Econometrica, Vol. 83, No. 4 (July, 2015), 1425–1465

COMMITMENT, FLEXIBILITY, AND OPTIMAL SCREENING OF TIME INCONSISTENCY

BY SIMONE GALPERTI[1]

This paper develops a theory of optimal provision of commitment devices to people who value both commitment and flexibility and whose preferences differ in the degree of time inconsistency. If time inconsistency is observable, both a planner and a monopolist provide devices that help each person commit to the efficient level of flexibility. However, the combination of unobservable time inconsistency and preference for flexibility causes an adverse-selection problem. To solve this problem, the monopolist and (possibly) the planner curtail flexibility in the device for a more inconsistent person at both ends of the efficient choice range; moreover, they may have to add unused options to the device for a less inconsistent person and also distort his actual choices. This theory has normative and positive implications for private and public provision of commitment devices.

KEYWORDS: Time inconsistency, self-control, commitment, flexibility, screening, unused options.

1. INTRODUCTION

EVIDENCE SUGGESTS THAT MANY PEOPLE have self-control problems (DellaVigna (2009)). Often aware of them, these people create a demand for commitment devices, which has received the attention of firms and governments.[2] However, people are usually uncertain about the future and hence demand devices that also allow for flexibility. Is it feasible for firms or governments to provide devices that satisfy individuals' conflicting desires for commitment and flexibility? If yes, is it optimal for them to do so? Moreover, self-control varies across people and is not immediately detectable. How does this affect the provision of commitment devices?

To answer these questions, this paper develops a theory of optimal provision of flexible commitment devices from the viewpoints of a profit-maximizing firm

[1]I am indebted to Eddie Dekel and Alessandro Pavan for many discussions that greatly improved the paper. I also thank S. Nageeb Ali, Gene Amromin, Stefano DellaVigna, Jeffrey Ely, William Fuchs, Garrett Johnson, Botond Koszegi, David Laibson, Santiago Oliveros, Jonathan Parker, Nicola Persico, Henrique Roscoe de Oliveira, Todd Sarver, Ron Siegel, Bruno Strulovici, Balazs Szentes, Rakesh Vohra, Asher Wolinsky, Michael Whinston, Leeat Yariv, and seminar participants at Northwestern, NYU, Berkeley, Duke, UCSD, Michigan, NYU Stern, LSE, EEA-ESEM 2012, ESSET 2013, and SAET 2013. I gratefully acknowledge financial support from the WCAS Center of Economic Theory of Northwestern University. All remaining errors are mine.

[2]Examples include saving devices such as individual retirement accounts and 401(k) plans designed by the U.S. government, or the devices offered by firms such as StickK and GymPact. Other examples are automatic drafts from checking to savings accounts, Christmas clubs, rotating savings and credit associations, microcredit savings accounts, fat farms, and programs to cut consumption of cigarettes, alcohol, or drugs (Ashraf, Gons, Karlan, and Yin (2003), Ashraf, Karlan, and Yin (2006), DellaVigna and Malmendier (2004), Bryan, Karlan, and Nelson (2010)).

 DOI: 10.3982/ECTA11851

and a welfare-maximizing planner. If devices can rely on rich transfer schemes, it is feasible to fully satisfy people's desires for commitment and flexibility. However, the *combination* of people's demand for flexibility and private information on their self-control causes an adverse-selection problem, which in turn leads to a trade-off between commitment and flexibility. The paper characterizes how, as a result, the firm and (possibly) the planner optimally curtail flexibility for people with self-control problems.

The model has two players—a provider (she) and an agent (he)—and two periods. At time 1, the provider offers the agent a commitment device, which consists of a menu of options. Each option involves a contractible action and a payment to the provider, which can be positive or negative. After committing to a device, at time 2 the agent chooses one of its options. An example would be a savings device through which the agent can save different amounts, each leading to a tax penalty or a subsidy. At time 1, the agent desires flexibility, because his preference over time-2 actions depends on an uncertain state; importantly, states are not contractible. At time 1, the agent can also desire commitment, because his preference can be time inconsistent. He correctly predicts his inconsistency (full sophistication), and this is common knowledge.[3] As in most of the literature, this paper uses time-1 preferences to measure efficiency. Finally, only the agent knows his degree of time inconsistency (or type) from time 1. For purposes of illustration, suppose he can be only consistent (type C) or inconsistent (I).

It is common to think that coexisting preferences for commitment and for flexibility must involve a trade-off, which may emerge in how commitment devices are designed, even when the agent's degree of inconsistency is observable (Amador, Werning, and Angeletos (2006), Ambrus and Egorov (2013), Bond and Sigurdsson (2013)). By contrast, Section 3 shows that, for any such degree, it is feasible to design a device that commits type I to a flexible plan of action which is efficient in each state. Roughly speaking, this difference arises because in the present model payments can be negative, and thus a device can penalize certain actions and at the same time reward others (see Section 5.1). Moreover, if types are observable, it is always optimal for the provider to offer I an efficient device.[4] The payments that such a device features are also characterized in Section 3. Section 1.1 illustrates these payments, together with the other key results, in an intuitive example.

Section 4 addresses the case of unobservable time inconsistency. In this case, a nontrivial adverse-selection problem arises because a less inconsistent agent values any flexible device strictly more than a more inconsistent agent. This is because, for example, an efficient savings device for I rewards high levels of savings and penalizes low levels. But if C takes it, being consistent, he predicts that he will incur fewer penalties and enjoy more rewards than I, and

[3]Section 6 discusses the case of a naive agent.
[4]This result generalizes a result of DellaVigna and Malmendier (2004) (Section II.F, p. 368).

hence he expects a higher payoff. As in the usual screening models, the device designed for C (hereinafter called the C-device) must then grant him information rents—for example, through a discount. Here, however, granting the minimal rents that render C indifferent between the C-device and the one designed for I (hereinafter called the I-device) can suddenly make I strictly prefer the C-device. This device does not feature payments that will solve I's inconsistency, and yet its discount can still be enough to lure I.

Section 4 then characterizes the devices that optimally screen time inconsistency. The I-device curtails flexibility at both ends of the efficient choice range: I reacts to both high and low states less than efficiently—resulting in a narrower choice range—and does not react at all to extreme states, again both high and low. This is because at high and low states flexibility turns out to cause the largest costs in terms of rents that the provider must grant C. To induce these inefficiencies, the provider modifies the payments I faces in an efficient device, as illustrated in Section 1.1.

The screening C-device may also have to depart from the optimal one in the case of observable types, so as to ward off I. The first strategy the provider adopts to bring this about is unconventional and is never useful in models with only consistent agents. To an otherwise efficient C-device, the provider adds options that C never uses but I views as temptations, thereby rendering the device less attractive for I.[5] But to ward off I, these unused options must be tempting enough; otherwise, the provider will also have to distort C's choices. This implies that screening time inconsistency violates the usual "no distortion at the top" property.[6]

This paper relates to the literature on the trade-off between preferences for commitment and those for flexibility.[7] As explained in Section 5.1, this literature usually restricts the payments that commitment devices can use and therefore the trade-off results from technology constraints. Here, by contrast, a tension between commitment and flexibility arises from information constraints. This literature also assumes that agents' degree of inconsistency is observable and does not explicitly model markets for commitment devices. Section 5.1 also compares this literature's and the paper's implications for commitment devices that involve savings.

Other papers have addressed the problem of designing contracts for agents with low self-control.[8] This paper, however, is the first to examine this problem when agents value both commitment and flexibility, are heterogeneous in their degree of inconsistency, and privately know this degree. In O'Donoghue and

[5]Related results appear in Esteban and Miyagawa (2006) and Bond and Sigurdsson (2013). See Section 5.2 for the differences from the present paper.

[6]See, for example, Mussa and Rosen (1978), Courty and Li (2000), and Battaglini (2005).

[7]See Amador, Werning, and Angeletos (2006), Ambrus and Egorov (2013), Bond and Sigurdsson (2013), and Halac and Yared (2014).

[8]See Kőszegi's (2013) survey on contracting with behaviorally biased agents.

Rabin (1999b) and DellaVigna and Malmendier (2004), the agent has no private information when offered a contract. In other papers, the agent has some private information from the outset but does not value flexibility. Moreover, in Eliaz and Spiegler (2006), the designer knows the agent's degree of inconsistency, but does not know if he is aware of his inconsistency. In Esteban and Miyagawa (2006), the designer screens the agent's usual willingness to pay—not his degree of inconsistency. Finally, Heidhues and Kőszegi (2010) focused on settings in which the optimal contracts with symmetric information also ensure self-selection with asymmetric information.

As discussed in Section 5.2, this paper is also related to sequential-screening models with only consistent agents. Since unused options can be a necessary part of optimal mechanisms, the paper highlights the following about direct mechanisms: With only consistent agents, we can always focus on mechanisms that describe only options on the truthful paths of play—and hence at each time ask the agent for only the information he acquires at that time (Myerson (1986)). With inconsistent agents, we also *have to* consider mechanisms that allow for off-path options, in the sense that an agent would choose them at time 2 only if he misreported his information at time 1. And to do so, these mechanisms may have to let the agent report again at time 2 information he received at time 1.[9] Finally, to find the optimal mechanism, the paper uses non-standard and recent methods. To handle the incentive constraints involving the agent's degree of inconsistency, it uses Lagrangian methods from Luenberger (1969), which differ from standard optimal-control methods and the standard dynamic-mechanism-design approach.[10] To ensure that the mechanism satisfies certain monotonicity properties, it adapts Toikka's (2011) generalization of Myerson's (1981) ironing method, to allow for off-path options.

Appendix A contains the proofs of the main results. All omitted proofs are in Appendix B of the Supplemental Material (Galperti (2015)).

1.1. *Illustrative Example*

This section presents an illustrative consumption–savings example. An agent is either time consistent (type C) or time inconsistent (type I). There are two periods. At time 2, the agent chooses consumption y and savings a. His budget set is $B = \{(y, a) \in \mathbb{R}^2_+ : y + a \leq m\}$, where m is his income. His utility function is $y + 2ts\sqrt{a}$, where $2\sqrt{a}$ is the benefit of a at retirement (after time 2), and t equals 1 for C and 0.9 for I. The agent privately learns state s at time 2, which is uniformly distributed on $[10, 15]$. At time 1, the agent knows m and t, but not s. Independently of his type, he evaluates his time-2 choices with the same utility function $y + 2s\sqrt{a}$. Thus at time 1, the agent would like to save $\mathbf{e}(s) = s^2$

[9]This is the case, for instance, if the support of time-2 information is independent of time-1 information.

[10]See, for example, Courty and Li (2000) and Pavan, Segal, and Toikka (2014).

at time 2—we will refer to **e** as the efficient outcome.[11] To be able to save at time 2, he has to obtain a savings device from a third party (provider) at time 1; without such a device, at time 2 he simply consumes his whole income, m. For each a, a device specifies a payment $p(a)$ from the agent to the provider, which can be negative. Thus after saving a, the agent will have $m - a - p(a)$ left for consumption. The provider designs her devices to maximize her expected revenues.

Suppose first that the provider observes t. Note that if a device induces the agent to follow plan **e** at time 2, it also maximizes how much he is willing to pay in expectation at time 1. Thus to type C the provider offers a device with a constant payment $p_{\mathbf{e}}^C$ (like a set-up fee) which is high enough that his expected utility equals that of the outside option (i.e., m). A device with constant p, however, induces type I to undersave relative to **e**. Indeed, the payment designed for the first-best I-device is $p_{\mathbf{e}}^I(a) = p_{\mathbf{e}}^C + k - 0.1a$, where $k > 0$ is set so that I expects to pay exactly $p_{\mathbf{e}}^C$.[12] It is easy to see that, by properly rewarding high levels of savings and penalizing low levels, $p_{\mathbf{e}}^I$ induces I to follow plan **e**. It also generates the same expected revenue as the C-device. In short, knowing t, the provider offers C and I different devices, but each induces the corresponding type to save efficiently (Proposition 3.1). Note that, without loss of generality, either device can specify payments only for savings in the efficient range $[100, 225]$ and rule out all other savings (formally, $p_{\mathbf{e}}^I(a) = p_{\mathbf{e}}^C(a) = +\infty$ for $a \notin [100, 225]$).

Now suppose the provider cannot observe t. If she offers the devices just described, at time 1 both types take the I-device (Proposition 4.1). C's expected utility from this device is about $m + 33$, which exceeds that from the C-device. This is because C predicts that he will oversave under the I-device, thereby incurring fewer penalties and enjoying more rewards than I. By contrast, I's expected utility from the C-device is about $m - 1.4$, which is lower than that from the I-device. Under either device, I expects to pay $p_{\mathbf{e}}^C$, but under the C-device he will undersave at time 2, which lowers his expected utility at time 1.

The provider wants to prevent C from gaming the system by taking the I-device. To do so, she can simply cut $p_{\mathbf{e}}^C$, offering the C-device at a discount d so that C will take it. The disadvantage of this approach is that the discount lowers the revenues from C. Moreover, simply cutting $p_{\mathbf{e}}^C$ by d cannot be optimal. The expected revenues from the I-device are maximal, so the provider is willing to modify this device if doing so allows her to reduce d.

To lower d, the provider has to make the I-device less beneficial for C. First of all, C benefits more than I from the rewards offered by $p_{\mathbf{e}}^I$ for high savings, as he is more likely to achieve them. Thus the provider could remove some of these rewards, at the cost of inducing I to undersave in some states. However, she can do more. C also benefits by avoiding penalties imposed by $p_{\mathbf{e}}^I$ for low

[11]Assume that m is large, so that consumption is always positive in what follows.

[12]All calculations for this example are available in Appendix C of the Supplemental Material.

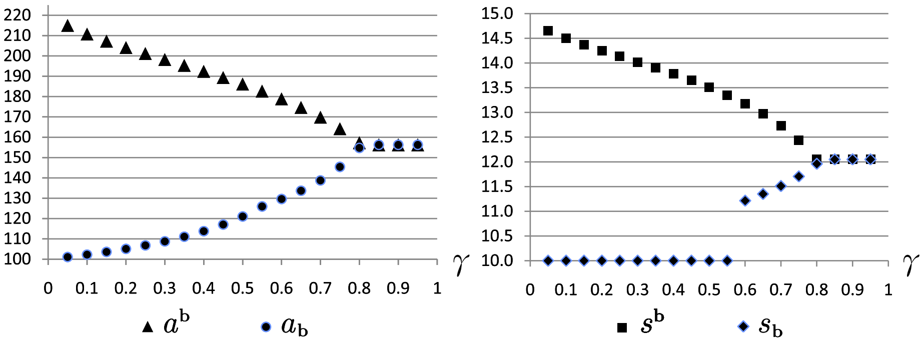

FIGURE 1.—Illustrative example: screening I-device.

savings. Thus the provider could also prevent I from saving some small amount, which requires large penalties for him to choose it only if s takes the smallest values. Overall, the provider will want to modify the I-device so as to curtail I's response to s (i.e., the flexibility afforded by the I-device) at both high levels of savings, which involve rewards, and low levels, which involve penalties (Theorem 4.1). These distortions allow her to limit the discount d, while minimizing the loss of surplus and hence revenues from the I-device.

In this example, the screening I-device induces I to save between some level $a_{\mathrm{b}} > 100$ and some level $a^{\mathrm{b}} < 225$, always responding to changes in $s \in (s_{\mathrm{b}}, s^{\mathrm{b}})$ with $10 \leq s_{\mathrm{b}} < s^{\mathrm{b}} < 15$ (Corollary 4.3). For $s \leq s_{\mathrm{b}}$, however, I always chooses a_{b}, and for $s \geq s^{\mathrm{b}}$ he always chooses a^{b} ("b" stands for "bunching"). Relative to $p^I_{\mathbf{e}}$, in this I-device penalties will grow faster as a falls below some $a_* > a_{\mathrm{b}}$, thus inducing I to oversave, and rewards will grow slower as a rises above some $a^* < a^{\mathrm{b}}$, thus inducing I to undersave (Corollary 4.2). I's savings range and bunching thresholds are represented in Figure 1 as a function of the probability γ that the agent's type is C.[13] As γ increases, the provider cares more about the revenues from C, and hence curtails the flexibility of the I-device to a greater extent, to a point where she removes it altogether.

An I-device without flexibility gives both types of agents the same expected utility, so no discount is needed to convince C to take the C-device. By contrast, as the I-device becomes more flexible and efficient (when $\gamma \downarrow 0$), it also becomes more attractive for C, and hence calls for a larger discount d on the C-device. Eventually, however, d becomes large enough that I would also take the C-device. Indeed, recall that C's expected utility from the efficient I-device is $m + 33$, so intuitively d approaches 33 as $\gamma \downarrow 0$. I's expected utility is, however, only 1.4 units lower than C's from *any* device with a constant payment and savings restricted to [100, 225]. Thus to ward off I, the provider has to fur-

[13]The discontinuity in s_{b} is due to specific properties of the uniform distribution (see Appendix B of the Supplemental Material).

ther reduce his expected utility from the C-device by making him undersave more. By restricting savings to $[100, 225]$, a C-device offers I some commitment in the form of a minimum savings. The provider can, however, allow for savings below 100 that only I would choose—and hence would not affect the choices of C and revenues. The C-device with the greatest deterrent has only one unused option—one that involves zero savings and payment $p^C - 100$ (the set-up fee minus a sort of refund for not using the device)—an option that I would choose for all $s < \frac{10}{0.9}$ (Proposition 4.3, Corollary 4.4). As calculated in Appendix C of the Supplemental Material, this option causes I's expected utility to fall by about 47 units relative to that from the C-device without unused options, which is more than enough to offset any discount $d \leq 33$.

2. MODEL

The model features two periods and a provider who supplies commitment devices to an agent.[14] The provider offers her devices at time 1 and maximizes a weighted sum of her profit and joint surplus (see below). A device consists of a menu of options (a, p), where $p \in \mathbb{R}$ is a payment and a is a contractible action in the feasible set $[\underline{a}, \overline{a}] \subset \mathbb{R}$ with $-\infty < \underline{a} < \overline{a} < +\infty$.[15] For each a, the provider incurs a cost at time 2, given by a function $c : [\underline{a}, \overline{a}] \to \mathbb{R}$ with derivatives that satisfy $c' \geq 0$ and $c'' \geq 0$. By assumption, the provider can fully commit to any device—which, if chosen at time 1, becomes binding for the agent—and at time 2 no third party can offer the agent contracts that might interfere with the provider's devices. Relaxing this assumption can undermine her ability to supply commitment devices and raises issues that, though important, are beyond the scope of this paper.

The agent may have time inconsistent preferences, in which case he is fully aware of his inconsistency (sophistication). Following Strotz (1956), the agent has two selves: *Self-1* lives at time 1 and chooses a device; *self-2* lives at time 2 and picks an option from the chosen device. The preferences of both selves depend on state s, which occurs at time 2. These states induce self-1 to desire flexibility (in the sense of Kreps (1979)) and are not contractible.[16] The distribution F of s is commonly known and has continuous, strictly positive density f over some interval $[\underline{s}, \overline{s}] \subset \mathbb{R}$ with $0 < \underline{s} < \overline{s} < +\infty$. Self-2's preference also depends on a scalar $t > 0$ that captures the agent's degree of time inconsistency (i.e., his type). Conditional on s, the direct utilities of self-1 and self-2 from action a are

$$u_1(a; s) = sb(a) - a \quad \text{and} \quad u_2(a; s, t) = tsb(a) - a,$$

[14] Section 6 discusses extensions of the model.

[15] Compactness of the set of feasible actions will be useful in the characterization of the optimal devices and the role of unused options in Section 4.

[16] If s were contractible, the agent could delegate his future choices to the provider and thus bypass his self-control problems.

where the function $b : [\underline{a}, \overline{a}] \to \mathbb{R}$ has derivatives that satisfy $b' > 0$ and $b'' < 0$. The corresponding total utilities are $u_1(a; s) - p$ and $u_2(a; s, t) - p$.

This formalization of time inconsistency captures in a tractable way a systematic shift (for $t \neq 1$) in how self-2 trades off a's direct costs and benefits. We can interpret this shift as arising from the timing of such costs and benefits.[17] In the example of Section 1.1, saving a reduces self-2's immediate consumption (by a) but yields benefits at retirement ($b(a)$). In this case, we typically have $t \leq 1$. Alternatively, a can capture *skipping* weekly workouts, which gives self-2 an immediate benefit ($b(a)$) but harms future health (by a). In this case, we typically have $t \geq 1$.

For the sake of clarity, in most of the paper the agent can be one of two types: C (consistent) has $t^C = 1$, and I (inconsistent) has $t^I \neq 1$. The agent knows t at time 1. The provider, by contrast, cannot observe t but knows the possible types and the probability $\gamma \in (0, 1)$ that the agent is of type C. At time 1, type j's overall utility from a device is the expected utility of self-2's ensuing decisions, computed using self-1's preference. If the agent rejects all devices, he gets the outside option, whose value is normalized to zero.[18]

As in other models with time inconsistency, the choice of a welfare criterion is delicate. Definition 2.1 follows the usual interpretation of self-1's preference as the agent's long-run preference.[19]

DEFINITION 2.1: In state s, the *surplus* is $u_1(a; s) - c(a)$, and the action that maximizes it, denoted by $\mathbf{e}(s)$, is called the *efficient outcome*.

For simplicity, assume that $u_1(\mathbf{e}(s); s) - c(\mathbf{e}(s)) > 0$ for all s. By the properties of u_1 and c, if the function $\mathbf{e}$ always takes interior values in $[\underline{a}, \overline{a}]$, it is strictly increasing. Such an $\mathbf{e}$ sets a clearer benchmark in terms of the efficient level of flexibility: efficiency always calls for different actions in different states. This leads to the following assumption.

ASSUMPTION 2.1: *For all s, $\mathbf{e}(s)$ is interior in $[\underline{a}, \overline{a}]$.*

Finally, at time 1, the provider designs her devices to maximize expected profit, weighted by $\sigma \in [0, 1]$, plus expected surplus, weighted by $1 - \sigma$. This is a convenient way to allow for a monopolist ($\sigma = 1$), as well as for a planner who may also have to worry about profitability ($\sigma < 1$). For example, the plan-

[17]This type of preference change is similar to that considered in β–δ discounting models (Phelps and Pollack (1968), Laibson (1997)) and is related to how DellaVigna and Malmendier (2004) modeled "investement" and "leisure" goods.

[18]The analysis can be extended to allow the value of the outside option to depend on the agent's type, without changing the paper's thrust (see Appendix D of the Supplemental Material).

[19]See, for example, O'Donoghue and Rabin (1999a) and DellaVigna and Malmendier (2004).

ner may exclusively provide commitment devices with limited funds to finance them—consider a government providing tax incentives for saving.[20]

3. OBSERVABLE TIME INCONSISTENCY

This section characterizes the provider's optimal devices when she can observe the agent's degree of inconsistency t, so as to better understand their properties when she cannot observe t. To do so, we can use direct mechanisms (DMs) that ensure that self-2 truthfully reports state s. Such a DM consists of an allocation $\boldsymbol{\alpha}^t : [\underline{s}, \overline{s}] \to [\underline{a}, \overline{a}]$ and a payment rule $\boldsymbol{\pi}^t : [\underline{s}, \overline{s}] \to \mathbb{R}$,[21] which must satisfy

$$\text{(IC)} \qquad u_2\big(\boldsymbol{\alpha}^t(s); s, t\big) - \boldsymbol{\pi}^t(s) \geq u_2\big(\boldsymbol{\alpha}^t\big(s'\big); s, t\big) - \boldsymbol{\pi}^t\big(s'\big) \quad \text{for all } s, s',$$

$$\text{(IR)} \qquad \int_{\underline{s}}^{\overline{s}} \big[u_1\big(\boldsymbol{\alpha}^t(s); s\big) - \boldsymbol{\pi}^t(s)\big] f(s)\, ds \geq 0.$$

Note that (IR) depends on self-1's preference, but (IC) depends on self-2's. Given these constraints, the provider maximizes

$$(1-\sigma) \int_{\underline{s}}^{\overline{s}} \big[u_1\big(\boldsymbol{\alpha}^t(s); s\big) - c\big(\boldsymbol{\alpha}^t(s)\big)\big] f(s)\, ds$$
$$+ \sigma \int_{\underline{s}}^{\overline{s}} \big[\boldsymbol{\pi}^t(s) - c\big(\boldsymbol{\alpha}^t(s)\big)\big] f(s)\, ds.$$

PROPOSITION 3.1—First Best: *If type t is observable, the optimal device sustains* $\mathbf{e}$ (*i.e.*, $\boldsymbol{\alpha}^t = \mathbf{e}$) *and yields the same expected profit, for any* $\sigma \in [0, 1]$ *and* $t > 0$.

Intuitively, since the agent has no private information at time 1, the provider can fully extract the utility that he expects from a device (i.e., *self-1*'s utility). For any σ, she then wants to maximize the expected surplus, which requires sustaining $\mathbf{e}$. Doing so is easy in standard models with only consistent agents: the provider can use a payment rule that will induce the agent to internalize the production cost, extracting the surplus with an entry fee. If the agent is inconsistent, however, she has to offer him tailored incentives ($\boldsymbol{\pi}^t$) so that *self-2* will comply with $\mathbf{e}$. In general, whether such incentives exist depends on the form

[20]Alternatively, we could let the planner maximize expected surplus subject to making some minimum profit. This would have the effect of making σ endogenous, without changing the thrust of the paper (the details are available upon request).

[21]As noted, the payments to the provider can be strictly negative. This is a key difference from Amador, Werning, and Angeletos (2006), Ambrus and Egorov (2013), and Bond and Sigurdsson (2013). Section 5.1 discusses this difference and its consequences.

of time inconsistency. In this model with $t > 0$, there is always a $\boldsymbol{\pi}^t$ that sustains $\mathbf{e}$, because self-2 prefers higher actions in higher states as prescribed by $\mathbf{e}$. But if t were negative, for example, then no $\boldsymbol{\pi}^t$ could sustain $\mathbf{e}$ with self-2.[22] This shows that in models with time inconsistency, even if information is symmetric and payments are unrestricted, efficiency may simply be infeasible.

To sustain $\mathbf{e}$, the incentives provided by a first-best device vary with t. Moreover, the first-best payment rule for $t \neq 1$ can be uniquely derived from the rule for the consistent type by adding a t-specific component $\mathbf{q}^t$ which offsets the agent's inconsistency. As this worsens (i.e., as t moves farther from 1), $\mathbf{q}^t$ decreases faster if $t < 1$—and increases faster if $t > 1$—when a increases.

COROLLARY 3.1: *Let $\boldsymbol{\pi}^t_{\mathbf{e}}$ be the first-best payment rule that sustains* $\mathbf{e}$ *with type t. Then, for $t \neq 1$, $\boldsymbol{\pi}^t_{\mathbf{e}}(s) = \boldsymbol{\pi}^1_{\mathbf{e}}(s) + \mathbf{q}^t(s)$, where*

$$\frac{d\boldsymbol{\pi}^1_{\mathbf{e}}(s)}{ds} = c'\big(\mathbf{e}(s)\big)\frac{d\mathbf{e}(s)}{ds} \quad \textit{and} \quad \frac{d\mathbf{q}^t(s)}{ds} = s(t-1)b'\big(\mathbf{e}(s)\big)\frac{d\mathbf{e}(s)}{ds}.$$

In words, $\boldsymbol{\pi}^1_{\mathbf{e}}$ induces C to internalize the cost c and hence choose efficiently. If $t^I < 1$, $\boldsymbol{\pi}^t_{\mathbf{e}}$ equals $\boldsymbol{\pi}^1_{\mathbf{e}}$ combined with payments $\mathbf{q}^t$ that fall in a—for example, a savings device with strictly increasing penalties for a below some $\hat{a}$ and rewards for a above $\hat{a}$ (Section 1.1). Similarly, if $t^I > 1$, $\boldsymbol{\pi}^t_{\mathbf{e}}$ equals $\boldsymbol{\pi}^1_{\mathbf{e}}$ combined with payments $\mathbf{q}^t$ that rise in a—for example, a gym plan with strictly increasing rewards for attended workouts or penalties for missed ones. To see this, note that $\frac{d\mathbf{q}^t(s)}{ds} < 0$ if and only if $t < 1$. By Proposition 3.1, expected revenues do not vary with t, hence $\int_{\underline{s}}^{\overline{s}} \mathbf{q}^t(s) f(s)\, ds = 0$ for all t. It follows that, for $t < 1$, $\mathbf{q}^t$ crosses zero only once—and does so from above, at some $\hat{s} \in (\underline{s}, \overline{s})$. Similarly, for $t > 1$, $\mathbf{q}^t$ crosses zero only once—and does so from below, at some $s' \in (\underline{s}, \overline{s})$.

By Proposition 3.1, the first-best devices induce efficient behaviors whose flexibility is invariant across types, hence no tension arises between commitment and flexibility. Moreover, profit maximization alone leads the provider to offer a full solution to the agent's inconsistency. And she is indifferent between trading with consistent agents and trading with inconsistent agents of any degree t, given their common time-1 preference. This generalizes a similar result of DellaVigna and Malmendier (2004), in which action a has only two feasible values (e.g., $a = 0$ and $a = 1$).

4. UNOBSERVABLE TIME INCONSISTENCY

4.1. *Adverse-Selection Problem*

When the provider cannot observe t, she has to design devices—one for each type j—that satisfy two incentive-compatibility conditions. First, j's self-1 must

[22] If $t < 0$, only decreasing allocations can be sustained with self-2.

select the device designed for j (the "*j-device*"). Second, given this device, j's self-2 must choose, in each state, the option designed for j to choose in that state.

As in models with only consistent agents (Myerson (1986)), we can analyze this problem using DMs that ensure that the agent will reveal his information sequentially: self-1's report corresponds to selecting a device, self-2's to choosing one of its options. But since here the agent can be inconsistent, we have to specify carefully what information self-2 *can* reveal. DMs in which self-1 reports t and self-2 can report only s entail a loss of generality. To see this, suppose s can take only two values. Then such DMs can describe only j-devices that contain at most two options, both chosen by j's self-2, but no third option that only another agent's self-2 would choose. To allow for such unused options while focusing on DMs, we must let self-2 report how his preference depends not only on s but also on t. In this way, given a time-1 report t^j, at time 2 there are enough messages for j's self-2 to choose his two options and another agent's self-2 to choose a third option through truthful reports.[23] Note that self-2's marginal valuation of a is pinned down by the product ts. Thus hereinafter let $\underline{v}^j = t^j\underline{s}$, $\overline{v}^j = t^j\overline{s}$, and $[\underline{v}, \overline{v}] = [\underline{v}^I, \overline{v}^C]$.

Without loss of generality, we can focus on DMs that assign a pair (a, p) to each sequential report of t and $v \in [\underline{v}, \overline{v}]$, and ensure that the agent truthfully reports t at time 1, and v at time 2 for any report on t. Formally, each DM is an array $(\mathbf{a}^j, \mathbf{p}^j)_{j=C,I}$, where $\mathbf{a}^j : [\underline{v}, \overline{v}] \to [\underline{a}, \overline{a}]$ is an allocation and $\mathbf{p}^j : [\underline{v}, \overline{v}] \to \mathbb{R}$ is a payment rule. Abusing notation, let $u_2(a; v) = u_2(a; s, t)$ for $v = ts$. When j reports t^j, denote j's expected utility at time 1 by $U^j(\mathbf{a}^j, \mathbf{p}^j)$, the expected profit by $\Pi^j(\mathbf{a}^j, \mathbf{p}^j)$, and the expected surplus by $W^j(\mathbf{a}^j)$. The provider's problem is

$$\mathcal{P} = \begin{cases} \max\limits_{(\mathbf{a}^j, \mathbf{p}^j)_{j=C,I}} \sigma\big[\gamma\Pi^C(\mathbf{a}^C, \mathbf{p}^C) + (1-\gamma)\Pi^I(\mathbf{a}^I, \mathbf{p}^I)\big] \\ \qquad + (1-\sigma)\big[\gamma W^C(\mathbf{a}^C) + (1-\gamma)W^I(\mathbf{a}^I)\big] \\ \text{s.t.,} \quad \text{for} \quad j = C, I \quad \text{and} \quad v, v' \in [\underline{v}, \overline{v}], \\ u_2(\mathbf{a}^j(v); v) - \mathbf{p}^j(v) \geq u_2(\mathbf{a}^j(v'); v) - \mathbf{p}^j(v'), & (\mathrm{IC}_2^j) \\ U^j(\mathbf{a}^j, \mathbf{p}^j) \geq U^j(\mathbf{a}^{-j}, \mathbf{p}^{-j}), & (\mathrm{IC}_1^j) \\ U^j(\mathbf{a}^j, \mathbf{p}^j) \geq 0. & (\mathrm{IR}^j) \end{cases}$$

The key to understanding the adverse-selection problem at the heart of the screening of time inconsistency is constraint (IC_1^j), which captures j's incentives to choose between devices at time 1. To study these incentives, write (IC_1^j) as

$$U^j(\mathbf{a}^j, \mathbf{p}^j) - U^{-j}(\mathbf{a}^{-j}, \mathbf{p}^{-j}) \geq U^j(\mathbf{a}^{-j}, \mathbf{p}^{-j}) - U^{-j}(\mathbf{a}^{-j}, \mathbf{p}^{-j}).$$

[23]For further details, see Proposition 4.3 and Section 5.2.

This expression says that if the difference on the right-hand side is strictly positive, then j's expected payoff from the mechanism must exceed $-j$'s (i.e., j enjoys information rents). That difference captures whether, at time 1, j expects to get a larger payoff than does $-j$ if he mimics $-j$. Thus we need to study C's and I's payoffs at time 1 when they choose the *same* device.

For any device, at time 1 C expects a weakly larger payoff than I, and a strictly larger one if and only if the device provides some flexibility.

PROPOSITION 4.1: *If* $(\mathbf{a}^j, \mathbf{p}^j)_{j=C,I}$ *satisfies* (IC_2^j), *then* $U^C(\mathbf{a}^j, \mathbf{p}^j) \geq U^I(\mathbf{a}^j, \mathbf{p}^j)$ *for* $j = C, I$, *with equality if and only if* $\mathbf{a}^j$ *is constant over* $(\underline{v}, \overline{v})$.[24]

This result has two implications. First, C is the "high" type in this model, as he values any device more than I. Second, I's demand for flexibility is a key determinant of the adverse-selection problem: if I demanded a device with only one option (no flexibility), C could not benefit by mimicking I, and so the provider would not have to grant C any information rents.

To see the intuition behind Proposition 4.1, fix a device $(\mathbf{a}^j, \mathbf{p}^j)$ that satisfies (IC_2^j), so that we know self-2's choices in each state. Recall that C's self-2 always chooses the best option from self-1's viewpoint, but I's self-2 may not. Since the two selves-1 have the same preference, C's self-1 must then be at least as well off as I's. Moreover, he must be strictly better off if C and I prefer and choose different options at time 2 with strictly positive probability. This always happens—as shown in the proof—unless $\mathbf{a}^j$ defines a device with only one option. The proof assumes only that either $t^I < t^C \leq 1$ or $t^I > t^C \geq 1$. Thus what drives the result is that C's ex post preference is "closer" to the common ex ante preference than is I's ex post preference.

We can now reduce problem $\mathcal{P}$ to simple conditions that characterize the optimal devices in terms of $\mathbf{a}^C$ and $\mathbf{a}^I$ only. By standard arguments, (IC_2^j) holds if and only if $\mathbf{a}^j$ is increasing and, for every $v \in [\underline{v}, \overline{v}]$,

$$\text{(E)} \qquad \mathbf{p}^j(v) = u_2\big(\mathbf{a}^j(v); v\big) + \int_v^{\overline{v}} b\big(\mathbf{a}^j(x)\big)\,dx + k^j,$$

with $k^j \in \mathbb{R}$. Since C's expected payoff must exceed I's, (IR^C) always holds. And since payoffs decrease and profits increase with payments, at the optimum both (IR^I) and (IC_1^C) must bind, as usual. These constraints then pin down k^C and k^I, for every $\mathbf{a}^C$ and $\mathbf{a}^I$.[25]

[24]At $\underline{v}$ and $\overline{v}$, $\mathbf{a}^j$ can jump without making $U^C(\mathbf{a}^j, \mathbf{p}^j) > U^I(\mathbf{a}^j, \mathbf{p}^j)$, simply because F is atomless. Since this indeterminacy has no economic content, hereinafter the paper focuses on the extension of $\mathbf{a}^j$ to $[\underline{v}, \overline{v}]$ by continuity whenever possible.

[25]When $\sigma = 0$, (IR^I) and (IC_1^C) can be slack at the optimum. However, assuming that they hold with equality entails no loss of generality as far as characterizing $\mathbf{a}^I$ and $\mathbf{a}^C$ is concerned.

However, nothing guarantees that (IC_1^I) is always slack. Here, I (the "low" type) may prefer the C- to the I-device, even though C (the "high" type) is indifferent between them (see Section 4.3). If we define[26]

$$R^{-j}(\mathbf{a}^j) = U^{-j}(\mathbf{a}^j, \mathbf{p}^j) - U^j(\mathbf{a}^j, \mathbf{p}^j),$$

then (IC_1^I) holds if and only if

$$\text{(R)} \qquad -R^I(\mathbf{a}^C) \geq R^C(\mathbf{a}^I).$$

That is, a dishonest I must expect to lose, relative to C's payoff, at least as much as a dishonest C expects to gain, relative to I's payoff.[27]

Finally, using $\Pi^j(\mathbf{a}^j, \mathbf{p}^j) = W^j(\mathbf{a}^j) - U^j(\mathbf{a}^j, \mathbf{p}^j)$, we obtain the following.

COROLLARY 4.1: *A mechanism* $(\mathbf{a}^j, \mathbf{p}^j)_{j=C,I}$ *solves* $\mathcal{P}$ *if and only if* $\mathbf{a}^C$ *and* $\mathbf{a}^I$ *solve*

$$\mathcal{P}' = \begin{cases} \max\limits_{\mathbf{a}^C, \mathbf{a}^I} \gamma W^C(\mathbf{a}^C) + (1-\gamma)\left[W^I(\mathbf{a}^I) - \sigma \dfrac{\gamma}{1-\gamma} R^C(\mathbf{a}^I)\right] \\ s.t. \quad \mathbf{a}^C, \mathbf{a}^I \textit{ are increasing} \quad \textit{and} \quad \text{(R)} \textit{ holds}. \end{cases}$$

Because of (R), $\mathcal{P}'$ is not separable across allocations, and hence we cannot solve for $\mathbf{a}^C$ independently of $\mathbf{a}^I$.

Nonetheless, we can characterize the properties of $\mathbf{a}^C$ and $\mathbf{a}^I$ by solving two distinct maximizations. Indeed, by Lemma A.1 in Appendix A, $(\mathbf{a}^C, \mathbf{a}^I)$ solves $\mathcal{P}'$ if and only if each solves

$$\text{(O)} \qquad \max_{\hat{\mathbf{a}}^j \text{ increasing}} W^j(\hat{\mathbf{a}}^j) - r^{-j} R^{-j}(\hat{\mathbf{a}}^j),$$

where

$$r^C = \sigma \frac{\gamma}{1-\gamma} + \frac{\mu}{1-\gamma} \quad \text{and} \quad r^I = \frac{\mu}{\gamma}$$

for some $\mu \geq 0$ such that $\mu[R^C(\mathbf{a}^I) + R^I(\mathbf{a}^C)] = 0$. Using (O), Sections 4.2 and 4.3 will characterize $\mathbf{a}^I$ and $\mathbf{a}^C$, each as a function of r^{-j}. As usual, r^C depends on the hazard rate between types, scaled by the profit weight σ. Moreover, r^C and r^I depend on a new term that captures whether (R) binds, as μ is a Lagrange multiplier associated with (R). This term links the maximizations that define $\mathbf{a}^C$ and $\mathbf{a}^I$, thereby reducing the nonseparability of $\mathcal{P}'$ to μ. This simplification relies on Lagrangian methods for infinite-dimensional policy spaces

[26] R^{-j} depends only on $\mathbf{a}^j$, because k^j enters additively in $U^i(\mathbf{a}^j, \mathbf{p}^j)$ (see the proof of Proposition 4.1).

[27] Type j is dishonest if at time 1 he reports a type different from j.

that globally identify the solutions to $\mathcal{P}'$ as saddle points of a Lagrangian function (Luenberger (1969)), which here is separable in $\mathbf{a}^C$ and $\mathbf{a}^I$. Combining the results in Sections 4.2 and 4.3, Section 4.4 will characterize μ.

4.2. *Optimal Device for Inconsistent Agents*

This section shows how the provider distorts the I-device, by curtailing flexibility at both ends of the efficient choice range, so as to limit C's information rents. For the sake of clarity, it analyzes the case where $t^I < 1$ and explains, at the end, why none of the results changes qualitatively in the case where $t^I > 1$.

By condition (O), designing the I-device involves a standard trade-off, which here leads to curtailment of flexibility. On the one hand, the provider wants to maximize the expected surplus with I, just as in the first-best I-device; this calls for a device with the efficient level of flexibility ($\mathbf{e}$). On the other hand, she wants to minimize C's rents, because they reduce profits and can induce I to mimic C (recall (R)); this calls for a device with no flexibility (Proposition 4.1). The optimal device should then strike a balance between these two polar cases, and hence flexibility should be curtailed below the efficient level. To gain greater intuition about this, recall that C's rents arise because, in each state, C and I have different valuations v and consequently behave differently under an efficient I-device. Thus curbing C's rents requires curtailing this difference in behavior, which depends on how $\mathbf{a}^I$ varies with v (i.e., its flexibility).

Even though a similar trade-off arises in standard screening models, those models offer no guidelines on how to optimally curtail flexibility. In Mussa and Rosen (1978), for instance, a seller has to lower the quality offered to low-valuation buyers below the efficient level, so as to extract more rents from high-valuation buyers. However, curtailing flexibility, in contrast to lowering quality, can be done in many ways, and the optimal one is not obvious.

As Theorem 4.1 below shows, relative to the first-best I-device, the optimal I-device curtails the flexibility of I's behavior as follows. First, I reacts less to extreme states (both high and low), so that his range of choices is a proper subset of the efficient one. Second, I does not react at all to states at the high end—and, in some environments, to states at the low end either; for such states, there is no flexibility at all. Denote the allocation that defines the first-best I-device (up to k^I) by $\mathbf{a}^I_{fb}$, with $\mathbf{a}^I_{fb}(v) = \mathbf{e}(v/t^I)$ for $v \in [\underline{v}^I, \overline{v}^I]$ and $\mathbf{a}^I_{fb}(v) = \mathbf{a}^I_{fb}(\overline{v}^I)$ otherwise.

THEOREM 4.1: *A solution $\mathbf{a}^I_{sb}$ to* (O) *exists*, *is unique*, *and is continuous in v and r^C. If $r^C > 0$, $\mathbf{a}^I_{sb}$ has the following properties*:

(a) Range reduction with "overconsumption" at the bottom and "underconsumption" at the top: *there are $v_* > \underline{v}^I$ and $v^* < \overline{v}^C$ such that $\mathbf{a}^I_{sb}(v) = \mathbf{a}^I_{fb}(v)$ at v_* and v^*, $\mathbf{a}^I_{sb}(v) > \mathbf{a}^I_{fb}(v)$ for $v < v_*$, and $\mathbf{a}^I_{sb}(v) < \mathbf{a}^I_{fb}(v)$ for $v > v^*$*;

(b) No flexibility at the top: *there is $v^{\mathrm{b}} < \overline{v}^I$ such that $\mathbf{a}^I_{sb}$ is constant over the interval $[v^{\mathrm{b}}, \overline{v}^C]$*;

(c) No flexibility at the bottom: *there exist distributions F and values of* t^I *and* r^C *such that* $\mathbf{a}^I_{sb}$ *is constant over the interval* $[\underline{v}^I, v_b]$, *where* $v_b > \underline{v}^I$. *A sufficient condition is that, for some* $s^\dagger > \underline{s}$ *and all* $s' > s$ *in* $[\underline{s}, s^\dagger]$,

$$\frac{F(s')/f(s') - F(s)/f(s)}{s' - s} \geq \frac{1}{t^I}\left(1 + \frac{1}{r^C}\right) - 1.$$

The proof is constructive. It first builds $\mathbf{a}^I_{sb}$ over $[\underline{v}^I, \overline{v}^I]$, relying on Toikka's (2011) generalization of Myerson's (1981) ironing method. It then explicitly builds the best extension of $\mathbf{a}^I_{sb}$ off path. Note that the condition that implies property (c) is more likely to hold when I is less inconsistent (t^I closer to1) and the provider cares more about C's rents (higher r^C). For instance, assuming that f is differentiable with bounded f' on $(\underline{s}, \overline{s})$, one can show that in the monopolist ($\sigma = 1$) case property (c) arises if $t^I > \frac{1}{2}$ and $\gamma > \frac{1}{2t^I}$.[28]

To gain some intuition behind Theorem 4.1, suppose the provider offers savings devices. Starting from $\mathbf{a}^I_{fb}$, focus on the problem of designing $\mathbf{a}^I(v)$ for a fixed v. Changing $\mathbf{a}^I(v)$ has a local effect on C's rents, which calls for increasing it, and a global effect, which calls for decreasing it. To see the first effect, note that, under $\mathbf{a}^I_{fb}$, C saves strictly more than I in state $s = v/t^I$ since $s > v$.[29] This difference causes C's rents and hence calls for increasing $\mathbf{a}^I_{fb}(v)$ toward $\mathbf{a}^I_{fb}(s)$. However, doing so raises $\mathbf{p}^I(v')$ for all $v' < v$, as $\mathbf{a}^I(v)$ appears in the integral in condition (E). This global effect boosts C's rents. To see why, recall that an efficient I-device involves penalties that prevent I's self-2 from saving $\mathbf{a}^I_{fb}(v')$ when his valuation is higher than v' (Corollary 3.1). This means that under $\mathbf{a}^I_{fb}$, in state $s' = v'/t^I$, I incurs a penalty which C avoids since $s' > v'$. Thus C benefits in state s' by avoiding this penalty, and raising it via $\mathbf{p}^I(v')$ enables him to benefit even more, boosting his rents. The relative strengths of these local and global effects ultimately determines whether $\mathbf{a}^I_{fb}(v)$ is distorted up or down.

To see why "overconsumption" and no flexibility arise at the bottom, consider Figure 2. For v close to $\underline{v}^I$, the local effect prevails over the global effect, because the probability that $v' < v$ is small. Thus it is optimal to distort $\mathbf{a}^I_{fb}$ upwards. As v rises, the global effect strengthens, shrinking the upward distortion until v reaches v_*, where it offsets the local effect and $\mathbf{a}^I(v_*) = \mathbf{a}^I_{fb}(v_*)$ (e.g., curve 1). Intuitively, how fast this happens depends on how fast $F^I(v)/f^I(v)$ grows.[30] If this ratio grows fast enough close to $\underline{v}^I$—as formally stated in (c)—the provider may want to distort $\mathbf{a}^I$ much more at some v than at some $v' > v$, leading to $\mathbf{a}^I(v') < \mathbf{a}^I(v)$ (e.g., curve 2). Since $\mathbf{a}^I$ must be increasing, however, she has to find a compromise and pool v's close to $\underline{v}^I$ (e.g., curve 3).

[28]As the example in Section 1.1 shows, there exist settings in which $\mathbf{a}^I_{sb}$ does not have property (c) (see also Corollary 4.3 and its proof).

[29]Recall that the valuation of C's self-2 coincides with s.

[30]Note that $F^I(v)/f^I(v)$ equals $t^I F(v/t^I)/f(v/t^I)$.

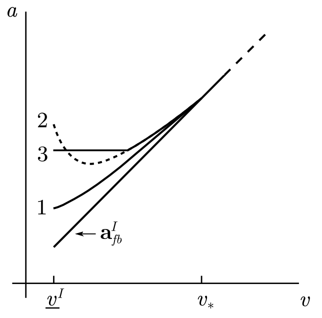

FIGURE 2.—"Overconsumption" and no flexibility at the bottom.

To see why "underconsumption" and no flexibility arise at the top, first note that $\mathbf{a}^I$ must be constant above $\overline{v}^I$. Options with $a > \mathbf{a}^I(\overline{v}^I)$ are useless for I but give a dishonest C more possibilities to save differently from I, thereby boosting his rents. Consequently, at $\overline{v}^I$ only the global effect matters, and it is optimal to distort $\mathbf{a}^I$ downwards. By the same argument as before, these downward distortions persist until v reaches v^*, where again $\mathbf{a}^I(v^*) = \mathbf{a}^I_{fb}(v^*)$. Second, suppose that no bunching occurs at the top and $\mathbf{a}^I$ is strictly increasing near $\overline{v}^I$. To see that $\mathbf{a}^I$ cannot be optimal, consider picking $v^{\mathrm{b}} < \overline{v}^I$ and lowering $\mathbf{a}^I(v)$ to $\mathbf{a}^I(v^{\mathrm{b}})$ for every $v > v^{\mathrm{b}}$. If v^{b} is close to $\overline{v}^I$, doing so causes a first-order drop in C's rents, but only a second-order welfare loss with I: it shrinks the gap between I's savings and a dishonest C's savings for every s in $[v^{\mathrm{b}}, \overline{s}]$, but it further distorts I's savings only for s in $[v^{\mathrm{b}}/t^I, \overline{s}]$.

To explain how the provider sustains I's distorted behavior, Corollary 4.2 below compares the payment rules associated with $\mathbf{a}^I_{sb}$ and $\mathbf{a}^I_{fb}$. At the bottom of $[\underline{v}^I, \overline{v}^I]$, $\mathbf{p}^I_{sb}$ either falls more slowly or rises more rapidly than $\mathbf{p}^I_{fb}$ for an equal decrease in a; this renders I less willing to decrease a, causing "overconsumption." At the top of $[\underline{v}^I, \overline{v}^I]$, $\mathbf{p}^I_{sb}$ either falls more slowly or rises more rapidly than $\mathbf{p}^I_{fb}$ for an equal increase in a; this renders I less willing to increase a, causing "underconsumption."

COROLLARY 4.2: *If* $d\mathbf{a}^I_{sb}/dv > 0$ *at* v, *then* $\frac{d\mathbf{p}^I_{sb}/dv}{d\mathbf{a}^I_{sb}/dv}$ *is strictly smaller than* $\frac{d\mathbf{p}^I_{fb}/dv}{d\mathbf{a}^I_{fb}/dv}$ *for* $v \in [\underline{v}^I, v_*)$, *and strictly larger for* $v \in (v^*, \overline{v}^I]$.

Proposition 4.2 below shows how the optimal I-device changes as r^C becomes very large or very small. Let a^{nf} be the ex ante efficient action if the agent cannot act on ex post information:

$$a^{\mathrm{nf}} = \arg\max_{a \in [\underline{a}, \overline{a}]} \mathbb{E}\big[u_1(a; s)\big] - c(a).$$

Then if the provider cares little about C's rents, she tends to offer an efficient I-device. In the opposite case, she tends to disregard I's desire for flexibility, in the limit offering a device with only a^{nf}—a radical reduction of flexibility vis-à-vis the first-best I-device.

PROPOSITION 4.2: *The allocation* $\mathbf{a}^I_{sb}$ *converges pointwise to* $\mathbf{a}^I_{fb}$ *as* $r^C \to 0$ *and uniformly to* a^{nf} *as* $r^C \to +\infty$.

Recall that r^C is increasing in the probability γ that the agent is of type C and the profit weight σ; if $\sigma > 0$, then $r^C \to +\infty$ as $\gamma \to 1$.

With more information on the distribution F, we can say more on the distortions in the I-device. This is because the virtual valuation that defines $\mathbf{a}^I_{sb}$ is in general complicated. For purposes of illustration, Corollary 4.3 considers uniform distributions. In this case, the range of choices of the I-device shrinks—and the no-flexibility regions $[\underline{v}^I, v_{\text{b}}]$ and $[v^{\text{b}}, \overline{v}^C]$ grow monotonically—as r^C rises (recall Figure 1).

COROLLARY 4.3: *Suppose* F *is uniform and* $t^I > \underline{s}/\overline{s}$. *Then* $\mathbf{a}^I_{sb}$ *crosses* $\mathbf{a}^I_{fb}$ *only once and is strictly increasing on* $[v_{\text{b}}, v^{\text{b}}]$. *As* r^C *rises*, v^{b} *and* $\mathbf{a}^I_{sb}(v^{\text{b}})$ *decrease and* $\mathbf{a}^I_{sb}(v_{\text{b}})$ *increases*; *when* $v_{\text{b}} > \underline{v}^I$, v_{b} *increases*.

Corollary 4.3 also highlights that the bunching at the top and bottom is due not to failures of standard regularity conditions—satisfied by uniform distributions—but precisely to the problem of optimally screening time inconsistency.

Case with $t^I > 1$. In this case, I's self-2 tends to overconsume a. As a result, the local and global effects described before work in the opposite direction relative to the case with $t^I < 1$. Under $\mathbf{a}^I_{fb}$, in state $s = v/t^I$, C consumes strictly less than I since $s < v$. Thus the local effect pushes $\mathbf{a}^I_{fb}(v)$ to decrease toward $\mathbf{a}^I_{fb}(s)$. On the other hand, decreasing $\mathbf{a}^I(v)$ causes $\mathbf{p}^I(v')$ to rise for v' *above* v.[31] In the first-best I-device, however, $\mathbf{p}^I_{fb}(v')$ also involves a penalty to prevent I's self-2 from consuming $\mathbf{a}^I_{fb}(v')$ when his valuation is below v'. Again, since C benefits by avoiding this penalty, the global effect now pushes $\mathbf{a}^I(v)$ to increase. Intuitively, the local effect prevails at the top of $[\underline{v}^I, \overline{v}^I]$, while the global effect dominates at the bottom. Thus the optimal way to curtail flexibility is qualitatively the same as in Theorem 4.1.

4.3. *Optimal Device for Consistent Agents*

This section shows that, to deter I from mimicking C, the provider may have to add unused options to an otherwise efficient C-device. If these options do not constitute a sufficient deterrent, she may also distort C's behavior. We will again analyze formally only the case with $t^I < 1$.

[31]By the standard arguments, $\mathbf{p}^I(v') = u_2(\mathbf{a}^I(v'), v') - \int_{\underline{v}}^{v'} b(\mathbf{a}^I(s))\, ds + \hat{k}^I$ for all v'.

By condition (O), when designing the C-device, the provider wants to maximize the surplus with C, but she also has to worry about jeopardizing I's incentives for revealing his inconsistency (recall condition (R)). A natural benchmark of an efficient C-device is the first-best one, defined (up to k^C) by $\mathbf{a}_{fb}^C$ with $\mathbf{a}_{fb}^C(v) = \mathbf{e}(v)$ for $v \in [\underline{v}^C, \overline{v}^C]$ and $\mathbf{a}_{fb}^C(v) = \mathbf{a}_{fb}^C(\underline{v}^C)$ otherwise. However, $\mathbf{a}_{fb}^C$ together with some $\mathbf{a}^I$ can violate (R).

By Lemma A.7 in Appendix A, $\mathbf{a}_{fb}^C$ and $\mathbf{a}_{fb}^I$ violate (R) and are therefore infeasible for a family $\mathcal{F}$ of distributions F. To see why, recall that, under the C-device, I loses relative to C's payoff to the extent that his behavior differs from C's. But the first-best C-device limits this difference, by featuring a minimum action which grants I some commitment: given $\mathbf{a}_{fb}^C$, for s close to $\underline{s}$, I's self-2 chooses $\mathbf{a}_{fb}^C(\underline{v}^C)$, which is close to C's choice $\mathbf{a}_{fb}^C(s)$. Thus if states close to $\underline{s}$ are likely enough, then ex ante I values the C-device almost as much as C. On the other hand, since $\mathbf{a}_{fb}^I$ is flexible, C's rents are positive and possibly large enough to induce I to mimic C. One can show, for example, that if $b(a) = \sqrt{a}$ and F is uniform, then $\mathbf{a}_{fb}^C$ and $\mathbf{a}_{fb}^I$ are infeasible (recall Section 1.1). Also, if $F \in \mathcal{F}$ and first-order stochastically dominates $\widetilde{F}$, then $\widetilde{F} \in \mathcal{F}$.[32]

Lemma A.7 focuses on $\mathbf{a}_{fb}^C$ and $\mathbf{a}_{fb}^I$, but its conclusion holds more generally. By Proposition 4.2, as $r^C \to +\infty$, $\mathbf{a}_{sb}^I$ converges to a constant allocation and hence C's rents vanish; thus by continuity, $\mathbf{a}_{fb}^C$ and $\mathbf{a}_{sb}^I$ are always feasible for sufficiently large r^C. However, as $r^C \to 0$, $\mathbf{a}_{sb}^I$ converges to $\mathbf{a}_{fb}^I$. Then by continuity, if $\mathbf{a}_{fb}^C$ and $\mathbf{a}_{fb}^I$ are infeasible, so are $\mathbf{a}_{fb}^C$ and $\mathbf{a}_{sb}^I$ for sufficiently small r^C. In this case, the question is how to optimally change $\mathbf{a}_{fb}^C$ so as to avoid I's mimicking C.

The first strategy is unconventional. In models with only consistent agents, the provider must distort the offer for one type whenever it induces another type to mimic. Here, by contrast, she may be able to avoid I's mimicking and at the same time offer an efficient C-device. To do this, she adds to the device options that C never uses but causes the behavior of a dishonest I to deviate even more from C's behavior, and hence to incur greater losses relative to C's payoff. Intuitively, these options remove the commitment that the first-best C-device gives a dishonest I. Being off path, they can be usefully employed in many ways. Proposition 4.3 shows the way which maximally deters I's mimicking: it involves only one unused option with action $\underline{a}$ that I would choose in sufficiently low states.

PROPOSITION 4.3: *An increasing $\mathbf{a}_{un}^C$ sustains $\mathbf{e}$ with C and maximizes $-R^I(\mathbf{a}^C)$ if and only if $\mathbf{a}_{un}^C(v) = \underline{a}$ for $v < v_{\mathrm{u}}$ and $\mathbf{a}_{un}^C(v) = \mathbf{a}_{fb}^C(v)$ for $v \geq v^{\mathrm{u}}$, where $\underline{v}^I < v_{\mathrm{u}} \leq v^{\mathrm{u}} \leq \underline{v}^C$.*

[32]This follows from the proof of Lemma A.7.

To see the intuition behind this, consider the unused option intended for a given $v < \underline{v}^C$. Once more, a local effect calls for decreasing $\mathbf{a}^C(v)$, but a global effect calls for increasing it. If we lower $\mathbf{a}^C(v)$, I's self-1 will view it as a worse temptation for his self-2 with valuation v, who always prefers lower actions ($t^I < 1$). This local effect favors setting $\mathbf{a}^C(v) = \underline{a}$. To do so, however, we have to render the unused options less expensive for every $v' < v$, and hence less of a deterrent for I's self-1: by condition (E), if we lower $\mathbf{a}^C(v)$, we must also lower $\mathbf{p}^C(v')$ for all $v' < v$. Thus this global effect favors raising $\mathbf{a}^C(v)$. Clearly, for v close to $\underline{v}^I$, the local effect prevails, as the probability that $v' < v$ is small, so we will set $\mathbf{a}^C$ equal to $\underline{a}$. This explains why we always have $v_{\mathrm{u}} > \underline{v}^I$. As v rises, the global effect strengthens, and, depending on F, it may force $\mathbf{a}^C$ to be above $\underline{a}$. When this happens at some $v^{\mathrm{u}} < \underline{v}^C$, the multiplicative structure of the agent's payoff causes $\mathbf{a}^C$ to jumps to $\mathbf{a}^C_{fb}(\underline{v}^C)$. Thus Proposition 4.3 shows that, to better deter I's mimicking, it may actually be useful to *restrict* the set of states in which I's self-2 would choose unused options from the C-device.

These options render I less willing to mimic C, but to ward off I altogether, they must correspond to a sufficiently high level of temptation. If they do, the provider will sustain $\mathbf{e}$ with C.

COROLLARY 4.4: *If $\mathbf{a}^C_{fb}$ and $\mathbf{a}^I_{sb}$ are infeasible, there is $\delta > 0$ such that $\mathbf{a}^C_{un}$ and $\mathbf{a}^I_{sb}$ are feasible if and only if $b(\underline{a}) \leq b(\mathbf{e}(\underline{s})) - \delta$.*

Action $\underline{a}$ can be tempting enough, relative to $\mathbf{e}(\underline{s})$, for several reasons. For instance, suppose $b(a)$ captures future consequences of some current action (such as splurging). Then $b(\underline{a})$ is likely to be far worse than $b(\mathbf{e}(\underline{s}))$, which takes into account the current utility $(-a)$ and the cost $c(a)$ (such as default costs). Section 1.1 provides an example where this holds.

Finally, when unused options alone cannot avoid I's mimicking C, the provider has to distort $\mathbf{a}^C$ on path. She does so to make I's behavior under the C-device differ even more from C's. This is the only way to make a dishonest I lose even more relative to C's payoff and thus satisfy condition (R). The optimal $\mathbf{a}^C$ now maximizes $W^C(\hat{\mathbf{a}}^C) - r^I R^I(\hat{\mathbf{a}}^C)$ with $r^I > 0$, which we can write as a virtual surplus to characterize $\mathbf{a}^C$. The resulting device, in general, can distort C's behavior both up and down relative to efficiency, in part also to cause the unused option $(\underline{a}, \underline{p})$ to appear to be an even worse temptation for I's self-1.

Case with $t^I > 1$. When I's self-2 tends to overconsume a, Proposition 4.3 and Corollary 4.4 change as follows: the unused option in the C-device that corresponds to the strongest deterrent now features $\overline{a}$ (rather than $\underline{a}$), because I's self-1 views $\overline{a}$ as the most tempting action for his self-2. Thus if $\mathbf{a}^C_{fb}$ and $\mathbf{a}^I_{sb}$ are infeasible, unused options allow the provider to sustain $\mathbf{e}$ with C and $\mathbf{a}^I_{sb}$ if and only if $b(\overline{a}) \geq b(\mathbf{e}(\overline{s})) + \hat{\delta}$ for a certain $\hat{\delta} > 0$.

4.4. *Overall Optimal Mechanism*

Sections 4.2 and 4.3 characterize the optimal devices for all r^C and r^I, which depend on the exogenous type distribution and profit weight (γ and σ, respectively) and the endogenous Lagrange multiplier μ. We can now characterize μ and its dependence on γ and σ. If unused options are always enough to deter I from mimicking C, then condition (R) never binds and $\mu = 0$. In this case, the provider always sustains an efficient $\mathbf{a}^C$ and $\mathbf{a}^I_{sb}$ with $r^C = \sigma\frac{\gamma}{1-\gamma}$. Thus suppose hereinafter that sustaining $\mathbf{e}$ with both types is infeasible, even when relying on unused options.

THEOREM 4.2: *Suppose $R^C(\mathbf{a}^I_{fb}) > -R^I(\mathbf{a}^C_{un})$. There exist r_1 and r_2, with $0 < r_1 < r_2 < +\infty$, such that the optimal C-device sustains $\mathbf{e}$ with C if and only if $\sigma\frac{\gamma}{1-\gamma} \geq r_1$, and must include unused options if and only if $\sigma\frac{\gamma}{1-\gamma} < r_2$. In all cases, the optimal I-device sustains $\mathbf{a}^I_{sb}$.*

Therefore, screening time inconsistency violates the "no distortion at the top" property—a property that is common in standard screening models. In those models, agents of the "highest" type always achieve efficient outcomes.[33]

The C-device is more likely to feature unused options and to be inefficient when type C is less likely (lower γ) or the provider cares less about profit (lower σ). In these cases, she is willing to grant C larger rents. Such rents, however, render I more willing to mimic C, and so require lowering the degree of commitment that I finds in the C-device. When $\sigma\frac{\gamma}{1-\gamma} < r_1$, condition (R) binds, and μ depends on σ and γ as follows:

PROPOSITION 4.4: *Let r_1 be as in Theorem 4.2. Then $\mu(\gamma, \sigma)$ is a continuous, bounded function and is strictly positive if and only if $\sigma\frac{\gamma}{1-\gamma} < r_1$. In this case, $\mu(\gamma, \sigma)$ is strictly decreasing in σ. Given $\sigma \in [0, 1]$, $\mu(\gamma, \sigma) \downarrow 0$ as either $\gamma \downarrow 0$ or $\gamma \uparrow \gamma^* = \frac{r_1}{\sigma + r_1}$.*

We can interpret μ as capturing how costly it is for the provider to satisfy (R). When $\sigma\frac{\gamma}{1-\gamma} < r_1$, she will sustain inefficient outcomes with both C and I. Consequently, the more she cares about welfare, the greater the increase in μ; moreover, $\mathbf{a}^I$ becomes *more* efficient, but $\mathbf{a}^C$ *less* efficient. To see this, note that by (O), $W^j(\mathbf{a}^j)$ is decreasing in r^{-j} at the optimum. (For the effects of r^C on $\mathbf{a}^I$, also recall Corollary 4.3.) Clearly, as σ falls, $r^I = \frac{\mu}{\gamma}$ rises. On the other hand, r^C cannot rise. Indeed, the optimal allocations for σ are feasible, and hence candidate solutions for $\hat{\sigma} < \sigma$. However, if the provider cares more about welfare, it cannot be optimal to design allocations both of which are less efficient.

[33]See, for example, Mussa and Rosen (1978), Courty and Li (2000), Battaglini (2005), and Pavan, Segal, and Toikka (2014).

Also note that, for $\sigma = 0$, $\mathbf{a}^C$ and $\mathbf{a}^I$ are inefficient for any nondegenerate distribution of types.

Now consider γ. When γ is small, so is μ, because the provider cares little about welfare with C and hence about distorting $\mathbf{a}^C$; moreover, she cares little about C's rents and hence sustains an almost efficient $\mathbf{a}^I$. On the other hand, when $\gamma \uparrow \gamma^*$ she cares relatively more about her payoff with C. Hence, she distorts $\mathbf{a}^I$ and reduces C's rents to the point where unused options are almost enough to prevent I's mimicking. She then has to distort $\mathbf{a}^C$ by only a little, so μ is again small. Finally, μ is highest for intermediate values of $\gamma < \gamma^*$, because the provider cares enough about her payoff with either type. Thus she is not willing to distort $\mathbf{a}^I$ by much, but then she has to distort $\mathbf{a}^C$ by more.

4.5. *Extension: Many Degrees of Inconsistency*

Consider the model in Section 2 but with $N > 2$ types, where N is finite and $1 \geq t^1 > t^2 > \cdots > t^N > 0$.[34] Let type j's probability be $\gamma^j > 0$.

Most of the analysis of the screening problem generalizes easily. A DM is an array $(\mathbf{a}^j, \mathbf{p}^j)_{j=1}^N$ with $\mathbf{a}^j : [\underline{v}, \overline{v}] \to [\underline{a}, \overline{a}]$ and $\mathbf{p}^j : [\underline{v}, \overline{v}] \to \mathbb{R}$, where $[\underline{v}, \overline{v}] = [\underline{s}t^N, \overline{s}t^1]$. As in problem $\mathcal{P}$, for every j and n a DM must satisfy (IC_2^j), (IR^j), and constraint IC_1^{jn} given by $U^j(\mathbf{a}^j, \mathbf{p}^j) \geq U^j(\mathbf{a}^n, \mathbf{p}^n)$. The provider's objective is unchanged, except that expectations are over N types. Finally, a generalization of Proposition 4.1 says that if (IC_2^j) holds, then $U^i(\mathbf{a}^j, \mathbf{p}^j) \geq U^n(\mathbf{a}^j, \mathbf{p}^j)$ for $i < n$, with equality if and only if $\mathbf{a}^j$ is constant over $(\underline{v}^n, \overline{v}^i)$ (where $\underline{v}^n = \underline{s}t^n$ and $\overline{v}^i = \overline{s}t^i$).[35]

With more than two degrees of inconsistency, however, screening raises one additional complication. Not only can the constraints IC_1^{jn} bind in both directions between adjacent types, as in the two-type model; they can also be violated for nonadjacent types, while holding between all adjacent ones. A related issue appears in dynamic-mechanism-design models with only consistent agents. This paper, however, handles it in a different way from that commonly used in the literature (see Section 5.2). It uses Lagrangian methods, which rely on having finitely many types.

Applying these methods reveals that, when designing each j-device, the provider again trades off the surplus with j against the rents that the device causes for (some of) the less inconsistent types. Moreover, she has to ensure that types more inconsistent than j do not mimic j (see Appendix B of the Supplemental Material). Based on Section 4.3, Proposition 4.5 considers the case in which unused options suffice to ensure this property.

[34]It follows from previous considerations that the case with $1 \leq t^1 < t^2 < \cdots < t^N < +\infty$ is analogous.

[35]This generalization follows easily from the proof of Proposition 4.1.

PROPOSITION 4.5: *Suppose $b(\underline{a})$ is low, so that unused options suffice to satisfy IC_1^{jn} for $j > n$. An optimal mechanism exists, with $\mathbf{a}^j$ unique over $(\underline{v}^j, \overline{v}^j)$ for every j. Moreover*, (i) $\mathbf{a}^1$ *sustains* $\mathbf{e}$; (ii) *each* $\mathbf{a}^j$ *sustains* $\mathbf{e}$ *if* $\sigma = 0$ (*otherwise at least* $\mathbf{a}^N$ *sustains a distorted outcome*); (iii) *if* $\mathbf{a}^j$ *is distorted, it curtails j's flexibility as in Theorem* 4.1; *and* (iv) *for* $j < N$, $\mathbf{a}^j$ *may have to feature unused options with* $a < \mathbf{a}^j(\underline{v}^j)$.

As long as the provider cares about profits ($\sigma > 0$), she will curtail flexibility for the most inconsistent type, as she did for type I. This is also true for an intermediate type n if, at the optimum, IC_1^{jn} binds for some less inconsistent j.

5. RELATION TO PREVIOUS LITERATURE

5.1. *Amador, Werning, and Angeletos (2006) (AWA)*

AWA studied the trade-offs between commitment and flexibility, using a consumption-savings problem similar to that in Section 1.1. At time 2, a present-biased self-2 chooses consumption y and savings a from the budget set $B = \{(y, a) \in \mathbb{R}^2_+ : y + a \leq m\}$. Anticipating this at time 1, self-1 wants to limit the deviations from his ideal state-contingent plan of self-2's actual choices while letting him act on time-2 information. AWA's model differs from the present paper in two key ways: (1) A commitment device is defined as any subset $D \subset B$ of options which self-2 is allowed to choose; (2) self-1 himself designs D (or, equivalently, a benevolent planner does that after observing the agent's type t). For $t < 1$, AWA showed that, under natural assumptions, the optimal D is a minimum-savings policy $D^* = \{(y, a) \in B : a \geq a_{\text{ms}}\}$ where a_{ms} exceeds the lowest savings in *self-1*'s ideal plan. (In terms of Section 1.1's example, $a_{\text{ms}} > 100$.)

Difference (1) concerns the commitment technology. To see how, take any $D \subset B$. Having a strictly increasing preference, self-2 considers only pairs $(y(a), a)$ in D such that y is maximal given a. Such a subset of D can be implemented by charging self-2 a *positive* payment for each a, interpreted as money burning. Indeed, for every $a \in [0, m]$, let $p(a) = m - a - y(a)$ if $(y(a), a) \in D$, and let $p(a)$ be large enough otherwise. For example, D^* can be implemented with $p^*(a) = 0$ if $a \geq a_{\text{ms}}$, and $p^*(a) = a_{\text{ms}} - a$ otherwise. Note that, with the suggested $p(\cdot)$, self-1's and self-2's relevant payoffs become $m - a - p(a) + sb(a)$ and $m - a - p(a) + tsb(a)$, which represent the same preferences as in Section 2. Thus AWA's commitment technology differs from that in the present model, because here payments can be strictly negative.[36]

As a consequence, AWA's results differ as follows. First, AWA's devices create a trade-off between commitment and flexibility, even though t is observable. As a result, D^* induces savings that are always inefficient: except when

[36]Examples of devices which we can view as featuring negative payments include existing special savings accounts with tax discounts that vary with how much one saves.

self-1 himself would choose a_{ms} (a zero-probability event), self-2 either undersaves or oversaves. Note that D^* curtails flexibility at the bottom of the efficient savings range, but for completely different reasons than those studied in Section 4. Second, a difference also arises when t is unobservable. AWA did not consider this case (difference (2)). One can see, however, that their devices cause no adverse-selection problem. An inconsistent agent's self-1 prefers D^* to all $D \subset B$. However, a consistent agent strictly prefers B to D^*, because D^* induces self-2 to oversave with strictly positive probability.

Thus AWA's implications for commitment policies involving savings also differ. First, policies such as Social Security and defined-benefit plans—which are better captured by their model—cause no adverse-selection problem. By contrast, devices such as defined-contribution plans—whose tax penalties and rewards are better captured by the present model—are subject to the adverse-selection problem identified here. Second, AWA gave a reason for penalizing inconsistent agents when they save less than some minimum amount, despite the resulting inefficiencies.[37] Here, the first-best I-device not only penalizes low levels of savings, but also rewards high levels, thereby avoiding inefficiencies. In both papers, however, penalties and rewards are justified by inconsistent agents' propensity to undersave. In contrast to AWA, this paper also gives a completely different reason for inefficiently restricting inconsistent agents' ability to save not only small but also large amounts: screening consistent from inconsistent agents. For examples of savings devices that resemble the differences between the screening C- and I-devices, one can compare the regular taxable accounts and the special "tax-shielded" accounts (such as IRAs and 401(k) plans) offered on the U.S. retirement market.[38]

Finally, AWA showed that, in a two-state setting, there is no commitment–flexibility trade-off if time inconsistency is weak. In this case, D^* contains only the two pairs (y, a) that self-1 finds optimal state by state. Intuitively, if self-2's bias is weak, it will not overcome the difference of appeal between states and induce him to always choose the option with lower savings. For a similar reason, in this paper, with finitely many states and t^I close to t^C, there is a *single* payment rule $\mathbf{p}$ that sustains $\mathbf{e}$ with both types. As a result, the provider will do so without having to worry about the agent's private information (see Appendix E of the Supplemental Material).

[37] See also Ambrus and Egorov (2013) and Bond and Sigurdsson (2013).

[38] For details on taxable and "tax-shielded" accounts related to this paper, see Amromin (2002, 2003), Holden, Sabelhaus, and Bass (2010), Munnell and Sundén (2006), and Holden and Schrass (2008, 2009, 2010). Other examples of savings devices that resemble the screening I-device include some Christmas-club and "individual-development" accounts (Ashraf et al. (2003)). In the exercising setting, there are devices that offer monetary incentives to work out regularly but also have restrictions similar to those of the I-device (see, e.g., GymPact.com).

5.2. *Sequential Screening and Time (In)consistency*

This paper points out some differences between sequential-screening models with and without inconsistent agents. For illustration, call the model in Section 2 TIM for "time inconsistency model." Consider also another model, called TCM for "time-consistency model," which is identical to TIM except that the agent's utility function is $u_2(a; s, t) - p$ in *both* periods.[39]

First, in TIM, unused options at time 2 can relax IC constraints at time 1 and can be essential in optimal mechanisms. This result is related to a key insight of Gul and Pesendorfer (GP) (2001): agents who are prone to temptations (or are time inconsistent) dislike menus with more options, as such menus exacerbate the woes of temptation. GP did not study mechanism-design problems, but it is easy to see how one could use their insight to relax IC constraints in settings with inconsistent agents. In TCM, by contrast, unused options can only make IC constraints harder to satisfy. Any option in the t-device not used by type t can only increase the payoff that another type t' expects at time 1 from choosing the t-device.

GP's insight is also behind results on mechanisms with unused options in Esteban and Miyagawa (EM) (2006). EM studied the classic monopolistic-screening problem in which, however, buyers have GP's temptation preferences with self-control. In EM, therefore, the monopolist screens buyers' usual valuation but not their degree of inconsistency. Also, buyers do not value flexibility, because they do not receive information over time—so EM did not study sequential screening. As a result, EM's mechanism relies on unused options but in a different way. First, they are always necessary for screening high-valuation buyers so as to maximize profits. Here, by contrast, they may not be necessary to keep I away from the efficient C-device: since C demands flexibility, such a device offers many options and can already create enough temptations from I's viewpoint. Second, one can show that, in EM, unused options never help to maximize welfare, whereas here they may be necessary. And last but not least, the arrival of future information creates a new trade-off, when designing unused options, between the probability of I's choosing them and how tempting they can be made to be (Proposition 4.3).

The second difference between TCM and TIM is closely related to the first and concerns how rich a DM must be to avoid any loss of generality. Based on the previous observations and Myerson (1986), in TCM we can use truthful DMs which, at each time, ask the agent only his "incremental" information: t at time 1, and only s at time 2. Such DMs can describe a device for each t and, for each t-device, the option that type t picks in each s. Figure 3(a) illustrates such a DM in a two-type example with two states. At time 1, the agent reports either t^C or t^I (i.e., he commits to one device); given this, at time 2, he reports either s_1 or s_2; each sequence of reports leads to an outcome $(\mathbf{a}_k^j, \mathbf{p}_k^j)$.

[39]A similar model appears, for example, in Courty and Li (2000).

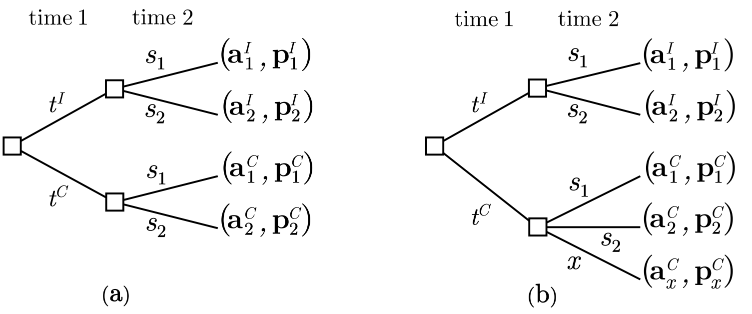

FIGURE 3.—Direct Mechanisms and Unused Options.

By asking only s at time 2, however, truthful DMs cannot describe devices with unused options. In TIM, such devices can be essential. To describe them through incentive-compatible DMs, we then have to consider DMs that let the agent reveal again, at time 2, information he received at time 1. To see this in the example, suppose that the optimal C-device has to include one option, besides $(\mathbf{a}^C_1, \mathbf{p}^C_1)$ and $(\mathbf{a}^C_2, \mathbf{p}^C_2)$, that only a dishonest I would choose in state s_2. To describe this device, a mechanism must involve at least three messages—say, s_1, s_2, and x—at the decision node following report t^C, as shown in Figure 3(b). Note that, by definition, only type I can send x after misreporting t^C at time 1 and observing s_2. Therefore, in the jargon of DMs, x is equivalent to the truthful, joint report (t^I, s_2) *at time 2*.

Finally, this paper departs from the literature in how it handles the IC constraints involving the agent's degree of inconsistency. As noted, these constraints can bind between adjacent types in both directions (Section 4.3) and between nonadjacent types (Section 4.5). Similar issues may also arise in TCM. This is because, in both models, at time 1 the agent chooses among multidimensional objects and has multidimensional information (recall that t affects beliefs about future payoffs). The literature has focused mostly on identifying restrictions on primitives such that local constraints bind only in one direction and imply global constraints.[40] In TIM, unfortunately, finding effective (let alone reasonable) restrictions is hard. Thus this paper has had to deal with all the IC constraints. Nonetheless, it characterizes the key properties of the optimal mechanism for screening time inconsistency, even when it is a priori impossible to tell which constraints will bind at the optimum (Section 4.5).

[40]For papers on this "local" approach to incentive compatibility, see Courty and Li (2000), Pavan, Segal, and Toikka (2014), and references therein. For a paper that discusses the importance of "global" approaches, see Battaglini and Lamba (2013).

6. DISCUSSION

The model in Section 2 can be extended in many directions. This section mentions only a few, leaving their proper treatment for future research.

First, we may consider other timings of when the agent learns t and s and interacts with the provider. Suppose that the provider offers her devices at time 1 but the agent learns t only at time 2 together with s. Then, despite symmetric information at time 1, no mechanism can achieve efficiency with both C and I. In short, at time 2 the agent's preference is fully determined by st. Thus no one-dimensional payment rule can elicit s separately from t, so as to achieve efficiency.[41] Alternatively, suppose that the agent learns t at an intermediate stage, before his preference changes from u_1 to u_2, and that unused options suffice to deter I from mimicking C. The provider will then achieve efficiency with both types: she will offer the agent the possibility of choosing, at the intermediate stage, between efficient C- and I-devices, and she will extract all rents at time 1.

Second, we may consider a model in which some agents are naive: at time 1, they unconsciously mispredict their self-control and cling to this prediction until they learn their time-2 preference.[42] Naïveté should not eliminate the adverse-selection problem identified here. At time 1, only self-1's belief about t matters—not whether it is correct. So if Alex *believes* himself to be less inconsistent than Bob, Alex should value any flexible device more than Bob if he *believes* himself to be more inconsistent. Therefore, if Bob gets a flexible device, Alex must get information rents. And curtailment of flexibility in the device for Bob should remain a key aspect of the optimal mechanism. With regard to unused options, they cannot help in screening a fully naive inconsistent agent, who believes himself to be consistent, from a genuinely consistent agent. With partial naïveté, designing such options requires caution, for a naive agent can mispredict how likely he is to pick them at time 2.

Finally, the paper assumes that the provider can offer commitment devices to begin with. However, this ability may be limited if she herself lacks commitment power or if agents are free to contract with other parties in the future. Gottlieb (2008) and Zhang (2012) tackled this issue, but only in settings in which agents' degree of inconsistency is observable.

7. CONCLUSION

In the absence of (ex ante) information asymmetry, devices that use unconstrained monetary incentives can fully solve agents' time inconsistency and avoid trade-offs between commitment and flexibility. Since such devices are

[41] This is related to Jehiel and Moldovanu (2001), which shows that efficiency cannot be achieved in contexts with interdependent valuations and multidimensional information.

[42] For models with naive, time inconsistent agents, see Kőszegi (2013).

valuable for agents who anticipate their inconsistency, both firms and benevolent governments have an incentive to supply them. However, the combination of agents' demand for flexibility and private information on their degree of inconsistency creates an adverse-selection problem. Its profit-maximizing solution involves devices for more inconsistent agents that curtail flexibility at both ends of the efficient choice range, and devices for less inconsistent agents that may include unused options and even distort behavior. These properties can also mark the welfare-maximizing solution.

These results can offer insights on how firms and governments (should) supply commitment devices when they cannot observe each agent's degree of inconsistency.

APPENDIX A: PROOFS OF THE MAIN RESULTS

All omitted proofs are in Appendix B of the Supplemental Material.

A.1. *Proof of Proposition 4.1*

Suppose $t^I < t^C \leq 1$; the proof when $t^I > t^C \geq 1$ is similar. By standard arguments, (IC_2^j) holds if and only if $\mathbf{a}^j$ is increasing and, for $v \in [\underline{v}, \overline{v}]$, $\mathbf{p}^j(v)$ satisfies a condition analogous to (E) in the main text. Given type i, let F^i be the distribution that F induces on $[\underline{v}, \overline{v}]$. We have

$$\begin{aligned}
U^i(\mathbf{a}^j, \mathbf{p}^j) &= \int_{\underline{v}^i}^{\overline{v}^i} \bigg[u_1(\mathbf{a}^j(v); v/t^i) - u_2(\mathbf{a}^j(v); v) \\
&\qquad - \int_v^{\overline{v}} b(\mathbf{a}^j(x))\,dx \bigg] dF^i(v) - k^j \qquad \text{(A.1)} \\
&= \int_{\underline{s}}^{\overline{s}} \bigg[sb(\mathbf{a}^j(t^i s)) - t^i sb(\mathbf{a}^j(t^i s)) \\
&\qquad - \int_{t^i s}^{\overline{v}} b(\mathbf{a}^j(x))\,dx \bigg] dF(s) - k^j. \qquad \text{(A.2)}
\end{aligned}$$

Using (A.2), we have

$$U^C(\mathbf{a}^j, \mathbf{p}^j) - U^I(\mathbf{a}^j, \mathbf{p}^j) = \int_{\underline{s}}^{\overline{s}} \Delta(s|\mathbf{a}^j)\,dF,$$

where

$$\begin{aligned}
\Delta(s|\mathbf{a}^j) &= s(1 - t^C)\big[b(\mathbf{a}^j(t^C s)) - b(\mathbf{a}^j(t^I s))\big] \qquad \text{(A.3)} \\
&\quad + \int_{t^I s}^{t^C s} \big[b(\mathbf{a}^j(x)) - b(\mathbf{a}^j(t^I s))\big]\,dx.
\end{aligned}$$

Since $\mathbf{a}^j$ is increasing and $t^I < t^C \leq 1$, $\Delta(s|\mathbf{a}^j) \geq 0$ on $(\underline{s}, \overline{s})$, with equality if $\mathbf{a}^j(v) = a$ on $(\underline{v}, \overline{v})$. So, $U^C(\mathbf{a}^j, \mathbf{p}^j) \geq U^I(\mathbf{a}^j, \mathbf{p}^j)$, with equality if $\mathbf{a}^j$ is constant on $(\underline{v}, \overline{v})$. If not, there is $\tilde{v} \in (\underline{v}, \overline{v})$ such that $v < \tilde{v} < v'$ implies $\mathbf{a}^j(v) < \mathbf{a}^j(v')$. Let $\tilde{s}_1 = \tilde{v}/t^C$ and $\tilde{s}_2 = \tilde{v}/t^I$, and consider $\mathcal{I} = (\tilde{s}_1, \tilde{s}_2) \cap [\underline{s}, \overline{s}] \neq \emptyset$. For $s \in \mathcal{I}$, $t^I s < \tilde{v} < t^C s$ and $\mathbf{a}^j(t^I s) < \mathbf{a}^j(t^C s)$, so

$$\int_{t^I s}^{t^C s} \left[b\big(\mathbf{a}^j(x)\big) - b\big(\mathbf{a}^j(t^I s)\big)\right] dx \geq \int_{\tilde{v}}^{t^C s} \left[b\big(\mathbf{a}^j(x)\big) - b\big(\mathbf{a}^j(t^I s)\big)\right] dx > 0,$$

where the first inequality follows from $\mathbf{a}^j$ being increasing. Since $dF > 0$,

$$\int_{\mathcal{I}} \left[\int_{t^I s}^{t^C s} \left[b\big(\mathbf{a}^j(x)\big) - b\big(\mathbf{a}^j(t^I s)\big)\right] dx\right] dF > 0,$$

which implies that $U^C(\mathbf{a}^j, \mathbf{p}^j) > U^I(\mathbf{a}^j, \mathbf{p}^j)$.

A.2. *Lemma A.1*

LEMMA A.1: $(\mathbf{a}^C, \mathbf{a}^I)$ *solves* $\mathcal{P}'$ *if and only if, for some* $\mu \geq 0$,

$$\mathbf{a}^j \in \arg \max_{\hat{\mathbf{a}}^j \text{ increasing}} \left\{W^j\big(\hat{\mathbf{a}}^j\big) - r^{-j} R^{-j}\big(\hat{\mathbf{a}}^j\big)\right\} \quad \textit{for} \quad j = C, I,$$

$$R^C\big(\mathbf{a}^I\big) + R^I\big(\mathbf{a}^C\big) \leq 0, \quad \textit{and} \quad \mu\big[R^C\big(\mathbf{a}^I\big) + R^I\big(\mathbf{a}^C\big)\big] = 0,$$

where $r^C = \sigma \frac{\gamma}{1-\gamma} + \frac{\mu}{1-\gamma}$ *and* $r^I = \frac{\mu}{\gamma}$.

PROOF: Let $\mathcal{B} = \{\mathbf{b} : [\underline{v}, \overline{v}] \to [b(\underline{a}), b(\overline{a})] \mid \mathbf{b} \text{ increasing}\}$. If $\mathbf{a}$ is increasing, then $\mathbf{b}(v) = b(\mathbf{a}(v)) \in \mathcal{B}$; if $\mathbf{b} \in \mathcal{B}$, then $\mathbf{a}(v) = b^{-1}(\mathbf{b}(v))$ is increasing. Let $\widetilde{W}^i(\mathbf{b}) = W^i(b^{-1}(\mathbf{b}))$ and $\widetilde{R}^i(\mathbf{b}^j) = R^i(b^{-1}(\mathbf{b}^j))$. Then $\mathcal{P}'$ is equivalent to

$$\mathcal{P}^{\mathbf{b}} = \begin{cases} \max \gamma \widetilde{W}^C\big(\mathbf{b}^C\big) + (1-\gamma)\left[\widetilde{W}^I\big(\mathbf{b}^I\big) - \dfrac{\sigma\gamma}{1-\gamma} \widetilde{R}^C\big(\mathbf{b}^I\big)\right] \\ \text{s.t.} \quad \mathbf{b}^C, \mathbf{b}^I \in \mathcal{B} \quad \text{and} \quad \widetilde{R}^C\big(\mathbf{b}^I\big) + \widetilde{R}^I\big(\mathbf{b}^C\big) \leq 0. \end{cases}$$

The space $\mathcal{X} = \{(\mathbf{b}^C, \mathbf{b}^I) \mid \mathbf{b}^i : [\underline{v}, \overline{v}] \to \mathbb{R}\}$ is linear and $\mathcal{Y} = \mathcal{B} \times \mathcal{B} \subset \mathcal{X}$ is convex. The objective is concave, as b^{-1} and c are convex; $\widetilde{R}^I(\cdot) + \widetilde{R}^C(\cdot)$ is linear. If $\mathbf{b}^C = b(\mathbf{a}^C_{fb})$ and $\mathbf{b}^I$ is constant, then $(\mathbf{b}^C, \mathbf{b}^I) \in \mathcal{Y}$ and $\widetilde{R}^I(\mathbf{b}^C) + \widetilde{R}^C(\mathbf{b}^I) < 0$. Given $\mu \geq 0$, define

$$\begin{aligned} L\big(\mathbf{b}^C, \mathbf{b}^I, \mu\big) &= \gamma \widetilde{W}^C\big(\mathbf{b}^C\big) + (1-\gamma)\left[\widetilde{W}^I\big(\mathbf{b}^I\big) - \frac{\sigma\gamma}{1-\gamma} \widetilde{R}^C\big(\mathbf{b}^I\big)\right] \\ &\quad - \mu\big[\widetilde{R}^I\big(\mathbf{b}^C\big) + \widetilde{R}^C\big(\mathbf{b}^I\big)\big] \\ &= \gamma\big[\widetilde{W}^C\big(\mathbf{b}^C\big) - r^I \widetilde{R}^I\big(\mathbf{b}^C\big)\big] + (1-\gamma)\big[\widetilde{W}^I\big(\mathbf{b}^I\big) - r^C \widetilde{R}^C\big(\mathbf{b}^I\big)\big]. \end{aligned} \tag{A.4}$$

By Corollary 1, p. 219, and Theorem 2, p. 221, of Luenberger (1969), $(\mathbf{b}_0^C, \mathbf{b}_0^I)$ solves $\mathcal{P}^{\mathbf{b}}$ if and only if there is $\mu_0 \geq 0$ such that (1) $(\mathbf{b}_0^C, \mathbf{b}_0^I)$ maximizes $L(\mathbf{b}^C, \mathbf{b}^I, \mu_0)$ over $\mathcal{Y}$ and (2) μ_0 minimizes $L(\mathbf{b}_0^C, \mathbf{b}_0^I, \mu)$ for $\mu \geq 0$. Condition (1) holds if and only if $\mathbf{b}_0^C$ and $\mathbf{b}_0^I$ maximize the first and second term in brackets of (A.4). Condition (2) holds if $\widetilde{R}^C(\mathbf{b}_0^I) + \widetilde{R}^I(\mathbf{b}_0^C) \leq 0$ and $\mu_0[\widetilde{R}^C(\mathbf{b}_0^I) + \widetilde{R}^I(\mathbf{b}_0^C)] = 0$. Finally, if $(\mathbf{b}_0^C, \mathbf{b}_0^I)$ solves $\mathcal{P}^{\mathbf{b}}$ and $\widetilde{R}^C(\mathbf{b}_0^I) + \widetilde{R}^I(\mathbf{b}_0^C) < 0$, then clearly $\mu_0 = 0$. Translating these conditions into those in Lemma A.1 is immediate. *Q.E.D.*

A.3. *Proof of Theorem 4.1 and Proposition 4.2*

Note that

$$W^i(\mathbf{a}^i) = \int_{\underline{v}}^{\overline{v}} \big[u_1\big(\mathbf{a}^i(v); v/t^i\big) - c\big(\mathbf{a}^i(v)\big)\big]\, dF^i. \tag{A.5}$$

Use (A.1) to express $R^C(\mathbf{a}^I)$. Changing order of integration and rearranging yields

$$R^C(\mathbf{a}^I) = -\int_{\overline{v}^I}^{\overline{v}^C} b\big(\mathbf{a}^I(v)\big) g^C(v)\, dv - \int_{\underline{v}^I}^{\overline{v}^I} b\big(\mathbf{a}^I(v)\big) G^I(v)\, dF^I, \tag{A.6}$$

where $g^C(v) : (\overline{v}^I, \overline{v}^C] \to \mathbb{R}$ and $G^I : [\underline{v}^I, \overline{v}^I] \to \mathbb{R}$ are given by

$$g^C(v) = \frac{t^C - 1}{t^C} v f^C(v) - \big(1 - F^C(v)\big) \quad \text{and}$$

$$G^I(v) = q^I(v) - \frac{f^C(v)}{f^I(v)} q^C(v),$$

with $q^i(v) = v/t^i - v - F^i(v)/f^i(v)$. By (A.5) and (A.6), $W^I(\mathbf{a}^I) - r^C R^C(\mathbf{a}^I)$ equals

$$\begin{aligned} VS^I\big(\mathbf{a}^I; r^C\big) = {} & \int_{\underline{v}^I}^{\overline{v}^I} \big[b\big(\mathbf{a}^I(v)\big) w^I\big(v; r^C\big) - \mathbf{a}^I(v) - c\big(\mathbf{a}^I(v)\big)\big]\, dF^I \\ & - r^C \int_{\overline{v}^I}^{\overline{v}^C} b\big(\mathbf{a}^I(v)\big)\big(1 - F^C(v)\big)\, dv, \end{aligned} \tag{A.7}$$

where $w^I(v; r^C) = v/t^I + r^C G^I(v)$. As in Lemma A.1's proof, it is convenient to work with $\mathbf{b}^I \in \mathcal{B}$, letting $\mathbf{a}^I = b^{-1}(\mathbf{b}^I)$.

Part 1: Existence and Uniqueness

Since $f > 0$, $(F^I)^{-1} : [0, 1] \to [\underline{v}^I, \overline{v}^I]$ is well-defined, strictly increasing, and continuous. Fix r^C and, for $x \in [0, 1]$, define $z(x; r^C) = w^I((F^I)^{-1}(x); r^C)$ and

$Z(x; r^C) = \int_0^x z(y; r^C)\, dy$. Then, z is continuous in x, except possibly at $x^m = F^I(v^m) > 0$, where $v^m = \min\{\overline{v}^I, \underline{v}^C\}$: if $t^C < 1$ and $\underline{v}^C < \overline{v}^I$,

$$\lim_{v \downarrow \underline{v}^C} w^I(v; r^C) = \lim_{v \uparrow \underline{v}^C} w^I(v; r^C) - r^C \frac{f^C(\underline{v}^C)}{f^I(\underline{v}^C)} \left[\frac{1 - t^C}{t^C} \underline{v}^C \right]. \tag{A.8}$$

Let $\Omega = \text{conv}(Z)$ (Rockafellar (1970, p. 36)). If $\Omega'(x; r^C)$ exists at $x \in [0, 1]$, let $\omega(x; r^C) = \Omega'(x; r^C)$. W.l.o.g., extend $\omega(x; r^C)$ by right-continuity on $[0, 1)$ and by left-continuity at 1.

LEMMA A.2: *ω is continuous in x and r^C.*

On $[\underline{v}^I, \overline{v}^I]$, let $\overline{w}^I(v; r^C) = \omega(F^I(v); r^C)$, which is increasing and continuous. Let $-\xi(\cdot) = b^{-1}(\cdot) + c(b^{-1}(\cdot))$ and replace w^I with $\overline{w}^I$ and $\mathbf{a} = b^{-1}(\mathbf{b})$ in VS^I to get

$$\overline{VS}^I(\mathbf{b}; r^C) = \int_{\underline{v}^I}^{\overline{v}^I} \left[\mathbf{b}(v) \overline{w}^I(v; r^C) + \xi(\mathbf{b}(v)) \right] dF^I + r^C \int_{\overline{v}^I}^{\overline{v}^C} \mathbf{b}(v) g^C(v)\, dv.$$

On $[\underline{v}^I, \overline{v}^I]$, let $\varphi(y, v; r^C) = y\overline{w}^I(v; r^C) + \xi(y)$ and

$$\overline{\mathbf{b}}^I(v; r^C) = \arg \max_{y \in [b(\underline{a}), b(\overline{a})]} \varphi(y, v; r^C). \tag{A.9}$$

On $(\overline{v}^I, \overline{v}^C]$, let $\overline{\mathbf{b}}^I(v; r^C) = b(\underline{a})$. Then, $\overline{\mathbf{b}}^I(r^C)$ is the unique pointwise maximizer of $\overline{VS}^I$, but it need not be increasing. Lemma A.3 characterizes any increasing maximizer of $\overline{VS}^I$.

LEMMA A.3: *If $\overline{VS}^I(\mathbf{b}^I; r^C) = \max_{\mathbf{b} \in \mathcal{B}} \overline{VS}^I(\mathbf{b}; r^C)$, $\mathbf{b}^I$ must satisfy $\mathbf{b}^I(v; r^C) = \overline{\mathbf{b}}^I(v; r^C)$ if $\underline{v}^I < v < v^{\mathrm{b}}$, and $\mathbf{b}^I(v; r^C) = y^{\mathrm{b}}(r^C)$ if $v^{\mathrm{b}} \leq v < \overline{v}^C$, where $v^{\mathrm{b}} \in [\underline{v}^I, \overline{v}^I]$ and $y^{\mathrm{b}}(r^C) \leq \overline{\mathbf{b}}^I(v^{\mathrm{b}}; r^C)$. If $v^{\mathrm{b}} > \underline{v}^I$, then $y^{\mathrm{b}}(r^C) = \overline{\mathbf{b}}^I(v^{\mathrm{b}}; r^C)$.*

PROOF: Drop r^C and suppose $\mathbf{b}^I \in \mathcal{B}$ maximizes $\overline{VS}^I$. First, $\mathbf{b}^I(v) = \mathbf{b}^I(\overline{v}^I)$ on $(\overline{v}^I, \overline{v}^C)$. If not, there is $v' \in (\overline{v}^I, \overline{v}^C)$ such that $\mathbf{b}^I(v) > \mathbf{b}^I(\overline{v}^I)$ for $v > v'$. Then $\mathbf{b}^I$ cannot be optimal, as $g^C(v) < 0$ on $(\overline{v}^I, \overline{v}^C)$. Consider $\mathbf{b}^I(v)$ on $[\underline{v}^I, \overline{v}^I]$. Recall that $\varphi(y, v)$ in (A.9) is strictly concave in y and continuous in v, and that $\overline{\mathbf{b}}^I(v)$ is continuous and increasing on $[\underline{v}^I, \overline{v}^I]$.

Case 1: $\mathbf{b}^I(\overline{v}^I) \leq \overline{\mathbf{b}}^I(\underline{v}^I)$. Then $\mathbf{b}^I(v) = \mathbf{b}^I(\overline{v}^I)$ on $(\underline{v}^I, \overline{v}^I]$. If not, there is $v' > \underline{v}^I$ such that $\mathbf{b}^I(v) < \mathbf{b}^I(\overline{v}^I) \leq \overline{\mathbf{b}}^I(v)$ for $v \leq v'$. By strict concavity, for

$v \in (\underline{v}^I, \overline{v}^I]$, $\varphi(\mathbf{b}^I(\overline{v}^I), v) \geq \varphi(\mathbf{b}^I(v), v)$, with strict inequality for $v \leq v'$. So $\int_{\underline{v}^I}^{\overline{v}^I} \varphi(\mathbf{b}^I(\overline{v}^I), v)\, dF^I > \int_{\underline{v}^I}^{\overline{v}^I} \varphi(\mathbf{b}^I(v), v)\, dF^I$, a contradiction.

Case 2: $\mathbf{b}^I(\overline{v}^I) = \overline{\mathbf{b}}^I(v^{\mathrm{b}}) > \overline{\mathbf{b}}^I(\underline{v}^I)$ for some $v^{\mathrm{b}} \in (\underline{v}^I, \overline{v}^I]$. So $\mathbf{b}^I(v) = \min\{\overline{\mathbf{b}}^I(v^{\mathrm{b}}), \overline{\mathbf{b}}^I(v)\}$ on $(\underline{v}^I, \overline{v}^I]$. Suppose not. First, consider $(v^{\mathrm{b}}, \overline{v}^I]$ and suppose $\mathbf{b}^I(v) < \overline{\mathbf{b}}^I(v^{\mathrm{b}})$ for some $v > v^{\mathrm{b}}$. By the argument in case 1, setting $\mathbf{b}^I(v) = \overline{\mathbf{b}}^I(v^{\mathrm{b}})$ on $(v^{\mathrm{b}}, \overline{v}^I]$ strictly improves on $\mathbf{b}^I$: the resulting function is in $\mathcal{B}$ and $\int_{v^{\mathrm{b}}}^{\overline{v}^I} \varphi(\overline{\mathbf{b}}^I(v^{\mathrm{b}}), v)\, dF^I > \int_{v^{\mathrm{b}}}^{\overline{v}^I} \varphi(\mathbf{b}^I(v), v)\, dF^I$. Second, consider $(\underline{v}^I, v^{\mathrm{b}}]$ and suppose $\mathbf{b}^I(v') \neq \overline{\mathbf{b}}^I(v')$ for some v'. If $\mathbf{b}^I(v') > \overline{\mathbf{b}}^I(v')$, then by continuity of $\overline{\mathbf{b}}^I$ and monotonicity of $\mathbf{b}^I$, there is $v'' > v'$ such that $\mathbf{b}^I(v) > \overline{\mathbf{b}}^I(v)$ on (v', v''). Similarly, if $\mathbf{b}^I(v') < \overline{\mathbf{b}}^I(v')$, then there is $v''' < v'$ such that $\mathbf{b}^I(v) < \overline{\mathbf{b}}^I(v)$ on (v''', v'). In either case, since $\overline{\mathbf{b}}^I$ is the unique maximizer of $\varphi(y, v)$, for $v \in (\underline{v}^I, v^{\mathrm{b}}]$, $\varphi(\overline{\mathbf{b}}^I(v), v) \geq \varphi(\mathbf{b}^I(v), v)$, with strict inequality for $v \in (v''', v')$ or $v \in (v', v'')$. So $\int_{\underline{v}^I}^{v^{\mathrm{b}}} \varphi(\overline{\mathbf{b}}^I(v), v)\, dF^I > \int_{\underline{v}^I}^{v^{\mathrm{b}}} \varphi(\mathbf{b}^I(v), v)\, dF^I$, a contradiction.

It remains to show that $\mathbf{b}^I(\overline{v}^I) \leq \overline{\mathbf{b}}^I(\overline{v}^I)$. Suppose not. By the argument in case 2, $\mathbf{b}^I(v) = \overline{\mathbf{b}}^I(v)$ on $(\underline{v}^I, \overline{v}^I)$. Then, $\mathbf{b}^I(\overline{v}^I) > \overline{\mathbf{b}}^I(\overline{v}^I)$ cannot be optimal: since $\mathbf{b}^I(v) = \mathbf{b}^I(\overline{v}^I)$ and $g^C(v) < 0$ on $(\overline{v}^I, \overline{v}^C)$, reducing $\mathbf{b}^I(\overline{v}^I)$ to $\overline{\mathbf{b}}^I(\overline{v}^I)$ satisfies monotonicity and strictly improves $\overline{VS}^I$. *Q.E.D.*

By Lemma A.3, $\mathbf{b}^I$ is continuous on $(\underline{v}^I, \overline{v}^C)$. At $\underline{v}^I$ and $\overline{v}^C$, $\mathbf{b}^I$ is not pinned down, but we can extend it by continuity w.l.o.g.

LEMMA A.4: *There is a unique* $\mathbf{b}^I \in \mathcal{B}$ *that maximizes* $\overline{VS}^I(\mathbf{b}; r^C)$.

PROOF: Drop r^C. By Lemma A.3, if a solution $\mathbf{b}^I$ exists, then either (1) $\mathbf{b}^I$ is constant at $y \leq \overline{\mathbf{b}}^I(\underline{v}^I)$ on $[\underline{v}^I, \overline{v}^C]$, or (2) $\mathbf{b}^I$ is constant at $\overline{\mathbf{b}}^I(v^{\mathrm{b}})$ on $[v^{\mathrm{b}}, \overline{v}^C]$, with $v^{\mathrm{b}} \leq \overline{v}^I$, and equals $\overline{\mathbf{b}}^I(v)$ for $v \leq v^{\mathrm{b}}$.

Case (1): We have $\overline{VS}^I(\mathbf{b}^I) = VS^I(\mathbf{b}^I) = \widetilde{W}^I(\mathbf{b}^I)$, since $\int_{\underline{v}^I}^{\overline{v}^I} [w^I(v) - \overline{w}^I(v)]\, dF^I = 0$ and by Proposition 4.1. Moreover,

$$\text{(A.10)} \qquad \widetilde{W}^I(\mathbf{b}^I) = y \int_{\underline{v}^I}^{\overline{v}^I} (v/t^I)\, dF^I + \xi(y).$$

So, there is a unique constant maximizer of $\overline{VS}^I(\mathbf{b}^I)$. Call it $\mathbf{b}_1^I$.

Case (2): Using $\varphi(y, v)$ in (A.9), $\overline{VS}^I$ equals

$$\text{(A.11)} \qquad Y(v^{\mathrm{b}}) = \int_{\underline{v}^I}^{v^{\mathrm{b}}} \varphi(\overline{\mathbf{b}}^I(v), v)\, dF^I + \int_{v^{\mathrm{b}}}^{\overline{v}^I} \varphi(\overline{\mathbf{b}}^I(v^{\mathrm{b}}), v)\, dF^I + \overline{\mathbf{b}}^I(v^{\mathrm{b}})K,$$

where $K = r^C \int_{\overline{v}^I}^{\overline{v}^C} g^C(v)\,dv$. By continuity of $\overline{\mathbf{b}}^I$, $Y(v^{\mathrm{b}})$ is continuous. So there is $v^{\mathrm{b}} \in [\underline{v}^I, \overline{v}^I]$ that fully identifies a maximizer for case (2). Since $\overline{\mathbf{b}}^I$ can be locally flat, v^{b} need not be unique. But any two optimal v_1^{b} and v_2^{b} must satisfy $\overline{\mathbf{b}}^I(v_1^{\mathrm{b}}) = \overline{\mathbf{b}}^I(v_2^{\mathrm{b}})$. To see this, suppose that $v_1^{\mathrm{b}} < v_2^{\mathrm{b}}$ both maximize $Y(v^{\mathrm{b}})$ and $\overline{\mathbf{b}}^I(v_1^{\mathrm{b}}) < \overline{\mathbf{b}}^I(v_2^{\mathrm{b}})$. W.l.o.g., let v_1^{b} be the largest v such that $\overline{\mathbf{b}}^I(v) = \overline{\mathbf{b}}^I(v_1^{\mathrm{b}})$. Let $\mathbf{b}_1$ and $\mathbf{b}_2$ be the functions identified by v_1^{b} and v_2^{b}, and for $\alpha \in (0,1)$, let $\widetilde{\mathbf{b}} = \alpha\mathbf{b}_1 + (1-\alpha)\mathbf{b}_2 \in \mathcal{B}$. On $(v_1^{\mathrm{b}}, \overline{v}^C]$, $\mathbf{b}_2(v) \neq \mathbf{b}_1(v)$; on $[\underline{v}^I, v_1^{\mathrm{b}}]$, $\mathbf{b}_2(v) = \mathbf{b}_1(v) = \overline{\mathbf{b}}^I(v)$. By strict concavity of $\varphi(y,v)$,

$$\int_{\underline{v}^I}^{v_1^{\mathrm{b}}} \varphi\big(\widetilde{\mathbf{b}}(v), v\big)\,dF^I + \int_{v_1^{\mathrm{b}}}^{\overline{v}^I} \varphi\big(\widetilde{\mathbf{b}}(v), v\big)\,dF^I + \widetilde{\mathbf{b}}(v)K$$
$$> \alpha Y\big(v_1^{\mathrm{b}}\big) + (1-\alpha)Y\big(v_2^{\mathrm{b}}\big).$$

Note that $\widetilde{\mathbf{b}}$ is constant on $[v_2^{\mathrm{b}}, \overline{v}^C]$ at some $\overline{\mathbf{b}}^I(\tilde{v}^{\mathrm{b}})$, with $\tilde{v}^{\mathrm{b}} \in (v_1^{\mathrm{b}}, v_2^{\mathrm{b}})$. So $\mathbf{b}^I(v)$ equals $\min\{\overline{\mathbf{b}}^I(\tilde{v}^{\mathrm{b}}), \overline{\mathbf{b}}^I(v)\}$, it satisfies case (2) and, by the argument in Lemma A.3 (Case 2),

$$Y\big(\tilde{v}^{\mathrm{b}}\big) \geq \int_{\underline{v}^I}^{v_1^{\mathrm{b}}} \varphi\big(\widetilde{\mathbf{b}}(v), v\big)\,dF^I + \int_{v_1^{\mathrm{b}}}^{\overline{v}^I} \varphi\big(\widetilde{\mathbf{b}}(v), v\big)\,dF^I + \widetilde{\mathbf{b}}(v)K > Y\big(v_1^{\mathrm{b}}\big).$$

The claim follows. So any maximizer of $Y(v^{\mathrm{b}})$ identifies a unique $\mathbf{b}^I$ for case (2). Call it $\mathbf{b}_2^I$.

By an argument similar to that for the uniqueness of $\mathbf{b}_2^I$, $\overline{VS}^I(\mathbf{b}_2^I) = \overline{VS}^I(\mathbf{b}_1^I)$ if and only if $\mathbf{b}_2^I = \mathbf{b}_1^I$ (on $(\underline{v}^I, \overline{v}^C)$). So the overall maximizer of $\overline{VS}^I$ is unique. *Q.E.D.*

Denote the unique maximizer of $\overline{VS}^I$ by $\mathbf{b}^I(r^C)$.

LEMMA A.5: $\mathbf{b}^I(r^C)$ *is the unique maximizer of* $VS^I(b^{-1}(\mathbf{b}))$.[43]

PROOF: Drop r^C. Since $\mathbf{b} \in \mathcal{B}$ and $\Omega \leq Z$ with equality at 0 and 1, integrating by parts, we get

$$\int_{\underline{v}^I}^{\overline{v}^I} \mathbf{b}(v)\big[w^I(v) - \overline{w}^I(v)\big]\,dF^I$$
$$= \mathbf{b}(v)\big[Z\big(F^I(v)\big) - \Omega\big(F^I(v)\big)\big]\Big|_{\underline{v}^I}^{\overline{v}^I}$$

[43]The argument modifies Toikka's (2011) proof of Theorem 3.7 and Corollary 3.9 to account for $(\overline{v}^I, \overline{v}^C]$.

$$-\int_{\underline{v}^I}^{\overline{v}^I} \left[Z\left(F^I(v)\right) - \Omega\left(F^I(v)\right)\right] d\mathbf{b}(v)$$

$$= \int_{\underline{v}^I}^{\overline{v}^I} \left[\Omega\left(F^I(v)\right) - Z\left(F^I(v)\right)\right] d\mathbf{b}(v) \leq 0.$$

Rewriting $VS^I(b^{-1}(\mathbf{b}))$, we get

$$\sup_{\mathbf{b}\in\mathcal{B}} VS^I\left(b^{-1}(\mathbf{b})\right) = \sup_{\mathbf{b}\in\mathcal{B}} \left\{\overline{VS}^I(\mathbf{b}) + \int_{\underline{v}^I}^{\overline{v}^I} \left[\Omega\left(F^I(v)\right) - Z\left(F^I(v)\right)\right] d\mathbf{b}(v)\right\}.$$

Since $\mathbf{b}^I \in \mathcal{B}$ and maximizes $\overline{VS}^I(\mathbf{b})$, we have to show that

$$\text{(A.12)} \quad \int_{\underline{v}^I}^{\overline{v}^I} \left[\Omega\left(F^I(v)\right) - Z\left(F^I(v)\right)\right] d\mathbf{b}^I(v) = 0.$$

If $\mathbf{b}^I$ is constant, then $d\mathbf{b}^I \equiv 0$. If not, consider $\overline{\mathbf{b}}^I$ on $[\underline{v}^I, \overline{v}^I]$ defined in (A.9) and v such that $\Omega(F^I(v)) < Z(F^I(v))$. For an open interval N around v, $\overline{w}^I(\cdot) = \omega(F^I(v))$ and $\overline{\mathbf{b}}^I$ is constant, so $d\overline{\mathbf{b}}^I(\cdot)$ assigns zero measure to N. For any such N, $d\mathbf{b}^I$ does the same. Consider $[v^{\mathrm{b}}, \overline{v}^C]$, on which $\mathbf{b}^I$ is constant. If $N \subset [v^{\mathrm{b}}, \overline{v}^C]$, the claim is immediate. The same holds if $N \cap [v^{\mathrm{b}}, \overline{v}^C] = \emptyset$, as then $\mathbf{b}^I(v) = \overline{\mathbf{b}}^I(v)$ on N. If $N \cap [v^{\mathrm{b}}, \overline{v}^C] \neq \emptyset$ and $N \cap [\underline{v}^I, v^{\mathrm{b}}) \neq \emptyset$ (so $v^{\mathrm{b}} > \underline{v}^I$), then $\mathbf{b}^I$ is constant on $[v^{\mathrm{b}}, \overline{v}^C] \cup N$.

By Lemma A.4, if $\widetilde{\mathbf{b}} \in \mathcal{B}$ differs from $\mathbf{b}^I$ on $(\underline{v}^I, \overline{v}^C)$, then $\overline{VS}^I(\widetilde{\mathbf{b}}) < \overline{VS}^I(\mathbf{b}^I)$. Uniqueness follows on $(\underline{v}^I, \overline{v}^C)$; extending it to $[\underline{v}^I, \overline{v}^C]$ is w.l.o.g. *Q.E.D.*

Finally, let $\mathbf{a}^I_{sb} = b^{-1}(\mathbf{b}^I(r^C))$.

Part 2: Continuity and Limit Behavior of $\mathbf{b}^I(r^C)$

Part 1 shows continuity in v. Consider continuity in r^C. By (A.9) and the Maximum Theorem, $\overline{\mathbf{b}}^I(v, \cdot)$ is continuous in r^C for $v \in [\underline{v}^I, \overline{v}^I]$. Now consider $Y(v^{\mathrm{b}}; r^C)$ in (A.11). By pointwise continuity of $\overline{w}^I(v; r^C)$ and $\overline{\mathbf{b}}^I(v; r^C)$, $Y(v^{\mathrm{b}}; r^C)$ is continuous in r^C and so $V^{\mathrm{b}}(r^C) = \arg\max_{v\in[\underline{v}^I, \overline{v}^I]} Y(v; r^C)$ is u.h.c. For $v, v' \in V^{\mathrm{b}}(r^C)$, $\overline{\mathbf{b}}^I(v; r^C) = \overline{\mathbf{b}}^I(v'; r^C)$. Take any sequence $\{r^C_n\}$ with $r^C_n \to r^C$. Then, $v^{\mathrm{b}}(r^C_n) \to v^{\mathrm{b}} \in V^{\mathrm{b}}(r^C)$. The candidate $\mathbf{b}^I_2(r^C)$ that maximizes $Y(v^{\mathrm{b}}; r^C)$ is such that $\mathbf{b}^I_2(v; r^C) = \min\{\overline{\mathbf{b}}^I(v^{\mathrm{b}}(r^C); r^C), \overline{\mathbf{b}}^I(v; r^C)\}$ on $[\underline{v}^I, \overline{v}^I]$, and $\mathbf{b}^I_2(v; r^C) = \overline{\mathbf{b}}^I(v^{\mathrm{b}}(r^C); r^C)$ for $v > \overline{v}^I$. So, by continuity of $\overline{\mathbf{b}}^I$, $\mathbf{b}^I_2(v; r^C_n) \to \mathbf{b}^I_2(v; r^C)$ on $[\underline{v}^I, \overline{v}^C]$. Finally, the constant $\mathbf{b}^I_1$ in the proof of Lemma A.4, as well as (A.10), is independent of r^C. It remains to show that $\mathbf{b}^I(r^C_n)$ converges pointwise to $\mathbf{b}^I(r^C)$. First, if $VS^I(b^{-1}(\mathbf{b}^I_1)) > VS^I(b^{-1}(\mathbf{b}^I_2(r^C))) = Y(v^{\mathrm{b}}(r^C); r^C)$, then by continuity of Y, there is N such that $n \geq N$ implies $VS^I(b^{-1}(\mathbf{b}_1^I)) > VS^I(b^{-1}(\mathbf{b}_2^I(r_n^C)))$. So, for $n \geq N$, $\mathbf{b}^I(v; r_n^C) = \mathbf{b}_1^I$ on $[\underline{v}^I, \overline{v}^C]$. Second, if $VS^I(b^{-1}(\mathbf{b}_1^I)) < VS^I(b^{-1}(\mathbf{b}_2^I(r^C)))$, then again for n large $\mathbf{b}^I(r_n^C) = \mathbf{b}_2^I(r_n^C)$, which converges pointwise to $\mathbf{b}^I(r^C)$. Finally, if $VS^I(b^{-1}(\mathbf{b}_1^I)) = VS^I(b^{-1}(\mathbf{b}_2^I(r^C)))$, then $\mathbf{b}_1^I \equiv \mathbf{b}_2^I(r^C)$. So $|\mathbf{b}_1^I - \mathbf{b}^I(v; r_n^C)| \leq \max\{0, |\mathbf{b}_1^I - \mathbf{b}_2^I(v; r_n^C)|\} \to 0$ as $n \to \infty$.

Note that $VS^I(b^{-1}(\mathbf{b}), 0) = \widetilde{W}^I(\mathbf{b})$, so $\mathbf{b}^I(v, 0) = \mathbf{b}_{fb}^I(v) = b(\mathbf{a}_{fb}^I(v))$ on $(\underline{v}^I, \overline{v}^I)$, which we can extend to $[\underline{v}^I, \overline{v}^C]$ by letting $\mathbf{b}^I(\underline{v}^I, 0) = \mathbf{b}_{fb}^I(\underline{v}^I)$ and $\mathbf{b}^I(v, 0) = \mathbf{b}_{fb}^I(\overline{v}^I)$ for $v \geq \overline{v}^I$. Therefore, $\mathbf{b}^I(r^C) \to \mathbf{b}_{fb}^I$ pointwise as $r^C \to 0$. To prove $\max_{[\underline{v}, \overline{v}]} |\mathbf{b}^I(v; r^C) - b(a^{\text{nf}})| \to 0$ as $r^C \to +\infty$, recall that $\mathbf{b}^I(r^C)$ maximizes $VS^I(b^{-1}(\mathbf{b}); r^C) = \widetilde{W}^I(\mathbf{b}) - r^C \widetilde{R}^C(\mathbf{b})$ and that, by Proposition 4.1, $\widetilde{R}^C(\mathbf{b}) > 0$ for any $\mathbf{b} \in \mathcal{B}$ not constant on $(\underline{v}^I, \overline{v}^C)$. Clearly, $\mathbf{b}^I(r^C)$ cannot converge to a constant function with value $y_0 \neq b(a^{\text{nf}})$. Now, suppose $\mathbf{b}^I(r^C) \to \mathbf{b}_\infty^I$ pointwise, where $\mathbf{b}_\infty^I$ is not constant on $(\underline{v}^I, \overline{v}^C)$. Then, there is $\hat{r}^C \geq 0$ such that, for $r^C > \hat{r}^C$, $b(a^{\text{nf}})$ strictly dominates $\mathbf{b}_\infty^I$: since $\widetilde{W}^I(\mathbf{b}_\infty^I)$ is bounded and $\widetilde{R}^C(\mathbf{b}_\infty^I) > 0$, $\widetilde{W}^I(\mathbf{b}_\infty^I) - \hat{r}^C \widetilde{R}^C(\mathbf{b}_\infty^I) \leq \widetilde{W}^I(b(a^{\text{nf}}))$. Finally, consider the extension of $\mathbf{b}^I(r^C)$ by continuity. By monotonicity,

$$\begin{aligned}
&\max_{[\underline{v}, \overline{v}]} \left|\mathbf{b}^I(v; r^C) - b(a^{\text{nf}})\right| \\
&\quad = \max\left\{\left|\mathbf{b}^I(\underline{v}; r^C) - b(a^{\text{nf}})\right|, \left|\mathbf{b}^I(\overline{v}; r^C) - b(a^{\text{nf}})\right|\right\}.
\end{aligned}$$

Part 3: Properties (a)–(c) of $\mathbf{b}^I(r^C)$

(b): Drop r^C. Recall that (1) $\mathbf{b}^I$ satisfies Lemma A.3, (2) on $[\underline{v}^I, \overline{v}^I]$, $\overline{\mathbf{b}}^I$ is defined by (A.9) and continuous. Suppose $v^{\mathrm{b}} > \underline{v}^I$. For $v < v^{\mathrm{b}}$, $\mathbf{b}^I(v) = \overline{\mathbf{b}}^I(v) < \overline{\mathbf{b}}^I(v^{\mathrm{b}}) = \mathbf{b}^I(v^{\mathrm{b}})$ by Lemma A.3 and $Y(v^{\mathrm{b}}) \geq Y(v)$ (see (A.11)). So

$$\begin{aligned}
\frac{Y(v^{\mathrm{b}}) - Y(v)}{\overline{\mathbf{b}}^I(v^{\mathrm{b}}) - \overline{\mathbf{b}}^I(v)} &= K + \int_v^{\overline{v}^I} \overline{w}^I(y)\, dF^I - \int_v^{v^{\mathrm{b}}} \overline{w}^I(y)\, dF^I \\
&\quad + \frac{\xi(\overline{\mathbf{b}}^I(v^{\mathrm{b}})) - \xi(\overline{\mathbf{b}}^I(v))}{\overline{\mathbf{b}}^I(v^{\mathrm{b}}) - \overline{\mathbf{b}}^I(v)}\left(1 - F^I(v)\right) \\
&\quad - \int_v^{v^{\mathrm{b}}} \frac{\xi(\overline{\mathbf{b}}^I(v^{\mathrm{b}})) - \xi(\overline{\mathbf{b}}^I(y))}{\overline{\mathbf{b}}^I(v^{\mathrm{b}}) - \overline{\mathbf{b}}^I(v)}\, dF^I \\
&\geq 0.
\end{aligned}$$

Since $\overline{\mathbf{b}}^I$ and $\overline{w}^I$ are increasing and continuous, we get

$$\lim_{v \uparrow v^{\mathrm{b}}} \frac{Y(v^{\mathrm{b}}) - Y(v)}{\overline{\mathbf{b}}^I(v^{\mathrm{b}}) - \overline{\mathbf{b}}^I(v)} = \int_{v^{\mathrm{b}}}^{\overline{v}^I} \left[\overline{w}^I(y) + \xi'\left(\overline{\mathbf{b}}^I(v^{\mathrm{b}})\right)\right] dF^I + K \geq 0. \tag{A.13}$$

Hence $v^{\mathrm{b}} < \overline{v}^I$, because $K < 0$ and $\xi'(\cdot) < 0$. I claim that $\overline{\mathbf{b}}^I(v) > \overline{\mathbf{b}}^I(v^{\mathrm{b}})$ for some $v \in (v^{\mathrm{b}}, \overline{v}^I]$. Suppose not. If $\overline{\mathbf{b}}^I(v^{\mathrm{b}})$ is interior, $\overline{w}^I(v) = -\xi'(\overline{\mathbf{b}}^I(v))$ for $v \geq v^{\mathrm{b}}$ and (A.13) is violated. If $\overline{\mathbf{b}}^I(v^{\mathrm{b}}) = b(\overline{a})$ and $V_{\overline{a}} = \{v \in [\underline{v}^I, \overline{v}^I] \mid \overline{w}^I(v) > -\xi'(b(\overline{a}))\}$ is nonempty (if it is, the previous case applies), then (A.13) is violated. Indeed, since v^{b} is the lowest v for which $\overline{w}^I(v) = -\xi'(b(\overline{a}))$, $\int_{v^{\mathrm{b}}}^{\overline{v}^I} [w^I(y) - \overline{w}^I(y)]\, dF^I = 0$; so

$$\begin{aligned}
&\int_{v^{\mathrm{b}}}^{\overline{v}^I} \big[\overline{w}^I(y) + \xi'\big(b(\overline{a})\big)\big]\, dF^I + K \\
&\quad = \int_{v^{\mathrm{b}}}^{\overline{v}^I} \big(y/t^I + \xi'\big(b(\overline{a})\big)\big)\, dF^I \\
&\qquad + r^C\bigg[\int_{\overline{v}^I}^{\overline{v}^C} g^C(y)\, dy + \int_{v^{\mathrm{b}}}^{\overline{v}^I} G^I(y)\, dF^I\bigg].
\end{aligned}$$

By Assumption 2.1 and concavity of ξ, $\overline{v}^I/t^I + \xi'(b(\overline{a})) < 0$. Moreover, integrating by parts,

$$\begin{aligned}
&\int_{\overline{v}^I}^{\overline{v}^C} g^C(v)\, dv = -\int_{\overline{v}^I}^{\overline{v}^C} \big(v/t^C\big)\, dF^C + \overline{v}^I\big(1 - F^C\big(\overline{v}^I\big)\big), \\
&\int_{v^{\mathrm{b}}}^{\overline{v}^I} q^i(v)\, dF^i = \int_{v^{\mathrm{b}}}^{\overline{v}^I} \big(v/t^i - v^{\mathrm{b}}\big)\, dF^i - \big(\overline{v}^I - v^{\mathrm{b}}\big)F^i\big(\overline{v}^I\big).
\end{aligned}$$

So, the term in brackets becomes

$$\begin{aligned}
\text{(A.14)}\quad &\int_{v^{\mathrm{b}}}^{\overline{v}^I} \big(v/t^I - v^{\mathrm{b}}\big)\, dF^I - \int_{v^{\mathrm{b}}}^{\overline{v}^C} \big(v/t^C - v^{\mathrm{b}}\big)\, dF^C(v) \\
&\quad = -\int_{v^{\mathrm{b}}/t^C}^{v^{\mathrm{b}}/t^I} \big(s - v^{\mathrm{b}}\big)\, dF(s) < 0,
\end{aligned}$$

where the inequality follows from $\overline{v}^I > v^{\mathrm{b}} > \underline{v}^I > 0$ and $0 < t^I < t^C \leq 1$.

Let $v^1 = \max\{v \mid \overline{\mathbf{b}}^I(v) = \overline{\mathbf{b}}^I(v^{\mathrm{b}})\} < \overline{v}^I$. For $v > v^1$, $\overline{\mathbf{b}}^I(v) > \overline{\mathbf{b}}^I(v^1)$ and $Y(v) \leq Y(v^1) = Y(v^{\mathrm{b}})$. By the argument giving (A.13),

$$\lim_{v \downarrow v^1} \frac{Y\big(v^1\big) - Y(v)}{\overline{\mathbf{b}}^I\big(v^1\big) - \overline{\mathbf{b}}^I(v)} = \int_{v^1}^{\overline{v}^I} \big[\overline{w}^I(y) + \xi'\big(\overline{\mathbf{b}}^I\big(v^1\big)\big)\big]\, dF^I + K \leq 0.$$

As $\overline{\mathbf{b}}^I(y)$ is interior and constant on $[v^{\mathrm{b}}, v^1]$, $\xi'(\overline{\mathbf{b}}^I(y)) = -\overline{w}^I(y) = -\overline{w}^I(v^{\mathrm{b}}) = \xi'(\overline{\mathbf{b}}^I(v^{\mathrm{b}}))$ and

$$-K \geq \int_{v^1}^{\overline{v}^I} \left[\overline{w}^I(y) - \overline{w}^I\left(v^1\right)\right] dF^I = \int_{v^{\mathrm{b}}}^{\overline{v}^I} \left[\overline{w}^I(y) - \overline{w}^I\left(v^{\mathrm{b}}\right)\right] dF^I \geq -K.$$

Finally, since $\int_{v^{\mathrm{b}}}^{\overline{v}^I} [w^I(y) - \overline{w}^I(y)]\, dF^I = 0$, the argument in Lemma A.2 yields $w^I(v^{\mathrm{b}}) = \overline{w}^I(v^{\mathrm{b}})$. Rearranging $w^I(y)$, we get

$$\text{(A.15)} \qquad \int_{v^{\mathrm{b}}}^{\overline{v}^I} \left[w^I\left(v^{\mathrm{b}}\right) - v/t^I\right] dF^I = r^C\left[\int_{v^{\mathrm{b}}}^{\overline{v}^I} G^I(v)\, dF^I + \int_{\overline{v}^I}^{\overline{v}^C} g^C(v)\, dv\right].$$

(a): Since $\mathbf{b}^I(r^C)$ and $\mathbf{b}^I_{fb} = b(\mathbf{a}^I_{fb})$ are continuous and increasing, it is enough to prove that $\mathbf{b}^I(\underline{v}^I; r^C) > \mathbf{b}^I_{fb}(\underline{v}^I)$ and $\mathbf{b}^I(\overline{v}^I; r^C) < \mathbf{b}^I_{fb}(\overline{v}^I)$. Hereafter, drop r^C. *Case 1*: $\mathbf{b}^I$ not constant. This implies $v^{\mathrm{b}} > \underline{v}^I$ and $\mathbf{b}^I(\overline{v}^I) = \mathbf{b}^I(v^{\mathrm{b}}) = \overline{\mathbf{b}}^I(v^{\mathrm{b}})$. By (A.14), (A.15), and $\overline{w}^I(v^{\mathrm{b}}) = w^I(v^{\mathrm{b}})$, we have $\overline{w}^I(v^{\mathrm{b}}) < \overline{v}^I/t^I$ and hence $\mathbf{b}^I(\overline{v}^I) < \mathbf{b}^I_{fb}(\overline{v}^I)$. Now let $v_{\mathrm{b}} = \max\{v \mid \overline{w}^I(v) = \overline{w}^I(\underline{v}^I)\}$.

LEMMA A.6: $\overline{w}^I(\underline{v}^I) \leq w^I(\underline{v}^I)$. *If the inequality is strict*, $v_{\mathrm{b}} > \underline{v}^I$.

So, if $\overline{w}^I(\underline{v}^I) = w^I(\underline{v}^I)$, it equals $(\underline{v}^I/t^I)(1 + r^C(1 - t^I)) + r^C[f^I(\underline{v}^I)]^{-1} > \underline{v}^I/t^I$. If $\overline{w}^I(\underline{v}^I) < w^I(\underline{v}^I)$, it equals $\overline{w}^I(v_{\mathrm{b}})$ on $[\underline{v}^I, v_{\mathrm{b}}]$. By definition, v_{b} must satisfy $\int_{\underline{v}^I}^{v_{\mathrm{b}}} [w^I(y) - \overline{w}^I(v_{\mathrm{b}})]\, dF^I = 0$, that is,

$$\text{(A.16)} \qquad \int_{\underline{v}^I}^{v_{\mathrm{b}}} \left(y/t^I - \overline{w}^I(v_{\mathrm{b}})\right) dF^I = -r^C \int_{\underline{v}^I}^{v_{\mathrm{b}}} G^I(y)\, dF^I.$$

Integrating by parts, $\int_{\underline{v}^I}^{v_{\mathrm{b}}} G^I(y)\, dF^I$ equals

$$\int_{\underline{v}^I}^{v_{\mathrm{b}}} \left(y/t^I - v_{\mathrm{b}}\right) dF^I - \int_{\underline{v}^I}^{v_{\mathrm{b}}} \left(y/t^C - v_{\mathrm{b}}\right) dF^C = \int_{v_{\mathrm{b}}/t^C}^{v_{\mathrm{b}}/t^I} (s - v_{\mathrm{b}})\, dF > 0,$$

where the inequality follows from $v_{\mathrm{b}} > \underline{v}^I > 0$ and $0 < t^I < t^C \leq 1$. So, by (A.16), $\overline{w}^I(v_{\mathrm{b}}; r^C) > \underline{v}^I/t^I$. In either case, $\mathbf{b}^I(\underline{v}^I) > \mathbf{b}^I_{fb}(\underline{v}^I)$.

Case 2: $\mathbf{b}^I$ constant. Then, $\mathbf{b}^I(v) = b(a^{\mathrm{nf}})$ on $[\underline{v}, \overline{v}]$. Since $\underline{v}^I/t^I < \mathbb{E}(s) < \overline{v}^I/t^I$, Assumption 2.1 implies $\mathbf{b}^I_{fb}(\underline{v}^I) < b(a^{\mathrm{nf}}) < \mathbf{b}^I_{fb}(\overline{v}^I)$.

(c): Drop r^C. Let $\underline{v}^I < v' \leq v^m$ (recall $v^m = \min\{\overline{v}^I, \underline{v}^C\}$) and consider

$$\text{(A.17)} \qquad w^I\left(v'\right) - w^I\left(\underline{v}^I\right) = \frac{v' - \underline{v}^I}{t^I}\left(1 + r^C\left(1 - t^I\right)\right) - r^C\left[\frac{F^I\left(v'\right)}{f^I\left(v'\right)} - \frac{F^I\left(\underline{v}^I\right)}{f^I\left(\underline{v}^I\right)}\right].$$

The first part of (A.17) is positive since $t^I \in (0,1)$, but the second can be negative. So $w(\cdot)$ can be decreasing in a neighborhood of $\underline{v}^I$. If it is, $\overline{w}^I(v)$ and $\mathbf{b}^I(r^C)$ are constant on $[\underline{v}^I, v_b] \neq \emptyset$. Since $\frac{F^I(v)}{f^I(v)} = \frac{t^I F(v/t^I)}{f(v/t^I)}$, the condition in Theorem 4.1(c) implies that, for $v' > v$ in $[\underline{v}^I, \min\{t^I s^\dagger, v^m\}]$, $w^I(v') \le w^I(v)$.

A.4. *Lemma A.7*

LEMMA A.7: *$R^C(\mathbf{a}^I_{fb}) > -R^I(\mathbf{a}^C_{fb})$ for a nonempty family $\mathcal{F}$ of distributions F.*

PROOF: Using (A.3),

$$-R^I(\mathbf{a}^C_{fb}) = \int_{\underline{s}}^{\underline{s}t^C/t^I} \Delta(s|\mathbf{a}^C_{fb})\,dF + \int_{\underline{s}t^C/t^I}^{\overline{s}} \Delta(s|\mathbf{a}^C_{fb})\,dF.$$

For $s \le \underline{s}\frac{t^C}{t^I}$, since $\mathbf{a}^C_{fb}(st^I) = \mathbf{a}^C_{fb}(\underline{s}t^C)$,

$$\Delta(s|\mathbf{a}^C_{fb}) = \underline{s}t^C b(\mathbf{a}^C_{fb}(\underline{s}t^C)) - st^C b(\mathbf{a}^C_{fb}(st^C)) + \int_{\underline{s}t^C}^{st^C} b(\mathbf{a}^C_{fb}(y))\,dy$$
$$+ s\big[b(\mathbf{a}^C_{fb}(st^C)) - b(\mathbf{a}^C_{fb}(\underline{s}t^C))\big].$$

Since $\mathbf{a}^C_{fb}$ is continuous, $\Delta(s|\mathbf{a}^C_{fb}) \to 0$ as $s \to \underline{s}$. Consider $R^C(\mathbf{a}^I_{fb})$. Since $st^I < st^C$, $\Delta(s|\mathbf{a}^I_{fb}) > 0$ for $s < \overline{s}$. Let $s_0 = \frac{1}{2}(\overline{s} + \underline{s})$. By continuity, $\min_{[\underline{s}, s_0]} \Delta(s|\mathbf{a}^I_{fb}) = \kappa > 0$. Choose $s_{\kappa/2} > \underline{s}$ so that $\Delta(s|\mathbf{a}^C_{fb}) \le \kappa/2$ for $s \in [\underline{s}, s_{\kappa/2}]$. Let $s_1 = \min\{s_0, s_{\kappa/2}\}$. Then,

$$R^C(\mathbf{a}^I_{fb}) \ge \int_{\underline{s}}^{s_1} \Delta(s|\mathbf{a}^I_{fb})\,dF \ge \kappa F(s_1),$$

$$-R^I(\mathbf{a}^C_{fb}) \le \sup_{s > s_1} \Delta(s|\mathbf{a}^C_{fb})\big(1 - F(s_1)\big) + \frac{\kappa}{2}F(s_1).$$

So $R^C(\mathbf{a}^I_{fb}) > -R^I(\mathbf{a}^C_{fb})$ if $F(s_1)/(1 - F(s_1)) > \frac{2}{\kappa}\sup_{s > s_1} \Delta(s|\mathbf{a}^C_{fb})$. *Q.E.D.*

A.5. *Proof of Proposition 4.3*

Using (A.1), changing order of integration, and rearranging yields

$$\text{(A.18)} \quad R^I(\mathbf{a}^C) = -\int_{\underline{v}^I}^{\underline{v}^C} b(\mathbf{a}^C(v)) g^I(v)\,dv + \int_{\underline{v}^C}^{\overline{v}^C} b(\mathbf{a}^C(v)) G^C(v)\,dF^C,$$

where $g^I : [\underline{v}^I, \underline{v}^C] \to \mathbb{R}$ and $G^C : [\underline{v}^C, \overline{v}^C] \to \mathbb{R}$ are given by

$$g^I(v) = \frac{t^I - 1}{t^I} v f^I(v) + F^I(v) \quad \text{and} \quad G^C(v) = \overline{q}^C(v) - \frac{f^I(v)}{f^C(v)}\overline{q}^I(v),$$

with $\overline{q}^j(v) = v - v/t^j - (1 - F^j(v))/f^j(v)$. Note that $g^I(\underline{v}^I) < 0$. Given $\mathbf{a}^C$ on $[\underline{v}^C, \overline{v}^C]$, we have to maximize $\int_{\overline{v}^C}^{\overline{v}^I} b(\mathbf{a}(v)) g^I(v)\, dv$ with $\mathbf{a} : [\underline{v}^I, \underline{v}^C) \to [\underline{a}, \mathbf{a}^C_{fb}(\underline{v}^C)]$ increasing. Let $\widetilde{F}$ be the uniform distribution on $[\underline{v}^I, \underline{v}^C]$ and $\widetilde{F}^{-1}$ its inverse. For $x \in [0, 1]$, let $z(x) = g^I(\widetilde{F}^{-1}(x))$ and $Z(x) = \int_0^x z(y)\, dy$. Let $\Omega = \mathrm{conv}(Z)$ and $\omega(x) = \Omega'(x)$, if $\Omega'(x)$ exists. W.l.o.g., extend $\omega(x)$ by right-continuity on $[0, 1)$ and by left-continuity at 1. On $[\underline{v}^I, \underline{v}^C]$, let $\overline{g}^I(v) = \omega(\widetilde{F}(v))$, which is increasing. Recall that $v^m = \min\{\overline{v}^I, \underline{v}^C\} > \underline{v}^I$, and let $x^m = \widetilde{F}(v^m) > 0$. Since g^I is continuous on $[\underline{v}^I, \overline{v}^I]$, by the logic behind Lemma A.2, ω is continuous on $[0, x^m]$. So $\overline{g}^I$ is continuous on $[\underline{v}^I, v^m]$. By the logic behind Lemma A.6, $\overline{g}^I(\underline{v}^I) \le g^I(\underline{v}^I)$. So $v_{\mathrm{u}} = \sup\{v \in [\underline{v}^I, \underline{v}^C] \mid \overline{g}^I(v) < 0\} > \underline{v}^I$. Similarly, define $v^{\mathrm{u}} = \inf\{v \in [\underline{v}^I, \underline{v}^C] \mid \overline{g}^I(v) > 0\}$; if the set is empty, set $v^{\mathrm{u}} = \underline{v}^C$. By Theorem 3.7 of Toikka (2011), $\mathbf{a}^C_{un}(v)$ must equal $\underline{a}$ for $v \in (\underline{v}^I, v_{\mathrm{u}})$ and $\mathbf{a}^C_{fb}(\underline{v}^C)$ for $v \in (v^{\mathrm{u}}, \underline{v}^C)$. Letting $\mathbf{a}^C_{un}(v^{\mathrm{u}}) = \mathbf{a}^C_{fb}(\underline{v}^C)$ is w.l.o.g. For completeness, on $[v_{\mathrm{u}}, v^{\mathrm{u}})$, $\mathbf{a}^C_{un}$ can be any increasing function, so long as it satisfies Toikka's *pooling property* (see Definition 3.5). By Corollary 3.8 of Toikka (2011), it is w.l.o.g. to set $\mathbf{a}^C_{un}(v) = \mathbf{a}^C_{fb}(\underline{v}^C)$ on $[v_{\mathrm{u}}, v^{\mathrm{u}})$.

A.6. *Proof of Theorem 4.2*

LEMMA A.8: *An increasing $\mathbf{a}^C(r^I)$ that maximizes $W^C(\hat{\mathbf{a}}^C) - r^I R^I(\hat{\mathbf{a}}^C)$ exists and the maximum is continuous in r^I. On $[\underline{v}^C, \overline{v}^C]$, $\mathbf{a}^C(r^I) = \mathbf{a}^C_{fb}$ if and only if $r^I = \mu/\gamma = 0$.*

Let $r^j > 0$ and $\mathbf{a}^{-j}(r^j)$ be the optimal allocation for $-j$. Then, $R^j(\mathbf{a}^{-j}(r^j))$ is decreasing in r^j, strictly if $j = C$ and $R^C(\mathbf{a}^I(r^C)) > 0$. Let $r^j_1 > r^j_2$ and $\mathbf{a}^{-j}_1$ and $\mathbf{a}^{-j}_2$ the corresponding allocations. Then,

$$W^{-j}(\mathbf{a}^{-j}_1) - r^j_1 R^j(\mathbf{a}^{-j}_1) \ge W^{-j}(\mathbf{a}^{-j}_2) - r^j_1 R^j(\mathbf{a}^{-j}_2),$$
$$W^{-j}(\mathbf{a}^{-j}_2) - r^j_2 R^j(\mathbf{a}^{-j}_2) \ge W^{-j}(\mathbf{a}^{-j}_1) - r^j_2 R^j(\mathbf{a}^{-j}_1).$$

Combining these conditions and rearranging, we get $R^j(\mathbf{a}^{-j}_2) \ge R^j(\mathbf{a}^{-j}_1)$. Let $j = C$ and $R^C(\mathbf{a}^I_2) > 0$. If $R^C(\mathbf{a}^I_2) = R^C(\mathbf{a}^I_1)$, then $W^I(\mathbf{a}^I_2) = W^I(\mathbf{a}^I_1)$ and both $\mathbf{a}^I_2$ and $\mathbf{a}^I_1$ are optimal for r^I_2. This is not possible by uniqueness of $\mathbf{a}^I_2$ (Theorem 4.1) and the fact that $\mathbf{a}^I_2$ and $\mathbf{a}^I_1$ must be different. To see this, recall that the right-hand side of (A.15) is strictly negative. So, since $r^I_2 < r^I_1$, either v^{b} or $w^I(v^{\mathrm{b}})$ must be different under $\mathbf{a}^I_1$ and $\mathbf{a}^I_2$, implying that bunching at the top involves either a different range of v's or a different action.

By Theorem 4.1 and Proposition 4.2, $R^C(\mathbf{a}^I(r^C))$ is then a continuous, decreasing function of r^C, with $R^C(\mathbf{a}^I(r^C)) \le R^C(\mathbf{a}^I_{fb})$ and $\lim_{r^C \to +\infty} R^C(\mathbf{a}^I(r^C)) = 0$. By Proposition 4.1, $R^I(\mathbf{a}^C_{fb}) < 0$. So, let $r_2 = \min\{r^C \ge 0 \mid R^C(\mathbf{a}^I(r^C)) + R^I(\mathbf{a}^C_{fb}) \le 0\}$. Clearly, $r_2 < +\infty$. Also, $r_2 > 0$ since $R^I(\mathbf{a}^C_{un}) < R^I(\mathbf{a}^C_{fb})$ and by assumption $R^C(\mathbf{a}^I_{fb}) + R^I(\mathbf{a}^C_{un}) > 0$. Now let $r_1 = \min\{r^C \ge 0 \mid R^C(\mathbf{a}^I(r^C)) +$

$R^I(\mathbf{a}^C_{un}) \leq 0\}$. By the same argument, $r_1 > 0$. Continuity of $R^C(\mathbf{a}^I(r^C))$ implies $r_1 < r_2$. The result follows from Lemma A.8.

A.7. *Proof of Proposition 4.4*

Recall the proof of Lemma A.1. By Theorem 1, p. 224, of Luenberger (1969), the optimal $(\mathbf{b}^C, \mathbf{b}^I, \mu)$ solves

$$\min_{\hat{\mu}\geq 0} \max_{\hat{\mathbf{b}}^C, \hat{\mathbf{b}}^I \text{ increasing}} L(\hat{\mathbf{b}}^C, \hat{\mathbf{b}}^I, \hat{\mu}),$$

where the minimum is achieved. Let $\mathcal{L}(\hat{\mu}; \gamma, \sigma) = L(\mathbf{b}^C, \mathbf{b}^I, \hat{\mu})$. Since r^C and r^I are continuous in μ, γ, and σ, by Theorem 4.1 and Lemma A.8, $\mathcal{L}(\mu; \gamma, \sigma)$ is continuous. W.l.o.g., we can assume that μ is bounded. To see this, note that for $\gamma \in (0, 1)$ both r^C and r^I are strictly increasing in μ. Also, $\widetilde{R}^j(\mathbf{b}^{-j}(r^j))$ is decreasing in r^j and $\widetilde{R}^I(\mathbf{b}^C(r^I)) < 0$ since $\mathbf{b}^C(r^I)$ is never constant on $[\underline{v}, \overline{v}]$. Since $\widetilde{R}^C(\mathbf{b}^I(r^C)) \to 0$ as $r^C \to +\infty$, there is $\overline{\mu} > 0$ such that $\widetilde{R}^C(\mathbf{b}^I(r^C)) + \widetilde{R}^I(\mathbf{b}^C(r^I)) < 0$ for $\mu > \overline{\mu}$. Moreover, $\mu(\gamma, \sigma)$ is unique. To see this, suppose $\mu_1 > \mu_2$ minimize $\mathcal{L}(\mu; \gamma, \sigma)$. Then, $r^C_1 > r^C_2$ and $r^I_1 > r^I_2$. Recall that $\widetilde{R}^j(\mathbf{b}^{-j}(r^j))$ is decreasing in r^j, strictly for $j = C$ if $\widetilde{R}^C(\mathbf{b}^I(r^C)) > 0$. Since $\widetilde{R}^C(\mathbf{b}^I(r^C_2)) + \widetilde{R}^I(\mathbf{b}^C(r^I_2)) \leq 0$, then $\widetilde{R}^C(\mathbf{b}^I(r^C_1)) + \widetilde{R}^I(\mathbf{b}^C(r^I_1)) < 0$ and $\mu_1 > 0$ cannot be optimal. Hence, by the Maximum Theorem, $\mu(\gamma, \sigma)$ is a continuous function with values in $[0, \overline{\mu}]$.

To show that $\mu(\gamma, \sigma)$ is strictly decreasing in σ, fix $\gamma \in (0, 1)$ and let $\sigma_1 < \sigma_2$. If $\mu(\gamma, \sigma_1) < \mu(\gamma, \sigma_2)$, then $r^C_1 < r^C_2$ and $r^I_1 < r^I_2$. By the same argument as before, $\mu(\gamma, \sigma_2) > 0$ cannot be optimal. If $\mu(\gamma, \sigma_1) = \mu(\gamma, \sigma_2) > 0$, then again $r^C_1 < r^C_2$ and $r^I_1 = r^I_2$. Hence, $\mu(\gamma, \sigma_2)$ cannot be optimal.

Let $\mu(1, \sigma) = \lim_{\gamma\to 1} \mu(\gamma, \sigma)$. Suppose $\mu(1, \sigma) > 0$. Since $r^I \to \mu(1, \sigma)$ as $\gamma \to 1$, by Lemma A.8, $\widetilde{W}^C(\mathbf{b}^C) < \widetilde{W}^C(\mathbf{b}^C_{fb})$ for γ close to 1. This cannot be optimal: letting $\mathbf{b}^I = \mathbf{b}^C = \mathbf{b}^C_{fb}$ yields a higher payoff for γ close enough to 1 and $\widetilde{R}^C(\mathbf{b}^C_{fb}) = -\widetilde{R}^I(\mathbf{b}^C_{fb})$ (recall Corollary 4.1). Now let $\mu(0, \sigma) = \lim_{\gamma\to 0} \mu(\gamma, \sigma)$. If $\mu(0, \sigma) > 0$, there is $\delta' > 0$ and $\varepsilon' > 0$ such that $\mu(\gamma, \sigma) \geq \varepsilon'$ for $\gamma \in (0, \delta')$. Then, for $\gamma \in (0, \delta')$, $r^C > 0$ which implies $\widetilde{W}^I(\mathbf{b}^I) < \widetilde{W}^I(\mathbf{b}^I_{fb})$ by Theorem 4.1. Again, this cannot be optimal: letting $\mathbf{b}^C = \mathbf{b}^I = \mathbf{b}^I_{fb}$ yields a higher payoff for γ close to 0 and $\widetilde{R}^C(\mathbf{b}^I_{fb}) = -\widetilde{R}^I(\mathbf{b}^I_{fb})$ (recall Corollary 4.1).

REFERENCES

AMADOR, M., I. WERNING, AND G. M. ANGELETOS (2006): "Commitment vs. Flexibility," *Econometrica*, 74 (2), 365–396. [1426,1427,1433,1446]

AMBRUS, A., AND G. EGOROV (2013): "A Comment on 'Commitment vs. Flexibility'," *Econometrica*, 81 (5), 2113–2124. [1426,1427,1433,1447]

AMROMIN, G. (2002): "Portfolio Allocation Choices in Taxable and Tax-Deferred Accounts: An Empirical Analysis of Tax Efficiency," Report. [1447]

——— (2003): "Household Portfolio Choices in Taxable and Tax-Deferred Accounts: Another Puzzle?" *European Finance Review*, 7, 547–582. [1447]

ASHRAF, N., N. GONS, D. S. KARLAN, AND W. YIN (2003): "A Review of Commitment Savings Products in Developing Countries," Working Paper 45, ERD. [1425,1447]

ASHRAF, N., D. S. KARLAN, AND W. YIN (2006): "Tying Odysseus to the Mast: Evidence From a Commitment Savings Product in the Philippines," *Quarterly Journal of Economics*, 121 (2), 635–672. [1425]

BATTAGLINI, M. (2005): "Long-Term Contracting With Markovian Consumers," *American Economic Review*, 95 (3), 637–658. [1427,1444]

BATTAGLINI, M., AND R. LAMBA (2013): "Optimal Dynamic Contracting: The First Order Approach and Beyond," Report, Princeton University. [1449]

BOND, P., AND G. SIGURDSSON (2013): "Commitment Contracts," Report, University of Minnesota. [1426,1427,1433,1447]

BRYAN, G., D. KARLAN, AND S. NELSON (2010): "Commitment Devices," *Annual Review of Economics*, 2 (1), 671–698. [1425]

COURTY, P., AND H. LI (2000): "Sequential Screening," *Review of Economic Studies*, 67 (4), 697–717. [1427,1428,1444,1448,1449]

DELLAVIGNA, S. (2009): "Psychology and Economics: Evidence From the Field," *Journal of Economic Literature*, 47 (2), 315–372. [1425]

DELLAVIGNA, S., AND U. MALMENDIER (2004): "Contract Design and Self-Control: Theory and Evidence," *Quarterly Journal of Economics*, 119 (2), 353–402. [1425,1426,1428,1432,1434]

ELIAZ, K., AND R. SPIEGLER (2006): "Contracting With Diversely Naive Agents," *Review of Economic Studies*, 73 (3), 689–714. [1428]

ESTEBAN, S., AND E. MIYAGAWA (2006): "Optimal Menu of Menus With Self-Control Preferences," Report, Universitat Autònoma de Barcelona. [1427,1428,1448]

GALPERTI, S. (2015): "Supplement to 'Commitment, Flexibility, and Optimal Screening of Time Inconsistency'," *Econometrica Supplemental Material*, 83, http://dx.doi.org/10.3982/ECTA11851. [1428]

GOTTLIEB, D. (2008): "Competition Over Time-Inconsistent Consumers," *Journal of Public Economic Theory*, 10 (4), 673–684. [1450]

GUL, F., AND W. PESENDORFER (2001): "Temptation and Self-Control," *Econometrica*, 69, 1403–1435. [1448]

HALAC, M., AND P. YARED (2014): "Fiscal Rules and Discretion Under Persistent Shocks," *Econometrica*, 82 (5), 1557–1614. [1427]

HEIDHUES, P., AND B. KŐSZEGI (2010): "Exploiting Naivete About Self-Control in the Credit Market," *American Economic Review*, 2279–2303. [1428]

HOLDEN, S., AND D. SCHRASS (2008): "The Role of IRAs in U.S. Households' Saving for Retirement, 2008," Investment Company Institute, 18, 1, Washington, DC. [1447]

——— (2009): "The Role of IRAs in U.S. Households' Saving for Retirement, 2009," Investment Company Institute, 19, 1, Washington, DC. [1447]

——— (2010): "The Role of IRAs in U.S. Households' Saving for Retirement, 2010," Investment Company Institute, 19, 8, Washington, DC. [1447]

HOLDEN, S., J. SABELHAUS, AND S. BASS (2010): "The IRA Investor Profile. Traditional IRA Investors' Contribution Activity, 2007 and 2008," Investment Company Institute, Washington, DC. [1447]

JEHIEL, P., AND B. MOLDOVANU (2001): "Efficient Design With Interdependent Valuations," *Econometrica*, 69 (5), 1237–1259. [1450]

KŐSZEGI, B. (2013): "Behavioral Contract Theory," *Journal of Economic Literature*, 52 (4), 1075–1118. [1427,1450]

KREPS, D. (1979): "A Representation Theorem for Preference for Flexibility," *Econometrica*, 47, 565–577. [1431]

LAIBSON, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112 (2), 443–478. [1432]

LUENBERGER, D. G. (1969): *Optimization by Vector Space Methods*. New York: Wiley. [1428,1438, 1453,1463]
MUNNELL, A. H., AND A. E. SUNDÉN (2006): *401(k) Plans Are Still Coming Up Short*. Washington, DC: Brookings Institute Press. [1447]
MUSSA, M., AND S. ROSEN (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18 (2), 301–317. [1427,1438,1444]
MYERSON, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6 (1), 58–73. [1428,1439]
——— (1986): "Multi Stage Games With Communication," *Econometrica*, 54 (2), 323–358. [1428,1435,1448]
O'DONOGHUE, T., AND M. RABIN (1999a): "Doing It Now or Later," *American Economic Review*, 89 (1), 103–124. [1432]
——— (1999b): "Incentives for Procrastinators," *Quarterly Journal of Economics*, 114 (3), 769–816. [1428]
PAVAN, A., I. SEGAL, AND J. TOIKKA (2014): "Dynamic Mechanism Design: A Myersonian Approach," *Econometrica*, 82 (2), 601–653. [1428,1444,1449]
PHELPS, E., AND R. POLLACK (1968): "On Second Best National Savings and Game-Equilibrium Growth," *Review of Economic Studies*, 35, 185–199. [1432]
ROCKAFELLAR, R. T. (1970): *Convex Analysis*. Princeton: Princeton University Press. [1454]
STROTZ, R. H. (1956): "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies*, 23, 165–180. [1431]
TOIKKA, J. (2011): "Ironing Without Control," *Journal of Economic Theory*, 146 (6), 2510–2526. [1428,1439,1456,1462]
ZHANG, W. (2012): "Endogenous Preference and Dynamic Contract Design," *The B.E. Journal of Theoretical Economics*, 12 (1). [1450]

Dept. of Economics, UCSD, La Jolla, CA 92093, U.S.A.; sgalperti@ucsd.edu.

Manuscript received August, 2013; final revision received March, 2015.