac{\varepsilon}{2}}{\varepsilon} + \theta \left( 1 - \frac{1}{20} \right) \frac{\frac{1}{11} - \frac{\varepsilon}{2}}{\varepsilon} & \text{if } 10 \leq \theta < 10 + \varepsilon \\ 1 & \text{if } 10 + \varepsilon \leq \theta \end{cases}.$$

Direct computations give the threshold  $\theta_p$  as a decreasing function of  $\varepsilon$  on  $(0, \frac{1}{10})$ , which monotonically increases from  $\frac{5620 - \sqrt{28754482}}{780} = 0.33$  to  $\frac{1}{2}$  as  $\varepsilon$  decreases from  $\frac{1}{10}$  to 0:

$$\theta_p(\varepsilon) = \frac{1}{390\varepsilon} \left( 1010\varepsilon + 180 - \sqrt{5} \sqrt{3861\varepsilon^3 + 202538\varepsilon^2 + 58680\varepsilon + 6480} \right).$$

In particular, this implies that all individuals with  $\theta \geq \frac{1}{2}$  are pooled.

Let us prove that this contract must involve money-burning for  $\varepsilon$  small enough for all individuals with  $\theta \geq \frac{1}{2}$ . Recall the values  $V$  and  $\tilde{V}$  we defined in Example 1 as the ex-ante payoff from the optimal contract and the optimal contract subject to no money burning in state  $\theta_h = 10$ ; we had  $V > \tilde{V}$ . In this example, for  $\varepsilon \in (0, \frac{1}{10})$ , let us define the ex-ante payoff from the optimal contract as  $V_\varepsilon$  and that from the optimal contract with the constraint that types  $\theta \geq \frac{1}{2}$  do not burn money (and thus types  $\theta > \theta_p(\varepsilon)$  do not burn money) by  $\tilde{V}_\varepsilon$ . We now prove that  $\liminf_{\varepsilon \rightarrow 0} V_\varepsilon \geq V$  and that  $\limsup_{\varepsilon \rightarrow 0} \tilde{V}_\varepsilon \leq \tilde{V}$ ; this would establish that for  $\varepsilon$  small enough, money-burning must be used for the types  $\theta \geq \frac{1}{2}$ .

We first prove that  $\liminf_{\varepsilon \rightarrow 0} V_\varepsilon \geq V$ . Let us take the optimal contract for the two-type case,  $\Xi = (c_l, k_l, c_h, k_h)$ , and provide these two options,  $(c_l, k_l)$  and  $(c_h, k_h)$ , to all types from  $\frac{1}{10} - \varepsilon$  to  $10 + \varepsilon$ . From Proposition 1, we know that type  $\theta_l = \frac{1}{10}$  is indifferent between the two contracts; then single-crossing considerations will imply that types  $\theta < \frac{1}{10}$  will choose  $(c_l, k_l)$  while types  $\theta > \frac{1}{10}$  will choose  $(c_h, k_h)$ . The ex-ante utility from such contract equals

$$V'_\varepsilon = \left( \frac{10}{11} - \frac{\varepsilon}{2} \right) \left( \left( \frac{1}{10} - \frac{\varepsilon}{2} \right) u_l + w_l \right) + \varepsilon \left( \left( 5 + \frac{1}{20} \right) u_h + w_h \right) + \left( \frac{1}{11} - \frac{\varepsilon}{2} \right) \left( \left( 10 + \frac{\varepsilon}{2} \right) u_h + w_h \right).$$

Clearly, we have  $\lim_{\varepsilon \rightarrow 0} V'_\varepsilon = V_\varepsilon$ . But we have taken some contract, not necessarily optimal, so  $V_\varepsilon \geq V'_\varepsilon$  for all  $\varepsilon$ . This implies  $\liminf_{\varepsilon \rightarrow 0} V_\varepsilon \geq V$ .

Let us now prove that  $\limsup_{\varepsilon \rightarrow 0} \tilde{V}_\varepsilon \leq \tilde{V}$ . Suppose this is not the case, and there exists  $\delta > 0$  and a monotonically decreasing sequence  $\varepsilon_1, \varepsilon_2, \dots$  with  $\lim_{n \rightarrow \infty} \varepsilon_n = 0$  such that  $\tilde{V}_{\varepsilon_n} > \tilde{V} + \delta$  for all  $n \in \mathbb{N}$ . Suppose that  $\Xi^{\varepsilon_n} = \{(c^{\varepsilon_n}(\theta), k^{\varepsilon_n}(\theta))\}_{\theta \in [\frac{1}{10} - \varepsilon_n, 10 + \varepsilon_n]}$  is the optimal contract for  $\varepsilon_n$ , subject to no money-burning for types  $\theta \geq \frac{1}{2}$ . Let us construct a binary contract  $(\tilde{c}_l^{\varepsilon_n}, \tilde{k}_l^{\varepsilon_n}, \tilde{c}_h^{\varepsilon_n}, \tilde{k}_h^{\varepsilon_n})$  in the following way. We let  $(c_h^{\varepsilon_n}, k_h^{\varepsilon_n}) = (c^{\varepsilon_n}(10), k^{\varepsilon_n}(10))$  be the contract that type  $\theta_h = 10$  chooses under  $\Xi^{\varepsilon_n}$  (as well as all types  $\theta > \theta_p(\varepsilon)$ ). We let  $(\tilde{c}_l^{\varepsilon_n}, \tilde{k}_l^{\varepsilon_n})$  be the contract that maximizes  $\max_{\theta \in [\frac{1}{10} - \varepsilon_n, \varepsilon_n]} \theta U(c^{\varepsilon_n}(\theta)) + W(k^{\varepsilon_n}(\theta))$  (the reason we do not take  $(c^{\varepsilon_n}(\frac{1}{10}), k^{\varepsilon_n}(\frac{1}{10}))$  is even in the optimal contract the type  $\theta_l$  may get a relatively low payoff, which is not a problem if this type has zero mass, but may be a problem if it has a mass of  $\frac{10}{11}$ ); suppose that this maximum is reached at  $\theta = \tilde{\theta}_{\varepsilon_n}$ .

Let us compute the ex-ante payoff from the following contract  $\tilde{\Xi}^{\varepsilon_n}: (\tilde{c}^{\varepsilon_n}(\theta), \tilde{k}^{\varepsilon_n}(\theta)) = (c_l^{\varepsilon_n}, k_l^{\varepsilon_n})$  if  $\theta \leq \frac{1}{10}$  and  $(\tilde{c}^{\varepsilon_n}(\theta), \tilde{k}^{\varepsilon_n}(\theta)) = (c_h^{\varepsilon_n}, k_h^{\varepsilon_n})$  if  $\theta > \frac{1}{10}$  for different distributions of  $\theta$ .

We first take  $f_{\varepsilon_n}$ ; the payoff from this contract (note that this contract need not be incentive compatible!) is

$$\tilde{V}'_{\varepsilon_n} = \left(\frac{10}{11} - \frac{\varepsilon_n}{2}\right) \left(\left(\frac{1}{10} - \frac{\varepsilon_n}{2}\right) u_l^{\varepsilon_n} + w_l^{\varepsilon_n}\right) + \varepsilon_n \left(\left(5 + \frac{1}{20}\right) u_h^{\varepsilon_n} + w_h^{\varepsilon_n}\right) + \left(\frac{1}{11} - \frac{\varepsilon_n}{2}\right) \left(\left(10 + \frac{\varepsilon_n}{2}\right) u_h^{\varepsilon_n} + w_h^{\varepsilon_n}\right).$$

(where  $u_l^{\varepsilon_n} = U(c_l^{\varepsilon_n}) = \sqrt{c_l^{\varepsilon_n}}$  etc are defined as usual). But under the contract  $\Xi^{\varepsilon_n}$ , types  $\theta > 10$  get exactly the same allocation as in  $\tilde{\Xi}^{\varepsilon_n}$ , and types  $\theta < \frac{1}{10}$  get payoff

$$\theta u_l^{\varepsilon_n} + w_l^{\varepsilon_n} \geq |\theta - \theta_{\varepsilon_n}| + \tilde{\theta}_{\varepsilon_n} u_l^{\varepsilon_n} + w_l^{\varepsilon_n} \geq |\theta - \theta_{\varepsilon_n}| + \theta U(c^{\varepsilon_n}(\theta)) + W(k^{\varepsilon_n}(\theta)),$$

since  $u_l^{\varepsilon_n} \in (0, 1)$ . Consequently,

$$\tilde{V}'_{\varepsilon_n} - \tilde{V}_{\varepsilon_n} \geq -\left(\frac{10}{11} - \frac{\varepsilon_n}{2}\right) \varepsilon_n - \varepsilon_n \left(5 + \frac{1}{20} + 1\right),$$

where the second term certainly exceeds the possible difference between  $\tilde{V}'_{\varepsilon}$  and  $\tilde{V}_{\varepsilon}$  coming from  $\theta \in (\frac{1}{10}, 10)$ . But the right-hand side tends to 0 as  $\varepsilon_n \rightarrow 0$ , so for  $n$  high enough,  $\tilde{V}'_{\varepsilon_n} > \tilde{V}_{\varepsilon_n} - \frac{\delta}{3}$ .

Let us now take the binary distribution as in Example 1 and consider the payoff under  $\tilde{\Xi}^{\varepsilon_n}$  (again, this contract need not be incentive compatible under this distribution). We have

$$\tilde{V}''_{\varepsilon_n} = \frac{10}{11} \left(\frac{1}{10} u_l^{\varepsilon_n} + w_l^{\varepsilon_n}\right) + \frac{1}{11} (10 u_h^{\varepsilon_n} + w_h^{\varepsilon_n}).$$

Clearly,

$$\tilde{V}''_{\varepsilon_n} - \tilde{V}'_{\varepsilon_n} \geq -\varepsilon_n \left(5 + \frac{1}{20} + 1\right) - \left(\frac{1}{11} - \frac{\varepsilon_n}{2}\right) \frac{\varepsilon_n}{2},$$

so for  $n$  high enough we have  $\tilde{V}''_{\varepsilon_n} > \tilde{V}'_{\varepsilon_n} - \frac{\delta}{3}$ .

Consider now the sequence of contracts  $\tilde{\Xi}^{\varepsilon_n}$ . It is characterized by two pairs  $(c_l^{\varepsilon_n}, k_l^{\varepsilon_n})$  and  $(c_h^{\varepsilon_n}, k_h^{\varepsilon_n})$ ; moreover,  $c_h^{\varepsilon_n} + k_h^{\varepsilon_n} = y$  is satisfied for every  $n$ . Let us pick a subsequence  $\{n_r\}$  such that  $(c_l^{\varepsilon_{n_r}}, k_l^{\varepsilon_{n_r}})$  and  $(c_h^{\varepsilon_{n_r}}, k_h^{\varepsilon_{n_r}})$  converge to some  $(\hat{c}_l, \hat{k}_l)$  and  $(\hat{c}_h, \hat{k}_h)$ ; this is possible since  $B$  is compact and, moreover, we have  $\hat{c}_h + \hat{k}_h = y$ . Denote the ex-ante payoff from this contract under the binary distribution by  $\hat{V}$ . We have:

$$\hat{V} = \frac{10}{11} \left(\frac{1}{10} \hat{u}_l + \hat{w}_l\right) + \frac{1}{11} (10 \hat{u}_h + \hat{w}_h);$$

here we used the fact that  $U(\cdot)$  and  $W(\cdot)$  are continuous. We have

$$\lim_{r \rightarrow \infty} \left(\hat{V} - \tilde{V}''_{\varepsilon_{n_r}}\right) = 0$$

by construction, and therefore for  $r$  high enough,  $\hat{V} > \tilde{V}''_{\varepsilon_{n_r}} - \frac{\delta}{3}$ .

This shows that there is some  $n$  such that  $\hat{V} > \tilde{V}_{\varepsilon_n} - \delta$ . But we took the sequence such that  $\tilde{V}_{\varepsilon_n} > \tilde{V} + \delta$  for all  $n$ , which implies that  $\hat{V} > \tilde{V}$ . Recall, however, that  $\tilde{V}$  is the ex-ante payoff

in the optimal contract with no money-burning for the high type, and  $\hat{V}$  is the ex-ante payoff in one of such contracts. We would get a contradiction if we prove that the contract  $(\hat{c}_l, \hat{k}_l)$  and  $(\hat{c}_h, \hat{k}_h)$  is incentive-compatible. To do so, let us write the following two incentive compatibility constraints that the contract  $\tilde{\Xi}^{\varepsilon_{n_r}}$  satisfies:

$$\begin{aligned}\tilde{\theta}_{\varepsilon_{n_r}} u_l^{\varepsilon_{n_r}} + \frac{1}{20} w_l^{\varepsilon_{n_r}} &\geq \tilde{\theta}_{\varepsilon_{n_r}} u_h^{\varepsilon_{n_r}} + \frac{1}{20} w_h^{\varepsilon_{n_r}}; \\ 10 u_h^{\varepsilon_{n_r}} + \frac{1}{20} w_h^{\varepsilon_{n_r}} &\geq 10 u_l^{\varepsilon_{n_r}} + \frac{1}{20} w_l^{\varepsilon_{n_r}}.\end{aligned}$$

Taking the limits as  $r \rightarrow \infty$  and using the fact that  $\tilde{\theta}_{\varepsilon_{n_r}} \in [\frac{1}{10} - \varepsilon_{n_r}, \frac{1}{10}]$  and thus tends to  $\frac{1}{10}$ , we get

$$\begin{aligned}\frac{1}{10} \hat{u}_l + \frac{1}{20} \hat{w}_l &\geq \frac{1}{10} \hat{u}_h + \frac{1}{20} \hat{w}_h; \\ 10 \hat{u}_h + \frac{1}{20} \hat{w}_h &\geq 10 \hat{u}_l + \frac{1}{20} \hat{w}_l.\end{aligned}$$

This proves that the contract  $(\hat{c}_l, \hat{k}_l, \hat{c}_h, \hat{k}_h)$  is incentive compatible, and thus  $\hat{V} \leq \tilde{V}$ . We have reached a contradiction which proves that  $\limsup_{\varepsilon \rightarrow 0} \tilde{V}_\varepsilon \leq \tilde{V}$ .

Consequently, we have established both  $\liminf_{\varepsilon \rightarrow 0} V_\varepsilon \geq V$  and  $\limsup_{\varepsilon \rightarrow 0} \tilde{V}_\varepsilon \leq \tilde{V}$ . But  $V > \tilde{V}$ , therefore, for  $\varepsilon$  close to 0,  $V_\varepsilon > \tilde{V}_\varepsilon$ . This means that there is  $\varepsilon > 0$  for which the optimal contract must involve money-burning in the allocation that types  $\theta > \theta_p(\varepsilon)$  get, and the mass of these agents is at least  $\frac{1}{11}$  (as  $\theta_p(\varepsilon) < \frac{1}{2}$ ). This completes the proof that  $w(\theta_p) < z(u(\theta_p))$  is possible.

To show that  $w(\theta_p) = z(u(\theta_p))$  is possible as well, we can refer to Proposition 7 which establishes sufficient conditions for this to be true. ■

**Proof of Proposition 7.** This result is a reformulation of Proposition 3 in AWA which is proven there correctly. The proof is thus omitted. ■

## 7 References

Amador, Manuel and Kyle Bagwell (2011): “The theory of optimal delegation with an application to tariff caps,” mimeo Stanford University.

Amador, Manuel, Iván Werning and George-Marios Angeletos (2006): “Commitment vs Flexibility,” *Econometrica*, 74, vol.2, 365-396.

Ambrus, A. and G. Egorov (2009): “Delegation and nonmonetary incentives,” mimeo, Harvard University.

Amromin, G. and P. Smith (2003): “What explains early withdrawals from retirement accounts? Evidence from a panel of taxpayers,” *National Tax Journal*, 56, 595-612.

Athey, S., A. Atkeson and P. Kehoe (2005): “The optimal degree of discretion in monetary policy,” *Econometrica*, 73, 1431-1475.

Athey, S., K. Bagwell and C. Sanchirico (2004): “Collusion and price rigidity,” *Review of Economic Studies*, 71, 317-349.

Dekel, E., B. Lipman and A. Rustichini (2001): “Representing preferences with a unique subjective state space,” *Econometrica* 69, 891-934.

Fudenberg, D. and D. Levine (2006): “A dual-self model of impulse control,” *American Economic Review*, 96, 1449-1476.

Gilkeson, J., J. List and C. Ruff (1999): “Evidence of early withdrawal in time deposit portfolios,” *Journal of Financial Services Research*, 15, 103-122.

Gilkeson, J. and C. Ruff (1996): “Valuing the withdrawal option in retail CD portfolios,” *Journal of Financial Services Research*, 10, 333-358.

Gul, F. and W. Pesendorfer (2001): “Temptation and self-control,” *Econometrica*, 69, 1403-1435.

Laibson, D. (1997): “Golden eggs and hyperbolic discounting,” *Quarterly Journal of Economics*, 112, 443-477.

Laibson, D. (1998): “Life-cycle consumption and hyperbolic discount functions,” *European Economic Review*, 42, 861-871.

Phelps, E. and R. Pollack (1968): “On second best national savings and game-equilibrium growth,” *Review of Economic Studies*, 35, 185-199.

Strotz, R. (1956): “Myopia and inconsistency in dynamic utility maximization,” *Review of Economic Studies*, 23, 165-180.