Commitment, Flexibility, and Optimal Screening of Time Inconsistency

Simone Galperti*

Northwestern University

October 1, 2012

Abstract

I study the optimal supply of flexible commitment devices to people who value both commitment and flexibility, and whose preferences exhibit varying degrees of time inconsistency. I find that, if time inconsistency is observable, then both a monopolist and a planner supply devices that enable each person to commit to the efficient level of flexibility. If instead time inconsistency is unobservable, then both face a screening problem. To screen a more time-inconsistent from a less time-inconsistent person, the monopolist and (possibly) the planner inefficiently curtail the flexibility of the device tailored to the first person, and include unused options in the device tailored to the second person. These results have important policy implications for designing special savings devices that use tax incentives to help time-inconsistent people adequately save for retirement.

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

JEL CLASSIFICATION: D42, D62, D82, D86, D91, G21, G23.

---

*Department of Economics, Northwestern University, Evanston, IL 60208 (E-mail: simonegalperti2008@u.northwestern.edu). I am indebted to Eddie Dekel and Alessandro Pavan for many, long, and fruitful discussions that greatly improved the paper. I also thank Gene Amromin, Stefano DellaVigna, Roland Eisenhuth, 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 University, New York University, UC Berkeley, EEA-ESEM 2012. I gratefully acknowledge financial support from the Center of Economic Theory of the Weinberg College of Arts and Sciences of Northwestern University. All remaining errors are mine.

1 Introduction

Evidence from economics and psychology shows that people have self-control problems (see, e.g., DellaVigna (2009)). Signs of such problems are, for example, chronic splurging, amassing excessive credit-card debt, and saving too little for retirement. People often understand their self-control problems and thus look for commitment devices. At the same time, they also face uncertainty about their future and thus they value flexibility.[1] These conflicting desires create a demand for flexible commitment devices. Examples include illiquid savings devices such as savings accounts, individual retirement accounts (IRAs), and 401(k) plans.[2]

To get a sense of the importance of such devices, consider the U.S. retirement market. As of June 2012, this market had assets of $18.5 trillion, which is roughly 120% of U.S. GDP and 36% of all U.S. household financial assets. Of these, $3.3 trillion (17.8%) was in 401(k) plans, and $5.1 trillion (26.5%) was in IRAs.[3] IRAs and 401(k) plans are examples of "tax-shielded" investment accounts, which obey special tax rules that set limits on contributions to their balances and penalties for withdrawals. Authorizing such special accounts is also particularly costly: For the fiscal years 2010-2014, the resulting burden on the U.S. federal budget amounts to $711 billion (12.7%) of the estimated tax expenditures.[4]

This paper offers a novel theoretical investigation of markets for flexible commitment devices. It is the first to consider simultaneously three essential aspects of such markets. First, people demand devices that provide both commitment and flexibility. Second, people differ in their degrees of self-control, so they demand different devices (see, e.g., Ashraf, Karlan, and Yin (2006)). Third, the provider of these devices—whether it is a profit-maximizing firm, or a welfare-maximizing government—cannot observe people's degrees of self-control when tailoring its supply to individual needs. Thus, it faces the problem of screening each individual according to his preference for commitment and for flexibility. Understanding this problem and characterizing its profit- and welfare-maximizing solutions represent the core of the present paper.[5]

I model a market for flexible commitment devices using a dynamic principal-agent framework with two periods.[6] In period 1, the principal offers the agent an incentive device, which allows him to choose among several actions in period 2, and for each action specifies a payment to the principal. As an example, consider a savings device that allows the agent to contribute to and withdraw from its balance by paying some fee. In period 1, agents desire flexible devices because their preference over period-2 actions depends on an unknown (non-contractible) state (Kreps (1979)). Moreover, in period 1 some agents also desire commitment, because for each state their

[1] For an alternative analysis of the trade-offs between commitment and flexibility see, e.g., Amador, Werning, and Angeletos (2006). I will compare my analysis with theirs in Section 6.2.

[2] Other examples include gym memberships with large upfront fees but no per-visit charge, automatic drafts from checking to investment accounts, Christmas clubs, rotating savings and credit associations, and microcredit savings accounts in developing countries. (See Ashraf, Gons, Karlan, and Yin (2003), Ashraf, Karlan, and Yin (2006), DellaVigna and Malmendier (2006), Bryan, Karlan, and Nelson (2010)).

[3] Investment Company Institute, "The U.S. Retirement Market, Second Quarter 2012" (September 2012).

[4] Source: Estimates of The Federal Tax Expenditures for the Fiscal Years 2010-2014 (2010)

[5] In their survey, Bryan, Karlan, and Nelson (2010) raise the question of whether markets can provide products to solve people's commitment problems, but the answer lies beyond the scope of their paper. Laibson (1998) questions the optimality of the retirement instruments created by the U.S. government. Amador, Werning, and Angeletos (2006) study the design of paternalistic commitment policies such as minimum-savings requirements.

[6] DellaVigna and Malmendier (2004), and Eliaz and Spiegler (2006) adopt a similar approach. Hereafter, the principal is a she and the agent is a he.

preference is time inconsistent (Strotz (1956)). Finally, the agents' self-control problems are unobservable, because they privately know their degree of time inconsistency (i.e., their type). For clarity's sake, in most of the paper I focus on a two-type case: agents of type C are time consistent, and agents of type I are all time inconsistent to the same degree. In Section 4.3 I extend the results to markets with more realistic heterogeneity.

In Section 3 I show that, if the agents' types were observable, then the principal would offer each type a (possibly different) device with a perfect mix of commitment and flexibility—independently of whether she maximizes profits or welfare. Such a device helps each type commit to a flexible plan of action that is efficient in each state from the viewpoint of his period-1 preference.[7] I call this outcome efficient, adopting the standard interpretation of the agent's period-1 preference as his long-run preference (O'Donoghue and Rabin (1999)). I also show that, in some cases, the principal can sustain the efficient outcome using the same device with all types—so their unobservability is irrelevant. Otherwise, it creates a screening problem.

This screening problem arises because, thanks to his stronger self-control, type C values any flexible device strictly more than does type I. So, if the principal wants to sustain a flexible outcome with type I, she must grant type C enough rents for him not to pretend to be I. Unfortunately, these rents can jeopardize the incentives of type I to reveal his time inconsistency. This second aspect of the screening problem, however, admits an unconventional remedy. Remember that type I desires commitment. Therefore, the principal can promote his honesty by adding to the device of type C tempting items that type C *never* uses, but type I would. These items make type I view the device of type C as offering less commitment, and thus deter I from pretending to be C.

I derive the optimal screening devices in Section 4.2. First, when designing the device for type I, the principal faces a standard trade-off between maximizing efficiency with type I and extracting rents from type C. Intuitively, since these rents depend on the flexibility of the device of type I, the principal should curtail it below the efficient level. She can, however, do so in many ways, and finding the optimal one represents a novel, challenging question. I illustrate the answer with a concrete example in Section 1.1 and present the general result in Section 4.2.2. Second, when designing the device of type C, the principal always wants to maximize its efficiency, but she must also take into account—as noted—that this device may induce type I to pretend to be C. To prevent this, she may have to modify the device of type C—relative to the symmetric-information, efficient one—in two ways: (1) she may add unused items;[8] and (2) she may also sustain an inefficient outcome with type C. Property (2) is a failure of the 'no distortion at the top' property, which marks previous screening models, both static (e.g., Mussa and Rosen (1978)) and dynamic (e.g., Courty and Li (2000)). Comparing the monopolist's and the planner's optimal device for type C, I show that if the monopolist has to do (1) (respectively (2)), then the planner does as well.

This paper contributes to the growing literature on behavioral contract theory.[9] It adds to this literature the result that, when a profit-maximizing firm faces diversely time-inconsistent agents, it screens them using contracts with different degrees of flexibility. Moreover, the paper characterizes the specific form that this screening takes: The contracts for time-inconsistent

[7] A related result appears in DellaVigna and Malmendier (2004), as discussed in Section 6.2.

[8] A related result appears in Esteban and Miyagawa (2005), as discussed in Section 6.2.

[9] This literature includes DellaVigna and Malmendier (2004, 2006); Esteban and Miyagawa (2005); Eliaz and Spiegler (2006, 2008); Heidhues and Koszegi (2010); and Spiegler (2011) among others.

agents inefficiently curtail flexibility as illustrated in Section 1.1; instead, the contracts for time-consistent agents may intentionally inflate flexibility by offering more options than such agents deem useful.

This paper also contributes to the study of optimal paternalism (O'Donoghue and Rabin (2003)). It demonstrates that paternalistic planners play an important role in providing flexible commitment devices to people with self-control problems: When such problems are unobservable, profit-maximizing firms create inefficiencies especially for the people who need commitment. The paper also derives the best paternalistic provision of such devices. To illustrate its properties, in Section 5 I address this question: How should the government design special savings devices (SDs), with specific tax rules that help time-inconsistent agents appropriately save for retirement? The main points can be summarized as follows: (1) like firms, the government cannot observe people's self-control problems, and therefore it faces a trade-off between the corrective and the redistributive role of taxation; (2) solving this trade-off calls for curtailing the liquidity of the special SDs relative to the standard SDs; and (3) the optimal way to do so involves limiting, through binding caps, the ability both to tap and to contribute to a special SD.

I find that my policy implications are consistent with some general features of the U.S. regulation of IRAs and 401(k) plans, which are special SDs with tax incentives. I also find that some evidence in Amromin (2002, 2003) is consistent with the principle that curtailing the liquidity of the special SDs—namely, the IRAs and the 401(k) plans—helps keep people with stronger self-control choosing the standard SDs—namely, the so-called 'regular taxable accounts.' I believe that my results are useful for understanding some effects of the current U.S. regulation of SDs, and for assessing its future revisions.

Finally, this paper relates to the literature on dynamic mechanism design. First, it highlights an important methodological point: In models with time-inconsistent agents, one cannot in general restrict attention to mechanisms that are defined only on the path of play. Indeed, to characterize the optimal devices, one *must* allow for options that are off path. This can never occur in standard models without time inconsistency. Second, deriving the devices that optimally screen the agents' time inconsistency requires some non-standard and recent techniques. For reasons that I will explain later, to handle the incentive constraints involving the agents' types, I use Lagrangian methods from Luenberger (1969). This approach differs from a standard optimal-control approach and the standard dynamic-mechanism-design approach (Courty and Li (2000); Pavan, Segal, and Toikka (2011)). Moreover, to ensure that each device satisfies a necessary monotonicity property in the period-2 states, I adapt Toikka's (2011) generalization of Myerson's (1981) ironing procedure to my setting with off-path options.

In Section 6, I provide a more detailed literature review; I also discuss the possibility of introducing overconfident agents into the model. Section 7 concludes. All proofs are in the appendix.

1.1 An Illustrative Example: Savings Devices

Consider the following situation. In period 1, the principal offers savings devices (SDs) to the agent, who is planning his future savings. An SD allows the agent to contribute to and withdraw from its balance in period 2 (his working life) at some fee, thereby providing a tool to smooth consumption between that period and some future period of retirement.[10] Specifically, given an SD, in period 2 the agent chooses the action a and pays the fee p, where $a < 0$ for contributions and $a > 0$ for withdrawals. For each a, SDs give a return at retirement, according to a rate that I assume fixed and exogenous. On the other hand, the principal can design different SDs by changing how the fee p depends on the action a. Finally, in this example, the principal incurs no cost to provide SDs.

The agent's choice among SDs depends on how he evaluates, in period 1, the decisions that he expects to make in period 2. In period 1 the agent is uncertain about what action a is optimal in period 2, as this decision hinges on an unknown state s—e.g., his period-2 health. Specifically, in period 1 he assigns to a the direct utility $sa + d(a)$ in state s; for instance, think of sa as the period-2 benefit of withdrawing from an SD, and of $d(a)$ as the resulting expected well-being at retirement. So, in period 1 the agent evaluates his expected period-2 choices using the utility $sa + d(a) - p$. However, when actually choosing a in period 2, the agent may use a different utility function. Specifically, if he is of type C, in period 2 he still assigns to a the direct utility $sa + d(a)$. Instead, if he is of type I, in period 2 he systematically overweights the immediate utility of withdrawals and disutility of contributions; he is therefore time inconsistent. To capture this inconsistency, I will let $\phi sa + d(a)$, with $\phi > 1$, be the direct utility that type I derives from a in period 2. In state s, the agent then chooses a based on its period-2 direct utility net of the associated fee p.

Suppose that, according to the period-1 utility $sa + d(a)$, for some states it is optimal to contribute to an SD ($a^* < 0$), but for others (e.g., unexpected health problems) it is optimal to withdraw from it ($a^* > 0$). For simplicity, let this optimal plan correspond to the function $\mathbf{a}^*(s) = s - s_0$, where s ranges from $\underline{s}$ to $\overline{s}$.[11] Note that $\mathbf{a}^*$ defines what I call the efficient outcome, since the principal's costs are zero.

Before describing the optimal SDs, I want to briefly give some intuition for why type C values strictly more any flexible SD provided to type I, than does I himself. Suppose that the principal wants to help type I commit to the plan $\mathbf{a}^*$. To do so, the SD of type I must feature penalties for withdrawals and rewards for deposits, since in period 2 type I cares systematically more about his immediate utility than about his retirement well-being. How does type C evaluate such an SD in period 1? Since type C has stronger self-control, roughly speaking, he expects to reap the rewards and avoid the penalties more often than does type I; so C expects a larger payoff from the SD of type I than does I himself. In contrast, suppose that the principal offers type I an SD with only one option, a^{nf}, i.e., no flexibility at all. Then, type I doesn't need any penalty or reward to commit to a^{nf}, and type C cannot benefit from his stronger self-control. More generally, by making the SD of type I less flexible, the principal can reduce the penalties and rewards that help type I commit. These changes curtail how much type C expects to benefit from lying in period 1, and consequently reduce the rents that keep C from doing so.

I shall now describe the SDs that the principal offers to screen the agents' types. I consider the case in which she maximizes her expected profits from the fees p. The detailed derivation is in Section 4.2.

Consider first the SD designed for type I. Figure 1 illustrates his resulting savings decisions—

[10] In the language of DellaVigna and Malmendier (2004), contributions are "investment" goods and withdrawals are "leisure" goods.

[11] This plan is optimal if, for example, $d(a) = -\frac{1}{2}(a + s_0)^2$ with $\underline{s} < s_0 < \overline{s}$.

the function denoted by $\boldsymbol{\chi}^I$—and compares them with the efficient plan $\mathbf{a}^*$.

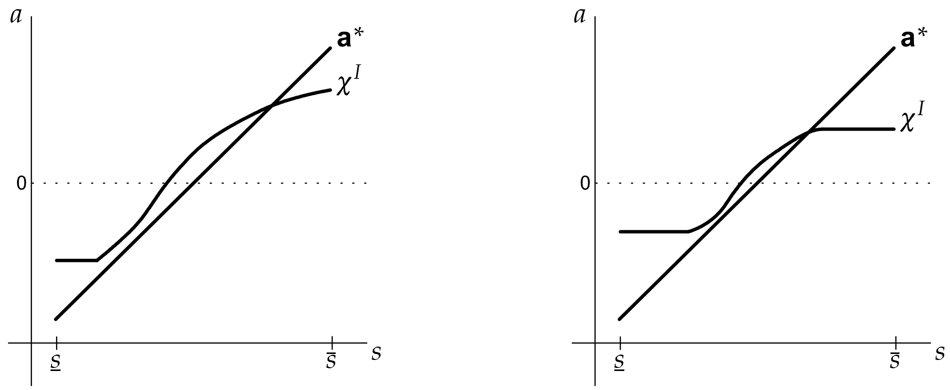

Figure 1: Efficient Outcome vs. Second Best with Type I

By comparing $\boldsymbol{\chi}^I$ and $\mathbf{a}^*$, we see exactly how the principal curtails the flexibility with which type I adjusts his savings decision to the period-2 states. Specifically, there are three key properties: (1) type I obtains fewer savings options than in the efficient plan, since the range of $\boldsymbol{\chi}^I$ is a strict subset of that of $\mathbf{a}^*$; (2) type I will contribute too little for small states, where $\boldsymbol{\chi}^I(s) > \mathbf{a}^*(s)$, and will withdraw too little for large states, where $\boldsymbol{\chi}^I(s) < \mathbf{a}^*(s)$; and (3) type I will not adjust at all his savings decisions to some extreme states, where $\boldsymbol{\chi}^I$ is constant. This last property always occurs for very small states; it also occurs for very large states (as in the right panel) if, for example, the share of type I in the market is small. As I show in the rest of the paper (see Proposition 3), properties (1)-(3) hold quite generally, as long as the principal's goal assigns some weight to profits. Concretely, we can think of the SD offered to type I as an account that, besides possibly rewarding contributions and punishing withdrawals, also sets limits on these transactions. Such limits appear, for example, in some individual-development accounts—a form of matched savings accounts—and Christmas-club accounts offered by some financial institutions in the U.S. (see also Ashraf, Gons, Karlan, and Yin (2003)).

Consider now the SD designed for type C. Figure 2 compares the savings options that this SD offers type C—the function denoted by $\mathbf{x}^C$—with the efficient plan $\mathbf{a}^*$.

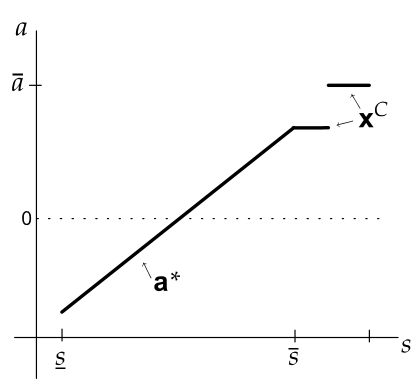

Figure 2: Efficient Outcome vs. Second Best with Type C

We see that the device of type C commits him to the efficient savings plan $\mathbf{a}^*$ on the relevant range $[\underline{s}, \overline{s}]$. But it also includes one option that involves an inefficiently high withdrawal ($\overline{a}$) in period 2, and that type C never uses because he deems it reasonable only for s above the largest possible state $\overline{s}$. I shall later discuss how the presence of unused options depends on the primitives of the model.

2 The Model

A principal interacts with an agent over two periods. In period 1 (the contracting stage), the principal offers the agent a device that gives him the opportunity to choose an action a in period 2 (the consumption stage).[12] The contractible action a belongs to a feasible set $A := [\underline{a}, \overline{a}] \subset \mathbb{R}$ with $-\infty < \underline{a} < \overline{a} < +\infty$. For each a, a device specifies a payment p from the agent to the principal. The principal incurs a production cost in period 2, which is given by the differentiable, non-decreasing, and convex function $c : A \to \mathbb{R}$. The principal can perfectly commit to a device, which—if accepted by the agent—is then binding for *both* parties. I also assume that there is no spot market for a in period 2.

Agents may have time-inconsistent preferences, but they are always sophisticated. To model time-inconsistent agents, I follow Strotz (1956). For each agent, I call *self-1* the self who chooses a device and *self-2* the self who chooses an option from the resulting menu. Both selves' preferences depend on a state s, which occurs in period 2 and reflects shocks to taste, income, or health. These shocks induce self-1 to desire flexibility; moreover, they are not contractible, for example because only the agent observes them.[13] Conditional on s, the *direct* utilities of self-1 and self-2 from action a are

$$u_1(a; s) := sa + d(a) \quad \text{and} \quad u_2(a; s, \phi) := \phi sa + d(a),$$

where the function $d : A \to \mathbb{R}$ is differentiable and strictly concave. Finally, agents' preferences are quasi-linear in the payment p.

The positive parameter ϕ determines the preferences' degree of time inconsistency, and leads self-1 to desire commitment. When $\phi \neq 1$, self-1 foresees that, in each state, his self-2 trades off the benefit and the cost of any action a in a systematically different manner. This modeling assumption is based on the idea, proposed by cognitive psychologists, of 'salience:' Decision-makers seem to perceive the benefit of their actions (sa) as more (or less) salient than their cost ($d(a)$) depending on whether the decision they face is immediate, or occurs in the future (see, e.g., Akerlof (1991) and the references therein).

The illustrative example in Section 1.1 gives one interpretation of the model for the case in which the principal offers savings devices. The model can however capture other situations. For instance, the principal may offer gym memberships as in the next example.[14]

[12] Alternatively, we could interpret a as the utils resulting from some underlying contractible action.

[13] If s were contractible, the perfect ability of the principal to commit would make the solution for the agents' time inconsistency trivial.

[14] The role of gym memberships as commitment devices has been extensively studied, e.g., by DellaVigna and Malmendier (2004, 2006).

Example 1 *In period 1, a gym offers the agent a membership contract. Becoming a member allows the agent to exercise in period 2 (e.g., the following month) at the gym's facility; the time he spends there determines a fee* p*. Exercising involves some immediate discomfort* $a < 0$ *for the agent, which is proportional to the time he exercises, but it improves his physical health by* $d(a)$*. In period 1, the agent's self-1 anticipates that his willingness to exercise will depend on how tired he will be after work or whether he will be sick* (s)*. Furthermore, the self-1 may foresee that, when the time comes, the self-2 systematically overweights the downsides of going to the gym and tends to exercise less than the self-1's goal. To capture the agent's preferences, we can use the functions* $sa + d(a) - p$ *for his self-1 and* $\phi sa + d(a) - p$*, with* $\phi > 1$*, for his self-2*

As far as period-2 states are concerned, s belongs to the set $S := [\underline{s}, \overline{s}] \subset \mathbb{R}$ with $0 < \underline{s} < \overline{s} < +\infty$. This assumption simply says that, in each state, the agent assigns a weight to the benefit and the cost of his actions which is bounded away from zero. States have distribution F, which has continuous and positive density f and is commonly known in period 1.

For clarity's sake, most of the analysis focusses on the case with only two types: agents of type C are time consistent, and agents of type I are all time inconsistent to the same degree.[15] I assume that type I systematically overweights the benefit of his actions, since agents with self-control problems typically tend to consume more, eat more, or exercise less than they would like to. Formally, type I has $\phi^I > 1$ and type C has $\phi^C = 1$; also, let the fraction of agents of type C be $\gamma \in (0, 1)$. Because the agents are sophisticated, they know their degree of time inconsistency ϕ before contracting occurs. In contrast, the principal cannot observe ϕ; she, however, knows the possible types and the share γ.

To evaluate any device in period 1, type i has to take into account what his self-2 does in period 2. Therefore, I restrict attention to devices for which self-2's optimal decisions are always well-defined, independently of the agent's type. So, the payoff to self-1 of any device is just the expected utility of the resulting decisions of self-2, computed using the period-1 preference. Finally, if an agent rejects all the principal's offers, he receives the outside option whose value is normalized to zero.[16]

I adopt the following definition of efficiency, which adheres to the standard interpretation of the self-1's preference as the long-run preference of the agent (O'Donoghue and Rabin (1999)).

Definition 1 (Efficiency) *For each state* s*, the* ex-ante social surplus *of action* a *is* $u_1(a; s) - c(a)$*, and the* efficient outcome *is*

$$\mathbf{a}^*(s) := \arg\max_{a \in A} u_1(a; s) - c(a).$$

I also assume that it is never efficient to induce the smallest and the largest action, and that the maximum ex-ante social surplus is positive in every state.

Assumption 1 *For every* s*,* $\mathbf{a}^*(s)$ *is interior and* $u_1(\mathbf{a}^*(s); s) - c(\mathbf{a}^*(s)) > 0$*.*

By standard arguments, the outcome $\mathbf{a}^*$ is already non-decreasing; by Assumption 1, $\mathbf{a}^*$ becomes increasing. Thus, $\mathbf{a}^*$ involves choosing a different action in each state and defines the efficient, benchmark level of flexibility.

[15] I extend my results to the case with more than two types in Section 4.3.

[16] I discuss the case in which the outside option has type-dependent values in Section 6.1.

Finally, I assume that, when designing her devices in period 1, the principal maximizes a weighted sum of her expected profits and the expected ex-ante social surplus, with weight $\pi \in [0,1]$ on profits and $1-\pi$ on welfare. This assumption is simply a convenient way to cover the case of a monopolist ($\pi = 1$), as well as the case of a paternalistic planner who may also have to worry about the profitability of her devices ($\pi < 1$). Intuitively, this situation may arise when the planner is the exclusive provider of commitment devices and has limited resources to finance them, or when she only regulates the market of such devices and has to ensure that third-party providers expect enough profits to enter the market. In these cases, I could alternatively let the planner maximize the expected ex-ante social surplus subject to the constraint of generating some minimum (possibly negative) level of profits. As I show in Appendix C, this alternative specification would only make π endogenous, without affecting my results.

3 Symmetric Information: A Benchmark

In this section I show that, if the agents' degree of time inconsistency ϕ were observable, then the principal would offer each type a device that sustains the efficient outcome, independently of how much she cares about profits. In so doing, she would provide a perfect mix of commitment and flexibility.

When the agent's type is observable, by the Revelation Principle, I can characterize optimal devices using direct revelation mechanisms (DMs) that induce the agent to report truthfully the state s in period 2. Formally, a DM is a pair of functions $(\mathbf{a}, \mathbf{p})$, where $\mathbf{a} : S \to A$ and $\mathbf{p} : S \to \mathbb{R}$. DMs must satisfy the constraints

$$\int_S \left[u_1\left(\mathbf{a}\left(s\right); s\right) - \mathbf{p}\left(s\right)\right] dF\left(s\right) \geq 0 \qquad (IR)$$

and

$$u_2\left(\mathbf{a}\left(s\right); s, \phi\right) - \mathbf{p}\left(s\right) \geq u_2\left(\mathbf{a}\left(s'\right); s, \phi\right) - \mathbf{p}\left(s'\right) \text{ for all } s, s'. \qquad (IC)$$

Note that the *IR* constraint uses the preference of self-1, whereas the *IC* constraints use the preference of self-2. Subject to these constraints, the principal then chooses $(\mathbf{a}, \mathbf{p})$ to maximize

$$(1-\pi)\int_S [u_1\left(\mathbf{a}\left(s\right); s\right) - c(\mathbf{a}\left(s\right))] dF\left(s\right) + \pi \int_S \left[\mathbf{p}\left(s\right) - c\left(\mathbf{a}\left(s\right)\right)\right] dF\left(s\right).$$

Although changing π modifies the principal's goal, it turns out that she always helps time-inconsistent agents fully solve their self-control problems.

Lemma 1 (First Best) *Suppose that the agents' degree of time inconsistency ϕ is observable. Then, for any ϕ, the principal offers a device that sustains the efficient outcome $\mathbf{a}^*$ for all $\pi \in [0,1]$.*

The intuition for Lemma 1 is as follows. Since the agent has no private information at the contracting stage, he has no advantage over the principal. Thus, the principal becomes the residual claimant of the utility that the agent expects when choosing a device, i.e., of the expected utility of the agent's *self-1.* Hence, independently of π, the principal's goal becomes to maximize the expected ex-ante surplus. There is, however, a twist here, compared to a model with only time-consistent agents: When $\phi > 1$, the principal has to make sure that the agent's *self-2* is willing to go along with her efficient plan. Doing so is feasible, despite time-inconsistency, because self-1 and self-2 agree that moving from any state to a higher one must correspond to choosing a (weakly) larger action. As I discuss in Section 6.2, DellaVigna and Malmendier (2004) prove a result similar to Lemma 1, which relies on the same intuition.

The goal of the present paper is to explain why and how the agent's private information about his time inconsistency affects the principal's supply of flexible commitment devices. As I show in Appendix A, if the number of states is finite and the gap between them is large enough relative to the degree ϕ^I, then the agent's private information plays no role with regard to sustaining the efficient outcome. However, with a continuum of states, private information always matters.

4 Asymmetric Information

4.1 Screening Problem and Implementable Outcomes

In this section, I explain how the agents' private information about their degree of time inconsistency ϕ causes a screening problem. Understanding this mechanism is interesting on its own, since it is not obvious a priori how ϕ shapes the agents' incentives to choose among commitment devices. Furthermore, to find the devices that the principal will offer, we first need to understand how asymmetric information about ϕ restricts the outcomes that she can implement.

The initial step consists in describing how agents choose from a menu of actions and payments in period 2, and how they consequently choose among devices (i.e., menus) in period 1. We can again do so using DMs. Note that, in each state, the self-2 preference of type i is pinned down by the single number $\phi^i s$. So, call these numbers the self-2 'valuations' of action a, and represent them with the set

$$\Theta^i := \{\theta \mid \theta = \phi^i s, s \in S\} = [\underline{\theta}^i, \overline{\theta}^i].$$

By the Revelation Principle (Myerson (1986)), we can then focus on DMs that satisfy two properties: (1) they assign an action-payment pair to the agent's sequential reports of ϕ and θ—these reports correspond to choosing a device and then one of its options; and (2) they ensure that it is optimal to report truthfully ϕ in period 1 and to report truthfully θ in period 2, *conditional* on having truthfully reported ϕ in period 1. To easily track the agents' behavior after *any* period-1 report, I consider DMs that assign action-payment pairs to coherent as well as incoherent reports, like ϕ^C followed by $\theta > \overline{\theta}^C$. Formally, define $\Theta := \left[\underline{\theta}, \overline{\theta}\right]$ where $\underline{\theta} := \underline{\theta}^C$ and $\overline{\theta} := \overline{\theta}^I$ (i.e., $\Theta = co(\Theta^C \cup \Theta^I)$). A DM is then a collection of functions $\{\mathbf{X}, \mathbf{T}\} = (\mathbf{x}^i, \mathbf{t}^i)_{i=C,I}$, where $\mathbf{x}^i : \Theta \to A$ is an *allocation rule* and $\mathbf{t}^i : \Theta \to \mathbb{R}$ is a *payment rule*. It remains to ensure that $\{\mathbf{X}, \mathbf{T}\}$ also satisfies property (2).

The self-1's incentives to report ϕ depend on what he expects the self-2 to choose afterward. So, we need to work backward, starting from the self-2's incentives to report truthfully θ for any self-1's report of ϕ. With a slight abuse of notation, let $u_2(a;\theta) := u_2(a; s, \phi)$ for $\theta = \phi s$. I require that the DMs satisfy the following property.[17]

[17] Although this property is stronger than the second part of property (2), it entails no loss of generality. To see this, fix the device that the principal designs for type j. Suppose that, *given such a device*, θ^j and θ^i, with $\theta^j \neq \theta^i$, both choose the pair (a, p). We can describe this behavior with a P2-IC DM that offers two copies

Definition 2 (Period-2 Incentive Compatibility) *A DM* $\{\mathbf{X},\mathbf{T}\}$ *is* period-2 incentive compatible (P2-IC) *if and only if, for* $i=C,I$ *and every* $\theta,\theta'\in\Theta$*,*

$$u_2(\mathbf{x}^i(\theta);\theta)-\mathbf{t}^i(\theta)\geq u_2(\mathbf{x}^i(\theta');\theta)-\mathbf{t}^i(\theta').$$

By standard arguments, $\{\mathbf{X},\mathbf{T}\}$ is P2-IC if and only if two conditions hold: for $i=C,I$,

$$\mathbf{x}^i\in\mathcal{M}:=\{\mathbf{x}:\Theta\to A\ \mid \mathbf{x}\text{ non-decreasing}\}, \qquad (MON^i)$$

and

$$\mathbf{t}^i(\theta)=u_2(\mathbf{x}^i(\theta);\theta)-\int_{\underline{\theta}}^{\theta}\mathbf{x}^i(y)\,dy+k^i\text{ for each }\theta, \qquad (ENV^i)$$

where $k^i:=\mathbf{t}^i(\underline{\theta})-u_2(\mathbf{x}^i(\underline{\theta});\underline{\theta})$. Hereafter I will consider only P2-IC DMs, which I will denote by $\{\mathbf{X},\mathbf{k}\}=(\mathbf{x}^i,k^i)_{i=C,I}$ and I may simply call DMs. Similarly, I will refer to the pair $(\mathbf{x}^i,k^i)$ as the *i-menu*, as it fully specifies the menu of options included in the device offered to type i.

Since DMs assign actions to coherent and incoherent reports of θ, we need to distinguish an allocation rule $\mathbf{x}^i$ from the outcome that it implies when type i chooses the i-menu.

Definition 3 (Outcome sustained by the i-menu) *Let* $\{\mathbf{X},\mathbf{k}\}$ *be a P2-IC DM and* $\boldsymbol{\chi}^i:\Theta^i\to A$*. I say that the i-menu* $(\mathbf{x}^i,k^i)$ *sustains the outcome* $\boldsymbol{\chi}^i$ *with type i if the restriction of the allocation rule* $\mathbf{x}^i$ *to* Θ^i *equals* $\boldsymbol{\chi}^i$ *(i.e.,* $\mathbf{x}^i|_{\Theta^i}=\boldsymbol{\chi}^i$*).*

In particular, menus that sustain the efficient outcome $\mathbf{a}^*$ will be of focal interest. To easily compare them with other menus, I express $\mathbf{a}^*$ in terms of θ by defining $\boldsymbol{\chi}^{i*}(\theta):=\mathbf{a}^*(\theta/\phi^i)$ for $\theta\in\Theta^i$ and $i=C,I$. Also, to describe i-menus containing only the actions required to sustain the outcome $\boldsymbol{\chi}^{i*}$, I will use the allocation rules $\mathbf{x}^{i*}$, which coincide with $\boldsymbol{\chi}^{i*}$ over Θ^i and satisfy $\mathbf{x}^{C*}(\theta)=\mathbf{x}^{C*}(\overline{\theta}^C)$ for $\theta>\overline{\theta}^C$, and $\mathbf{x}^{I*}(\theta)=\mathbf{x}^{I*}(\underline{\theta}^I)$ for $\theta<\underline{\theta}^I$.

Having restricted attention to P2-IC DMs, we can now easily compute the utility that each type expects when choosing a menu, and hence study his incentives to report ϕ truthfully. Specifically, let F^i be the distribution that F induces on Θ conditional on being type i, and let H^i be its inverse hazard rate. The utility that i's self-1 expects from the j-menu is

$$\begin{aligned} U^i(\mathbf{x}^j,k^j) &= E\left[u_1(\mathbf{x}^j(\theta);\theta/\phi)-\mathbf{t}^j(\theta)\mid\phi=\phi^i\right] \qquad (EU)\\ &= \int_{\underline{\theta}^i}^{\overline{\theta}^i}\mathbf{x}^j(\theta)\,v^i(\theta)\,dF^i(\theta)+\int_{\underline{\theta}}^{\underline{\theta}^i}\mathbf{x}^j(\theta)\,d\theta-k^j, \end{aligned}$$

where I used ENV^j to substitute for $\mathbf{t}^j(\theta)$ and $v^i(\theta)=\theta/\phi^i-(\theta-H^i(\theta))$. Then, to ensure truthful reports of ϕ as well as participation, a DM $\{\mathbf{X},\mathbf{k}\}$ must satisfy the standard constraints

$$U^i(\mathbf{x}^i,k^i)\geq U^i(\mathbf{x}^j,k^j)\text{ for }i\neq j, \qquad (IC^i)$$

and

$$U^i(\mathbf{x}^i,k^i)\geq 0\text{ for }i=C,I. \qquad (IR^i)$$

The constraints IC^i are the key to understanding the problem of screening the agents' degrees of time inconsistency, and the resulting restrictions on the set of implementable outcomes. Examining such constraints will therefore be the focus of the rest of this section.

of (a,p) after a report of ϕ^i, one indexed by θ^i and one by θ^j. The same applies if θ^i and θ^j make different decisions. For more details see, e.g., Pavan (2007).

To ensure that type i truthfully reports his degree of time inconsistency ϕ^i, we first need to understand his incentives to pretend to be type j. To do so, define

$$R^i(\mathbf{x}^j) := U^i(\mathbf{x}^j, k^j) - U^j(\mathbf{x}^j, k^j).$$

and re-write IC^i as

$$U^i(\mathbf{x}^i, k^i) \geq U^j(\mathbf{x}^j, k^j) + R^i(\mathbf{x}^j) \text{ for } i \neq j. \qquad (R\text{-}IC^i)$$

We see that i's willingness to mimic j depends on two things: the first, $U^j(\mathbf{x}^j, k^j)$, is the overall expected payoff to j; the second, $R^i(\mathbf{x}^j)$, is how much i expects to gain or lose, relative to j, by facing the j-menu while being of type i. The function $R^i(\mathbf{x}^j)$ is crucial, because it determines whether type i will enjoy information rents due to the allocation sustained with type j. By analogy, recall that in the screening model of Mussa and Rosen (1978), a high-valuation buyer has an incentive to misreport his type because he values more any quality offered to low-valuation buyers (the analog of $\mathbf{x}^j$) than they do. This advantage ultimately determines his information rents.

In my model, it is type C who enjoys an information advantage at the contracting stage, provided that the I-menu involves some flexibility. The next proposition presents this result; an intuitive explanation follows.

Proposition 1 *Time-consistent agents enjoy an information advantage over time-inconsistent agents: For any direct revelation mechanism* $\{\mathbf{X}, \mathbf{T}\}$ *that is period-2 incentive compatible,* $R^C(\mathbf{x}^I) \geq 0$ *and* $R^I(\mathbf{x}^C) \leq 0$*. Furthermore, the (dis)advantage of one type disappears if and only if the menu of the other type involves no flexibility: For* $i \neq j$, $R^i(\mathbf{x}^j) = 0$ *if and only if* $\mathbf{x}^j$ *is constant over* $(\underline{\theta}, \overline{\theta})$.[18]

To see the intuition for Proposition 1, recall our initial example about SDs and, for simplicity, suppose that there are only two states: $s_w > s_d$. Consider first the I-menu, and suppose that it makes type I withdraw $a_w > 0$ in s_w and deposit $a_d < 0$ in s_d. To do so, the I-menu must provide I's self-2 with the right incentives: Intuitively, the payment 'premium' $p_w - p_d$ must be big enough to prevent I's self-2 from choosing a_w in s_d. Since this premium is tailored to the preference of I's self-2, it may induce C's self-2 to make different contingent choices, if C faces the I-menu. The key observation is that C's choices reflect the common self-1 preference. Therefore, if in period 2 C's choices from the I-menu differ from I's, then in period 1 C must assign to the I-menu a strictly higher expected payoff than does I (i.e., $R^C(\mathbf{x}^I) > 0$). Conversely, suppose that the C-menu also makes type C withdraw a_w in s_w and deposit a_d in s_d. To do so, now the premium $\hat{p}_w - \hat{p}_d$ must be tailored to the preference of C's self-2. Therefore, it may induce I's self-2 to make different contingent choices, if I faces the C-menu. It follows that, if in period 2 I's choices from the C-menu differ from C's, then in period 1 I must assign to the C-menu a strictly lower expected payoff than does C (i.e., $R^I(\mathbf{x}^C) < 0$). The proof of Proposition 1 generalizes these observations by showing that, with a continuum of states, C's and I's choices from the j-menu differ with positive probability unless $\mathbf{x}^j$ corresponds to a singleton menu.[19]

[18] The allocation $\mathbf{x}^j$ can jump at $\underline{\theta}$ and $\overline{\theta}$ without making $R^i(\mathbf{x}^j) \neq 0$ simply because the distribution F is atomless. Since this indeterminacy has no economic content, hereafter I focus on the extension of $\mathbf{x}^j$ to $[\underline{\theta}, \overline{\theta}]$ by continuity, whenever possible.

[19] In the two-state case, Lemma 8 in Appendix A implies that C's self-2 will always choose (a_d, p_d) from the I-menu and I's self-2 will always choose $(a_w, \hat{p}_w)$ from the C-menu, provided that ϕ^I is large enough.

The bottom line is that the cause of i's (dis)advantage is—and can only be—the different behavior of i and of j given the j-menu. This also explains why it is essential that $\mathbf{x}^j$ features some flexibility. Furthermore, the two-state example shows that using payments to achieve flexible commitment plays an important role. As noted, to commit I to the flexible plan (a_w, a_d), the I-menu must provide his self-2 with the right prizes and penalties—the 'premium' $p_w - p_d$—to carry it out. However, I's self-1 would rather reap the prizes and avoid the penalties. Since C manages to do so, he has an advantage.[20]

Finally, to further understand Proposition 1, it is useful to think about what would change if agents were heterogeneous, but all time consistent. Specifically, consider a model that is identical to mine except that the agents' utility function is $u_2(a; s, \phi) - p$ in *both* periods, and suppose that $\underline{a} > 0$.[21] Call I_c the *time-consistent* agent with $\phi = \phi^I$, and C the one with $\phi = \phi^C$. As did I before, I_c expects to have a systematically higher valuation than does C. But now I_c has an advantage over C, because I_c enjoys a more already in period 1. Moreover, I_c's advantage is at work even if C's menu is a singleton.

Using Proposition 1, we can now express a condition that, together with monotonicity (MON^i), fully describes the set of feasible allocations. To do so, given non-decreasing $\mathbf{x}^C$ and $\mathbf{x}^I$, construct payment rules $\mathbf{t}^C$ and $\mathbf{t}^I$ using ENV^j. Combining the resulting *R-IC* constraints, we get that a DM can implement $\mathbf{x}^C$ and $\mathbf{x}^I$ if and only if

$$R^C(\mathbf{x}^I) \leq -R^I(\mathbf{x}^C). \tag{RR}$$

In words, $\mathbf{x}^C$ and $\mathbf{x}^I$ are feasible if and only if C's advantage under the I-menu does not exceed I's disadvantage under the C-menu.

Not all non-decreasing $\mathbf{x}^C$ and $\mathbf{x}^I$ meet this condition. For instance, if $\mathbf{x}^C$ is constant and $\mathbf{x}^I$ jumps at $\theta \in (\underline{\theta}, \overline{\theta})$, then $R^C(\mathbf{x}^I) > 0 = R^I(\mathbf{x}^C)$. More generally, the RR condition can fail when both allocation rules sustain a flexible outcome—even the same one. To see this, recall our SD example with two states, and suppose that C always chooses (a_d, p_d) from the I-menu, whereas I always chooses $(a_w, \hat{p}_w)$ from the C-menu. On the one hand, since the I-menu is flexible, the C-menu must grant type C a discount on its options—to avoid C mimicking I—by lowering the payments $\hat{p}_w$ and $\hat{p}_d$ (see EU and $R\text{-}IC^C$). On the other hand, I may worry only a little about choosing $(a_w, \hat{p}_w)$ in state s_d, when C chooses $(a_d, \hat{p}_d)$ instead; for instance, s_d may be very unlikely. So, if the discount on the C-menu is large enough, then I may overall prefer the C- to the I-menu, even though I dislikes the C-menu for being a weak commitment device.

Building on this intuition, the next lemma shows that also the efficient pair $(\mathbf{x}^{C*}, \mathbf{x}^{I*})$ may violate the RR condition and thus be infeasible.

Lemma 2 *Let $\mathbf{x}^{C*}$ and $\mathbf{x}^{I*}$ be the allocation rules associated to the efficient outcome $\mathbf{a}^*$. There is a family of distributions F, such that $(\mathbf{x}^{C*}, \mathbf{x}^{I*})$ is infeasible since $R^C(\mathbf{x}^{I*}) > -R^I(\mathbf{x}^{C*})$.*

Under $\mathbf{x}^{C*}$, type I and type C behave similarly in states close to $\overline{s}$. Thus, if such states are likely enough according to F, then in period 1 type I values the C-menu almost as much as does type C. And since $R^C(\mathbf{x}^{I*}) > 0$, the discount on the C-menu lures I to mimic C. For instance, one can show that, in our initial example about SDs, $(\mathbf{x}^{C*}, \mathbf{x}^{I*})$ is infeasible if F is uniform.

[20] A further discussion of this lest point appears in Section 6.2, in relation to Amador, Werning, and Angeletos' (2006) model which rules out payments across states.

[21] A similar model appears in Courty and Li (2000).

The second part of Proposition 1 hints at the possibility of relaxing I's incentives to lie about ϕ—and hence the RR condition—by adding to the C-menu items that type C never uses. The next lemma shows how to do so in the most effective way; an intuitive explanation follows.[22]

Lemma 3 (Usefulness of Unused Items) *Fix a non-decreasing outcome $\boldsymbol{\chi}^C$ with $\boldsymbol{\chi}^C(\overline{\theta}^C) < \overline{a}$. Let the allocation rule $\mathbf{x}_{\mathrm{u}}^C$ satisfy $\mathbf{x}_{\mathrm{u}}^C\big|_{\Theta^C} = \boldsymbol{\chi}^C$ and*

$$\mathbf{x}_{\mathrm{u}}^C(\theta) = \begin{cases} \boldsymbol{\chi}^C(\overline{\theta}^C) & \text{if } \overline{\theta}^C < \theta \leq \theta_{\mathrm{u}} \\ \overline{a} & \text{if } \theta > \theta_{\mathrm{u}} \end{cases}.$$

Then, there exists $\theta_{\mathrm{u}} < \overline{\theta}^I$ such that $\mathbf{x}_{\mathrm{u}}^C$ minimizes $R^I(\mathbf{x}^C)$. Furthermore, if $\hat{\mathbf{x}}^C$ is non-decreasing, it sustains $\boldsymbol{\chi}^C$, and it minimizes $R^I(\mathbf{x}^C)$, then $\hat{\mathbf{x}}^C(\theta) = \mathbf{x}_{\mathrm{u}}^C(\theta)$ for $\theta \in (\overline{\theta}^C, \theta_{\mathrm{d}}) \cup (\theta_{\mathrm{u}}, \overline{\theta}^I)$, where $\theta_{\mathrm{d}} \leq \theta_{\mathrm{u}}$ and depend only on ϕ^I and F.

To relax the R-IC^I constraint, the principal adds to the C-menu one item that type C never chooses, but deters I from mimicking C. To see why, consider our SD example with two states, in which the C-menu had only two savings options (a_w and a_d). Assume that, given such a menu, I would always withdraw a_w. Now, suppose we add to the C-menu the option to withdraw more than a_w, say $\overline{a}$. We can design payments (the fees) so that C still chooses a_w and a_d as before, yet I, who has less self-control, would withdraw $\overline{a}$ in some state. So, the new C-menu makes I and C behave differently not only in state s_d, but also in state s_w; thus, it makes lying in period 1 even less attractive for I.

However, to make lying the least attractive for type I, the C-menu need not induce I to give in to temptation whenever his self-2 valuation exceeds $\overline{\theta}^C$—for instance, if F is uniform, one can show that $\theta_{\mathrm{u}} = \theta_{\mathrm{d}} > \overline{\theta}^C$. Intuitively, the fee charged for withdrawing $\overline{a}$—our 'tempting' option—controls both the probability and the disappointment that I assigns ex ante to paying it when C doesn't. A low fee—tailored to a θ close to $\overline{\theta}^C$—makes I expect that he would withdraw $\overline{a}$ with high probability, but with little disappointment. Instead, a high fee—tailored to a θ close to $\overline{\theta}^I$—makes I think that he would withdraw $\overline{a}$ with low probability, but with much disappointment. Clearly, depending on the distribution F, charging the high fee for $\overline{a}$ may deter mimicking more.[23]

Although enlarging the C-menu as shown in Lemma 3 maximally relaxes the R-IC^I constraint, it may still fail to make a pair of outcomes implementable. Indeed, meeting the RR condition requires that the unused action $\overline{a}$ be large enough.

Corollary 1 *Suppose that the non-decreasing outcomes $\boldsymbol{\chi}^C$ and $\boldsymbol{\chi}^I$ are not implementable without exploiting unused items. Then, enlarging the C-menu as in Lemma 3 allows to implement them if and only if $\overline{a} - \boldsymbol{\chi}^C(\overline{\theta}^C) \geq D(\boldsymbol{\chi}^C, \boldsymbol{\chi}^I) > 0$.*

Thus, if the principal can freely enlarge the C-menu, asymmetric information about time inconsistency adds *no* restrictions on the set of outcomes that she can sustain. Moreover, offering type C unused items may be *necessary and sufficient* to sustain the efficient outcome with both types.

[22] I discuss how my results relate to Esteban and Miyagawa's (2005) use of "decorated" menus in Section 6.2.

[23] It is easy to see that enlarging the I-menu with unused items never helps screening. Such items would create more opportunities for C to behave differently than does I, given the I-menu, thus strengthening C's advantage.

4.2 Optimal Screening of Time Inconsistency

4.2.1 Trade-offs in Designing the Screening Devices

With asymmetric information, the advantage of time-consistent agents complicates the task of supplying flexible commitment devices. My goal here is to describe the trade-offs that the principal faces when designing the C- and I-menus to optimally screen the agents. I first provide an intuitive argument based on the insights from the previous section; the formal argument follows in Lemma 4.

With regard to the I-menu, in general, the principal faces a trade-off between maximizing the surplus with type I and extracting rents from type C. If the principal offers a flexible I-menu that sustains the efficient outcome with type I, then she maximizes not only the surplus with I's self-1, but also the profitability of the menu. But, by Proposition 1, a flexible I-menu triggers the advantage of type C, thereby boosting his information rents. Since such rents make any C-menu less profitable, the principal faces a trade-off. Similarly, in Mussa and Rosen (1978), serving the low-valuation buyers efficiently requires granting a high-valuation buyer rents that make trading with him less profitable. In the present model, however, the principal also has to worry about the rents that her I-menu creates for type C because they may jeopardize I's incentives to choose the I-menu (recall the RR condition).

With regard to the C-menu, the principal wants to maximize the surplus with type C, but again she has to worry about how her C-menu affects I's incentives to reveal his time-inconsistency. As for the I-menu, sustaining the efficient outcome with type C is necessary to maximize the surplus with him and hence the profitability of the C-menu. However, although for any flexible C-menu I's expected payoff is strictly lower than C's (Proposition 1), this gap may not be enough to deter I from mimicking C (again, recall RR and Lemma 2).

To turn these intuitive considerations into a formal argument, we first need to describe the problem of finding the optimal C- and I-menus under the restrictions of Section 4.1. To do so, recall that the ex-ante social surplus of action a in state $s = \theta/\phi^i$ is $u_1(a;\theta/\phi^i) - c(a)$ (Definition 1). So, let the ex-ante social surplus that type i expects from choosing the i-menu be

$$W^i(\mathbf{x}^i) := \int_{\underline{\theta}}^{\overline{\theta}} [u_1(\mathbf{x}^i(\theta);\theta/\phi^i) - c(\mathbf{x}^i(\theta))]dF^i(\theta). \qquad (ES)$$

Then, if a DM $\{\mathbf{X}, \mathbf{k}\}$ ensures participation and truthful reports of ϕ, its expected ex-ante social surplus is

$$W(\mathbf{X}) = \gamma W^C(\mathbf{x}^C) + (1-\gamma)W^I(\mathbf{x}^I),$$

and its expected profits are

$$\begin{aligned}\Pi(\mathbf{X},\mathbf{k}) &= \gamma \int_{\underline{\theta}}^{\overline{\theta}} \left[\mathbf{t}^C(\theta) - c(\mathbf{x}^C(\theta))\right] dF^C(\theta) + (1-\gamma)\int_{\underline{\theta}}^{\overline{\theta}} \left[\mathbf{t}^I(\theta) - c(\mathbf{x}^I(\theta))\right] dF^I(\theta) \\ &= \gamma\left[W^C(\mathbf{x}^C) - U^C(\mathbf{x}^C, k^C)\right] + (1-\gamma)\left[W^I(\mathbf{x}^I) - U^I(\mathbf{x}^I, k^I)\right].\end{aligned}$$

Therefore, we can write the principal's problem as

$$\mathcal{P} := \begin{cases} \max_{\{\mathbf{X},\mathbf{k}\}}(1-\pi)W(\mathbf{X}) + \pi\Pi(\mathbf{X},\mathbf{k}) \\ \text{s.t. } \mathbf{x}^i \in \mathcal{M},\ IR^i, \text{ and } R\text{-}IC^i \text{ for } i = C, I. \end{cases}$$

We can simplify $\mathcal{P}$ with a few standard steps. By Proposition 1, IR^I and $R\text{-}IC^C$ imply IR^C. Now, recall that $U^i(x^i, k^i)$ is decreasing in k^i (see *EU*). So, if $\pi > 0$, $R\text{-}IC^C$ must bind, and $R\text{-}IC^I$ becomes $R^I(\mathbf{x}^C) + R^C(\mathbf{x}^I) \leq 0$, which is the *RR* condition. Similarly, if $\pi > 0$, IR^I must also bind. If $\pi = 0$, we can safely focus on DMs that make $R\text{-}IC^C$ and IR^I hold with equality, since we are ultimately interested in the optimal $\mathbf{X}$. Given $\mathbf{X}$, we can then uniquely pin down $\mathbf{k}$ using the conditions $U^I(\mathbf{x}^I, k^I) = 0$ and $U^C(\mathbf{x}^C, k^C) = R^C(\mathbf{x}^I)$. Thus, the problem $\mathcal{P}$ reduces to

$$\mathcal{P}' = \begin{cases} \max_{\mathbf{x}} \gamma W^C(\mathbf{x}^C) + (1-\gamma)\left[W^I(\mathbf{x}^I) - \frac{\pi\gamma}{1-\gamma} R^C(\mathbf{x}^I)\right] \\ \text{s.t. } \mathbf{x}^C, \mathbf{x}^I \in \mathcal{M} \text{ and } RR. \end{cases}$$

The next lemma gives necessary and sufficient conditions for $\mathbf{x}^C$ and $\mathbf{x}^I$ to solve $\mathcal{P}'$. It relies on Lagrangian techniques that don't require assuming any property about $\mathbf{x}^C$ and $\mathbf{x}^I$ beyond the necessary monotonicity condition. (For simplicity, for $i = I$ let $-i = C$, and vice versa.)

Lemma 4 (Optimality) *The allocation rules $\mathbf{x}^C$ and $\mathbf{x}^I$ solve the principal's problem $\mathcal{P}'$ if and only if, for some real number $\mu \geq 0$, $(\mathbf{x}^C, \mathbf{x}^I, \mu)$ satisfies*

$$\mathbf{x}^i \in \arg\max_{\mathbf{x}\in\mathcal{M}} W^i(\mathbf{x}) - \rho^{-i} R^{-i}(\mathbf{x}) \;\; \textit{for } i = C, I,$$

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

where $\rho^C := \frac{\pi\gamma+\mu}{1-\gamma}$ and $\rho^I := \frac{\mu}{\gamma}$.

The first condition in Lemma 4 formally captures the trade-offs that I intuitively explained before. With regard to the I-menu, the principal maximizes the surplus with type I net of the resulting rents of type C, weighted by ρ^C. Such a weight depends on three things: (1) how much the principal cares about profits, i.e., π; (2) how much type C matters for her profits, i.e., γ; and (3) whether C's rents interfere with I's incentives to choose the I-menu, i.e., μ. Note that a positive weight on profits ($\pi > 0$) always implies a positive weight ρ^C on C's rents. However, ρ^C may be positive even for $\pi = 0$; that is, even for a planner with no resource constraint. Finally, with regard to the C-menu, again the principal maximizes the surplus with type C, but she may also take into account—through ρ^I—the extent to which the C-menu deters I from mimicking C, namely, $R^I(\mathbf{x}^C)$.

The form of the maximizations in Lemma 4 suggests deriving each $\mathbf{x}^i$ and its properties directly as a function of ρ^{-i}. For this reason, it is worth observing at this point that the weight ρ^C is increasing in π, γ, and μ. I consider $\mathbf{x}^I$ first.

4.2.2 The Optimal Device for Time-Inconsistent Agents

I now characterize how the principal distorts the I-menu to limit the rents to type C. These distortions always involve curtailing the flexibility of the I-menu as described in the initial example about SDs.

To gain some intuition for why the trade-off between surplus with type I and rents to type C leads to distorting the flexibility of the I-menu, consider first an analogy with Mussa and Rosen (1978). In their model, to curb the rents to high-valuation buyers, the monopolist distorts the quality for low-valuation buyers below the efficient level. This is because lowering such a quality curbs the advantage of the high-valuation buyers. Similarly, in the present model, the principal should distort the feature of the I-menu that causes C's advantage, namely, its flexibility. To further help intuition, consider now our SD example with two states. Recall that C gains when he mimics I, because C avoids the premium $(p_w - p_d)$ that commits I to not withdrawing a_w in state s_d. Reducing such a premium should then lessen C's advantage. Also, suppose that I's incentive to choose a_w over a_d increases in how much I can withdraw. Then, allowing I to withdraw less than a_w, say a', does the job: Deterring I from withdrawing a' in state s_d requires a smaller premium; and C gains less from being able to avoid it. Finally, if efficiency calls for I to withdraw a_w in s_w and to deposit a_d in s_d, then having only the options a_d and a' amounts to inefficiently low flexibility.

The bottom line is then this: Reducing the gap between I's and C's choices from the I-menu curbs C's advantage, *through* the induced changes in the prizes and penalties ($\mathbf{t}$) that help I achieve commitment. If we start from the efficient flexible I-menu, this strategy requires limiting I's ability to respond to future information. The optimal way to do so, however, is less immediate than lowering quality as in Mussa and Rosen. We may curtail I's flexibility for intermediate states, but leave it for extreme states, where it seems more valuable. Of course, another possibility is to do the opposite. To see which strategy is better, we need to delve more deeply into the model.

The first step is to write the objective $W^I(\mathbf{x}^I) - \rho^C R^C(\mathbf{x}^I)$, derived in Lemma 4, as an expected virtual surplus. Recall that i's expected utility $U^i(\mathbf{x}^I, k^I)$ depends on $\mathbf{x}^I$ through the function $v^i(\cdot)$, as shown in expression *EU*.

Lemma 5 (Virtual Surplus with Type I) *The objective $W^I(\mathbf{x}^I) - \rho^C R^C(\mathbf{x}^I)$ equals the expected virtual surplus*

$$VS^I(\mathbf{x}^I;\rho^C) := \rho^C \int_{\underline{\theta}^C}^{\underline{\theta}^I} \mathbf{x}^I(\theta) F^C(\theta)\, d\theta + \int_{\underline{\theta}^I}^{\overline{\theta}^I} \left[\mathbf{x}^I(\theta) w^I(\theta;\rho^C) + d(\mathbf{x}^I(\theta)) - c(\mathbf{x}^I(\theta))\right] dF^I(\theta).$$

For $\theta \in \Theta^I$, $w^I(\theta;\rho^C) := \theta/\phi^I + \rho^C V^I(\theta)$ is the virtual valuation *of $\mathbf{x}^I(\theta)$, where*

$$V^I(\theta) := v^I(\theta) - \frac{f^C(\theta)}{f^I(\theta)} v^C(\theta).$$

We can interpret VS^I as follows. The first term measures how the I-menu affects C's rents, through what C expects to choose when $\theta < \underline{\theta}^I$. Intuitively, if $\mathbf{x}^I(\theta)$ increases by one unit, its payment rises by θ (see ENV^I). But the direct benefit to C's self-1 also rises by θ, offsetting the local change in $\mathbf{t}^I$. However, the payments $\mathbf{t}^I$ also increase by one unit for all $\theta' < \theta$ (see ENV^I). This global effect reduces R^C, and its magnitude depends on the mass $F^C(\theta)$ of $\theta' < \theta$. The second term of VS^I is the virtual surplus generated by I's choices, which correspond to self-2 valuations in Θ^I. The virtual valuation $w^I(\theta;\rho^C)$ and the self-1 valuation θ/ϕ^I differ by the wedge $\rho^C V^I(\theta)$, which accounts for the effects of $\mathbf{x}^I(\theta)$ on C's rents. Intuitively, $\mathbf{x}^I(\theta)$ has an indirect effect—captured by $v^I(\theta)$—if $\mathbf{x}^I(\theta)$ implies that, in state $s = \theta/\phi^I$, C and I choose differently from the I-menu. But, for $\theta \in \Theta^C \cap \Theta^I$, $\mathbf{x}^I(\theta)$ has also a direct effect—captured by $-\frac{f^C(\theta)}{f^I(\theta)} v^C(\theta)$—because for such θs, $\mathbf{x}^I(\theta)$ enters directly in C's expected payoff from the I-menu.

I now show that a non-decreasing maximizer of VS^I exists and is unique; I also describe its behavior as the weight ρ^C gets arbitrarily small or large.[24] Recall that $\mathbf{x}^{I*}$ corresponds to an

[24] For instance, if $\pi > 0$, this happens as the share γ of type C goes to zero or to one, provided that the constraint RR doesn't bind ($\mu = 0$).

I-menu that sustains the efficient outcome $\mathbf{a}^*$. Also, let

$$a^{\mathrm{nf}} := \arg\max_{a \in A} aE(s) + d(a) - c(a)$$

be the ex-ante efficient action if the agent is not allowed to respond to future information at all—'nf' stands for 'no flexibility.'

Proposition 2 (Existence and Uniqueness) *For every $\rho^C > 0$, there exists an allocation rule $\mathbf{x}^I(\rho^C)$ that is non-decreasing in θ and maximizes VS^I. Furthermore, $\mathbf{x}^I(\rho^C)$ is unique, continuous in θ and ρ^C, and satisfies*

$$\lim_{\rho^C \to +\infty} \max_{\Theta} \left| \mathbf{x}^I(\theta;\rho^C) - a^{\mathrm{nf}} \right| = 0 \quad \text{and} \quad \lim_{\rho^C \to 0} \mathbf{x}^I(\theta;\rho^C) = \mathbf{x}^{I*}(\theta) \text{ for } \theta \in \Theta.$$

The proof builds on Toikka's (2011) generalization of Myerson's (1981) ironing technique, and explicitly constructs the optimal extension of $\mathbf{x}^I$ off path. Uniqueness follows from strict concavity of the function d. The first limit says that, as the principal becomes extremely concerned about C's rents, she tends to disregard I's desire for flexibility, in the limit offering I a menu with only a^{nf}—a radical reduction of flexibility vis-à-vis the first best. For example, whenever the principal cares about profits, she tends to induce such an outcome if time-consistent agents represent almost the entire market (if $\pi > 0$, $\rho^C \to +\infty$ as $\gamma \to 1$).

The next proposition contains the main characterization result of the paper. It confirms that, for positive ρ^C, the principal inefficiently curtails the flexibility of the I-menu; it also shows the key features of how she does so. To prove Property (c), I assume that the distribution F satisfies the following minor regularity property.

Assumption 2 *The inverse hazard rate $H(s) = \frac{1-F(s)}{f(s)}$ is non-increasing over $\left[s^\dagger, \overline{s}\right]$ for some $s^\dagger < \overline{s}$.*

Assumption 2 is meant to rule out the possibility that H keeps oscillating on its way to zero as s converges to $\overline{s}$. It is implied by the standard regularity assumption of Myerson (1981). It is also satisfied by any distribution with continuously differentiable, positive density.

Proposition 3 (Distortions of the I-menu) *For every $\rho^C > 0$, the optimal I-menu has the following properties:*

(a) Range Reduction with 'Overconsumption' at the Bottom and 'Underconsumption' at the Top: $\mathbf{x}^I(\theta;\rho^C) > \mathbf{x}^{I*}(\theta)$ *over* $[\underline{\theta}^I, \theta_*) \neq \varnothing$, *and* $\mathbf{x}^I(\theta;\rho^C) < \mathbf{x}^{I*}(\theta)$ *over* $(\theta^*, \overline{\theta}^I] \neq \varnothing$;

(b) No Flexibility at the Bottom: $\mathbf{x}^I(\rho^C)$ *is constant over* $[\underline{\theta}^C, \theta_{\mathrm{b}}]$ *with* $\theta_{\mathrm{b}} > \underline{\theta}^I$. *Furthermore, if* $\theta_{\mathrm{b}} < \overline{\theta}^I$, *it must satisfy*

$$\int_{\underline{\theta}^I}^{\theta_{\mathrm{b}}} [w^I(\theta_{\mathrm{b}};\rho^C) - \theta/\phi^I] dF^I(\theta) = \rho^C \left[\int_{\underline{\theta}^C}^{\underline{\theta}^I} F^C(\theta)\, d\theta + \int_{\underline{\theta}^I}^{\theta_{\mathrm{b}}} V^I(\theta)\, dF^I(\theta) \right]; \qquad (NFB)$$

(c) No Flexibility at the Top: *There exists a bounded $\overline{\rho}^C$ such that, if $\rho^C > \overline{\rho}^C$, then $\mathbf{x}^I(\rho^C)$ is constant over $[\theta^{\mathrm{b}}, \overline{\theta}^I]$ with $\theta^{\mathrm{b}} < \overline{\theta}^I$.*

By Property (a), the screening I-menu always restricts I's choices to a strict subset of the efficient set of actions, which corresponds to the range of $\mathbf{x}^{I*}$. The I-menu may, however, let type I act on his period-2 information: $\mathbf{x}^I$ need not be constant at a^{nf}. By continuity of $\mathbf{x}^I(\rho^C)$, this flexibility extends over a range of states with a rich set of options. The I-menu features some flexibility, for example, when the share γ of time-consistent agents is small and the largest feasible action $\overline{a}$ is big (see Propositions 2 and Corollary 1).

The principal makes type I 'overconsume' over small states and 'underconsume' over large states for the following reason. Recall that to curb C's rents, she has to shrink the gap between what C and I choose from the I-menu in each state, as doing so lowers the payment premia separating their choices. Therefore, the key is understanding how distorting $\mathbf{x}^I$ affects $\mathbf{t}^I$. To do so, fix $\theta \in \Theta^I$ and recall ENV^I. On the one hand, raising $\mathbf{x}^I(\theta)$ has a positive local effect on $\mathbf{t}^I(\theta)$ and a negative global effect on $\mathbf{t}^I(\theta')$ for every θ' above θ; so, the premia between the choices at θ and θ' shrink for every $\theta' > \theta$, but rise for every $\theta' < \theta$. On the other hand, reducing $\mathbf{x}^I(\theta)$ has the opposite effects. Thus, for θ close to $\underline{\theta}^I$, curbing C's rents involves raising $\mathbf{x}^I(\theta)$ because the mass of $\theta' > \theta$ prevails over that of $\theta' < \theta$. Instead, for θ close to $\overline{\theta}^I$, curbing C's rents involves reducing $\mathbf{x}^I(\theta)$ because the mass of $\theta' < \theta$ prevails over that of $\theta' > \theta$. Of course, the principal weighs the beneficial effects on R^C of distorting $\mathbf{x}^I$ against its welfare costs—through the virtual valuation w^I. Nevertheless, the direction of the distortions does not change.

By Property (b), the screening I-menu always makes type I not respond to information over a range of small states. This bunching yields an efficiency loss—captured by the left hand side of *NFB*—because a self-1 with valuation θ/ϕ^I is forced to choose the allocation assigned to the higher valuation $w^I(\theta_{\text{b}}; \rho^C)$. However, bunching also yields a benefit—captured by the right hand side of *NFB*—as it curtails C's rents for the same reason as before. To see why, fix an I-menu and identify the lowest action that I chooses, i.e., $\mathbf{x}^I(\underline{\theta}^I)$. If there are items with $a < \mathbf{x}^I(\underline{\theta}^I)$, then tossing them doesn't change the outcome with I, but curbs C's rents: In the states below $\underline{\theta}^I$, a dishonest C would now choose the larger action $\mathbf{x}^I(\underline{\theta}^I)$, which is closer to what I chooses. This argument applies *a fortiori* if, over those states, C would pick an even larger action, say $a_{\text{b}} > \mathbf{x}^I(\underline{\theta}^I)$. This extra step, however, requires that also I chooses a larger action for some θs above $\underline{\theta}^I$. The smaller is the set of θs affected by this change, the lower is the loss in I's welfare. So, the best strategy is to bunch at a_{b} every θ that was choosing a smaller action—formally, every $\theta < \theta_{\text{b}}$—and not to change the allocation for all other θs. Finally, for every threshold $\theta_{\text{b}} > \underline{\theta}^I$, a dishonest C would behave more similarly to an honest I for the states in $[\underline{s}, \theta_{\text{b}}]$, whereas I must choose a larger action only for the states in $[\underline{s}, \theta_{\text{b}}/\phi^I]$. Therefore, for θ_{b} close enough to $\underline{\theta}^I = \phi^I \underline{s}$, I's welfare falls less than do C's rents, and some bunching is always optimal.

Finally, by Property (c), the screening I-menu also makes type I not respond to information over a range of large states, provided that the principal cares enough about C's rents. By Property (a), the principal curbs C's rents by inefficiently reducing $\mathbf{x}^I(\theta)$ for large θs. Also, recall that the global effect on $\mathbf{t}^I$ of changing $\mathbf{x}^I$ limits these reductions, but loses power as θ tends to $\overline{\theta}^I$. So, intuitively, the principal wants to reduce $\mathbf{x}^I(\theta)$ more, the higher is θ. Now, recall that what justifies distorting $\mathbf{x}^I$ is the weight ρ^C on C's rents. So, intuitively, a higher ρ^C justifies larger distortions: in particular, reductions in $\mathbf{x}^I(\theta)$ that grow faster as θ tends to $\overline{\theta}^I$. By this mechanism, the principal may wish that a higher θ chose a *smaller* action than does a lower θ. Since this is impossible (by MON^I), at most the principal can pool top self-2 valuations of type I at the same action. However, the principal cares also about efficiency with type I, which requires that $\mathbf{x}^I$ be strictly increasing. Which interest prevails ultimately depends on the virtual valuation w^I. As I show in Appendix B, for ρ^C large enough, w^I is *decreasing* at the top of Θ^I, which causes bunching.

To achieve a more precise comparison of the screening I-menu (i.e., $\mathbf{x}^I(\rho^C)$) with the efficient one (i.e., $\mathbf{x}^{I*}$), it is necessary to introduce further assumptions about the distribution F; this is because of the complexity of the virtual valuation w^I. Nonetheless, Appendix B offers a complete derivation of $\mathbf{x}^I(\rho^C)$. For illustrative purposes, I consider here the case with uniform F. In this case, a simple relationship also emerges between the weight ρ^C and the set of actions in the I-menu, as well as the range of states involving flexibility (i.e., $[\theta_{\mathrm{b}}, \theta^{\mathrm{b}}]$).

Lemma 6 *Suppose that s is uniformly distributed and $\phi^I < \overline{s}/\underline{s}$. Then, the allocation rule $\mathbf{x}^I(\rho^C)$ crosses $\mathbf{x}^{I*}$ only once and is increasing over $[\theta_{\mathrm{b}}, \theta^{\mathrm{b}}]$. Furthermore, as ρ^C rises, $\mathbf{x}^I(\rho^C)$ changes as follows: θ_{b} and $\mathbf{x}^I(\theta_{\mathrm{b}}; \rho^C)$ increase, and $\mathbf{x}^I(\theta^{\mathrm{b}}; \rho^C)$ decreases; when $\theta^{\mathrm{b}} < \overline{\theta}^I$, then θ^{b} decreases as well.*

In general, more complicated patterns can occur, including bunching at intermediate points for standard ironing reasons. The main point is that, in the present model, bunching at the bottom and at the top arise for new reasons, which hold for a general class of distributions F.

4.2.3 The Optimal Device for Time-Consistent Agents

Proposition 1 suggests that, in the present model, time-consistent agents are the 'strong' type. In standard screening models, the agents of the 'strongest' type usually achieve an undistorted outcome, as if information were symmetric; this familiar result is known as the 'no distortion at the top' (NDT) property. In Mussa and Rosen (1978), for instance, the buyers with the highest valuation always trade efficiently with the monopolist. This is because, in their setup, the monopolist can always trade efficiently with the 'strongest' buyers without triggering mimicking by weaker buyers. In the present model, instead, sustaining the efficient outcome with C may jeopardize I's incentives to behave honestly (see Section 4.1). The goal here is to study what this implies for the optimal screening C-menu.

As a preliminary step, I show that the screening C-menu sustains the efficient outcome with type C if and only if the RR constraint does not bind.

Lemma 7 *There exists $\mathbf{x}^C \in \mathcal{M}$ that sustains the efficient outcome $\boldsymbol{\chi}^{C*}$ and maximizes $W^C(\mathbf{x}) - \rho^I R^I(\mathbf{x})$ if and only if $\rho^I = 0$.*

In other words, the NDT property holds if and only if sustaining $\boldsymbol{\chi}^{C*}$ with C, while offering I the I-menu of Section 4.2.2, does not induce I to mimic C.

This equilibrium may be infeasible—independently of π—if the C-menu doesn't include unused items, i.e., if $\mathbf{x}^C = \mathbf{x}^{C*}$. To see why, recall that as the principal cares less about C's rents (i.e., $\rho^C \to 0$), she tends to offer an I-menu that resembles the first-best one, i.e., $\mathbf{x}^{I*}$ (Proposition 2). By Lemma 2, however, in some environments the principal can't implement $\mathbf{x}^{C*}$ and $\mathbf{x}^{I*}$. Therefore, by continuity, she can't also sustain $\mathbf{x}^{C*}$ and $\mathbf{x}^I(\rho^C)$ for ρ^C small enough.

Although the principal can relax a binding RR constraint by adding unused items to the efficient C-menu (Lemma 3), in the present model with time inconsistency the NDT property fails. To actually satisfy RR, the unused action must be deterring enough for I not to mimic C: As Corollary 1 shows, its gap with the largest efficient action $\boldsymbol{\chi}^{C*}(\overline{\theta}^C)$ must be at least $D(\boldsymbol{\chi}^{C*}, \boldsymbol{\chi}^I)$. Therefore, if $D(\boldsymbol{\chi}^{C*}, \boldsymbol{\chi}^I)$ is too large, the principal must also distort the outcome with type C, since any enlargement of the efficient C-menu is too weak a deterrent for I. Otherwise, although the C-menu may differ from the symmetric-information one, it still sustains the efficient outcome with C. In our initial example about SDs, Figure 2 corresponds to this second case.

It turns out that the less the principal cares about profits, the more likely she is to include unused items in her optimal C-menu and to violate the NDT property. Intuitively, as the weight π falls, the principal cares less about C's rents. Therefore, she distorts less the I-menu, and has to grant a larger discount on the C-menu to convince C to choose it. But this larger discount also makes I more willing to mimic C. Thus, as π decreases, the principal has to increase the lack of commitment that I associates to the C-menu.

Proposition 4 (C-menu and Profits Weight) *If for profits weight $\hat{\pi}$ the optimal C-menu must include unused items, so does for any $\pi < \hat{\pi}$. Furthermore, if for $\hat{\pi}$ the optimal C-menu violates the 'no distortion at the top' property, so does for any $\pi < \hat{\pi}$.*

In particular, by Proposition 4, if the monopolist ($\pi = 1$) has to add unused items to her C-menu, so does the planner ($\pi < 1$). Similarly, if the monopolist's solution violates the NDT property, so does the planner's.

The largest feasible action $\overline{a}$ can be sufficiently bigger than the largest efficient one $\mathbf{a}^*(\overline{s})$—so that the NDT property holds—for several reasons. The principal may be able to offer very attractive actions, without any technological or legal constraint. More generally, if we view a as the immediate gratification of some underlying action—e.g., shopping with credit cards—the maximal gratification $\overline{a}$ is likely far bigger than the efficient one $\mathbf{a}^*(\overline{s})$, which takes into account the disutility d and the cost c—e.g., repaying the debt and bearing the default risk.

Finally, when unused items alone cannot ensure truthfulness by type I, the allocation rule $\mathbf{x}^C$ must maximize $W^C(\mathbf{x}) - \rho^I R^I(\mathbf{x})$ with $\rho^I > 0$. The weight on R^I calls for designing the C-menu so as to *worsen* I's disadvantage, which depends on the gap between I's and C's choices from the C-menu. To gain some intuition, consider again the SD example with two states. Recall that the three-option C-menu deters I from mimicking C only if it charges a high enough fee for withdrawing the larger amount $\overline{a}$. Now, suppose that I enjoys more withdrawing $\overline{a}$ than a_w, but only by a little. To be able to charge a high fee for $\overline{a}$, the principal must then lower the amount a of the middle and (possibly) of the bottom option below efficiency. When there are many states and self-2's valuations overlap across types, deriving the optimal $\mathbf{x}^C$ requires more work. The principle is, however, the same: make I's and C's choices from the C-menu *more* different, to increase the payment premia separating them.

4.3 Extension: Many Degrees of Time Inconsistency

I now extend the analysis of the two-type model to a model with more realistic variety in self-control problems. Screening any two agents, who have different degrees of time inconsistency, raises the same issues as screening type C and type I. However, with more than two types, deriving the optimal supply of menus is more intricate. Nonetheless, I show that the menu of the least time-inconsistent agents—who represent the 'strongest' type—sustains the same outcome as does the C-menu in the two-type model. The menus of the other agents, instead, may inefficiently curtail flexibility; if so, they again induce 'over-consumption' and no flexibility at the bottom, and 'under-consumption' and possibly no flexibility at the top. Finally, not only the menu of the least time-inconsistent agents may feature unused items, but also the menus of intermediate types.

To consider a population featuring more than two degrees of time inconsistency, I modify the model in Section 2 as follows. Let the set of types be $\{\phi^1, \ldots, \phi^N\}$, with $N > 2$ finite; for convenience, order the types so that a higher index coincides with a stronger degree of time inconsistency, i.e., $1 \leq \phi^1 < \phi^2 < \ldots < \phi^N < +\infty$. Finally, let the share of type i be $\gamma_i > 0$. All the other primitives remain as in Section 2.

As in Section 4.1, we can characterize implementable outcomes using direct revelation mechanisms. A DM is an array $\{\mathbf{X}, \mathbf{T}\} = (\mathbf{x}^i, \mathbf{t}^i)_{i=1}^N$, with $\mathbf{x}^i : \Theta \to A$ and $\mathbf{t}^i : \Theta \to \mathbb{R}$, where $\Theta = co\left(\bigcup_{i=1}^N \Theta^i\right)$. It is still without loss of generality to consider only period-2 incentive compatible DMs (Definition 2), which I still denote by $\{\mathbf{X}, \mathbf{k}\} = (\mathbf{x}^i, k^i)_{i=1}^N$. Also, for any j-menu $(\mathbf{x}^j, k^j)$, the expected utility to type i, $U^i(\mathbf{x}^j, k^j)$, is still given by *EU*. Thus, the period-1 information and participation constraints are

$$R\text{-}IC^{i,j}: \quad U^i(\mathbf{x}^i, k^i) \geq U^j(\mathbf{x}^j, k^j) + R^i(\mathbf{x}^j) \quad \text{and} \quad IR^i: \quad U^i(\mathbf{x}^i, k^i) \geq 0.$$

For convenience, I have already written the standard $IC^{i,j}$ constraints in terms of the functions $R^i(\mathbf{x}^j)$, as in Section 4.1.

With regard to each type's information (dis)advantage, Proposition 1 generalizes as follows. A less time-inconsistent agent enjoys an information advantage over any more time-inconsistent agent; i.e., $R^i(\mathbf{x}^j) \geq 0$ if $i < j$, and $R^i(\mathbf{x}^j) \leq 0$ if $i > j$. This is because Proposition 1 relies only on $1 \leq \phi^C < \phi^I$ (see the proof in Appendix B). Moreover, i's (dis)advantage disappears if and only if the j-menu makes i and j behave (almost) always identically; i.e., $R^i(\mathbf{x}^j) = 0$ if and only if $\mathbf{x}^j$ is constant over $(\min\{\underline{\theta}^i, \underline{\theta}^j\}, \max\{\overline{\theta}^i, \overline{\theta}^j\})$. For θ outside this range, neither i nor j get to choose $\mathbf{x}^j(\theta)$, so $R^i(\mathbf{x}^j)$ cannot depend on it.

With regard to period-1 incentive compatibility, two other insights extend to the general model. First, the information rents granted to a less time-inconsistent agent can jeopardize the incentives of a more time-inconsistent agent to disclose his self-control problems. Second, adding unused items to the menu of a less time-inconsistent agent helps promote truthfulness by a more time-inconsistent agent. Moreover, the most effective way to do so is still given by Lemma 3, except that the thresholds θ_{d} and θ_{u} now depend on ϕ^i, ϕ^j, and F.

Consider now the principal's problem of finding an optimal DM. Similarly to Section 4.2.1, we can express this problem—using the ENV^i conditions, *EU*, and *ES*—as

$$\mathcal{P}^N := \begin{cases} \max_{\{\mathbf{X},\mathbf{k}\}} (1-\pi) W(\mathbf{X}) + \pi \Pi(\mathbf{X}, \mathbf{k}) \\ \text{s.t., } \mathbf{x}^i \in \mathcal{M},\ R\text{-}IC^{i,j}, \text{ and } IR^i, \text{ for all } i, j. \end{cases},$$

where

$$W(\mathbf{X}) = \textstyle\sum_{i=1}^N \gamma_i W^i(\mathbf{x}^i) \quad \text{and} \quad \Pi(\mathbf{X}, \mathbf{k}) = \textstyle\sum_{i=1}^N \gamma_i [W^i(\mathbf{x}^i) - U^i(\mathbf{x}^i, k^i)].$$

With more than two types, the problem $\mathcal{P}^N$ raises new difficulties, which I solve with a novel approach. First, in general, local necessary conditions for period-1 incentive compatibility do not imply global incentive compatibility. A similar issue appears in the literature on dynamic mechanism design with only time-consistent agents, which has developed a specific approach to it (Courty and Li (2000); Pavan, Segal, and Toikka (2011)). This approach involves introducing additional, potentially ad-hoc, restrictions on the primitives so that local conditions ensure global incentive compatibility. I find using this approach for studying a new screening problem unappealing; moreover, finding effective—let alone reasonable—restrictions is particularly hard in my context. Second, in my context, there isn't an ordering of types that allows one to identify, a priori, which constraints will bind. For these two reasons, to solve $\mathcal{P}^N$, I take a different approach that does not introduce new restrictions on the primitives, and deals with period-1 incentive constraints all at once. My approach builds on Lagrangian methods and relies on having a finite number of constraints.

Applying these methods reveals that, when designing each j-menu, the principal trades off the surplus with type j and the rents that the menu causes for (some of) the less time-inconsistent agents (see Appendix B). Such a trade-off generalizes the one highlighted in the two-type model. Furthermore, the principal has to ensure that each j-menu deters any type who is more time inconsistent than j from mimicking j. Consistently with Section 4.2.3, I focus on the case in which unused items suffice to guarantee this property.

Proposition 5 *Suppose $\overline{a}$ is large so that unused items suffice to satisfy R-IC^{i,j} for $i > j$. A solution $\{\mathbf{X}, \mathbf{k}\}$ to $\mathcal{P}^N$ exists with $\mathbf{x}^i$ unique over $(\underline{\theta}^i, \overline{\theta}^i)$, for every i. Furthermore, we have:*

1) The 1-menu always sustains the efficient outcome;

2) If $\pi = 0$, then all menus sustain the efficient outcome, otherwise at least the N-menu sustains a distorted outcome;

3) If the i-menu sustains a distorted outcome, it features properties (a), (b), and possibly (c), as in Proposition 3;

4) For every $i < N$, the i-menu may have to include unused items with $a > \mathbf{x}^i(\overline{\theta}^i)$.

As long as the principal cares about profits ($\pi > 0$), she will distort the menu of the most time-inconsistent agents, and possibly of other intermediate types. Whether intermediate types obtain a distorted menu depends on the exact configuration of the active incentive constraints, which cannot be predicted a priori.

5 Policy Implications for Paternalistic Savings Devices

In this section, I discuss the implications of my analysis for how a paternalistic government should design special SDs, to tackle the tendency of some people to save too little for retirement. I argue that my policy implications are consistent with some broad features of the U.S. regulation of 'tax-shielded' and taxable investment accounts.[25]

[25] Of course, I do not intend this section to be seen as providing an explanation for all the complicated rules governing these types of accounts in the U.S..

People's tendency to undersave is a well-documented phenomenon about retirement income, which has been usually attributed to self-control problems (Diamond (1977), Thaler (1994), Laibson (1998), Bernheim (2002)) and has attracted governments' attention. To help people with self-control problems save adequately, a paternalistic government may want to create special SDs that provide appropriate tax incentives. However, when trying to sustain efficient savings decisions with all individuals, the government faces two natural constraints. First, it can usually allocate limited fiscal resources to finance its policies. Second, casual evidence suggests that people value flexibility, because they are uncertain about their best savings decisions in the future; moreover, they have different tendencies to undersave, because their self-control varies. Since the government cannot observe people's self-control, it has to induce each individual to select voluntarily the SD that best fits his desires for commitment and for flexibility. The question is then how the government should design its SDs, given these two constraints.

My model helps answer this question. To see how, recall our initial example about SDs. Again, each SD allows the agents to contribute to ($a < 0$) and withdraw from ($a > 0$) its balance. Now, however, each choice of a leads to a tax burden or benefit p—think of p as the fiscal consequences of a distribution from or a contribution to an SD. As in our initial example, it seems natural that efficiency calls for distributions ($a > 0$) when spending in period 2 has high enough priority ($s > s_0$), and for contributions ($a < 0$) otherwise. Finally, since the government can allocate limited resources to its SDs, it seems natural that it also takes into account their overall profitability, i.e., $0 < \pi < 1$.

The results of Section 4 have the following implications:

- *The unobservability of people's self-control offers one economic justification for government intervention in the market for SDs.* Indeed, asymmetric information leads profit-maximizing firms ($\pi = 1$) to supply SDs inefficiently, especially to the agents who actually need commitment devices.

- *When designing its SDs, a financially constrained government faces a trade-off between corrective and redistributive taxation; its solution calls for curtailing the liquidity of the special SDs, relative to the standard SDs.* By Proposition 1, special SDs—tailored to time-inconsistent agents—make time-consistent agents less willing to choose the standard SDs tailored to them. To make the time-consistent agents choose the standard SDs, it is necessary to grant a tax discount on them. Doing so drains fiscal resources, *in addition to* any tax incentives for the time-inconsistent agents. Curbing the discount on the standard SDs calls for curtailing the flexibility, or liquidity, of the special SDs.

- *To optimally curtail the liquidity of the special SDs, the government should set limits on the owners' freedom to make contributions (in small states) and distributions (in large states).* In small states, the special SDs should induce inefficiently low contributions by setting a binding cap (Proposition 3, Properties (a) and (b); Proposition 5); in large states, they should induce inefficiently low distributions, possibly using again a binding cap (Proposition 3, Properties (a) and (c); Proposition 5). It is important to remark that, according to the present analysis, the reason for setting limits on distributions from the special SDs is *not* to help the time-inconsistent agents avoid depleting their savings. Indeed, if the government could observe the agents' self-control, its special SDs could and should allow time-inconsistent agents to efficiently respond to all future states (recall Section 3).

- *The government should not take for granted that time-inconsistent agents will never choose the standard SDs tailored to time-consistent agents.* The tax discount on the standard SDs may lure the time-inconsistent agents into choosing them. To avoid this, the standard SD should include the possibility of taking large distributions in period 2, but make it so costly that only time-inconsistent agents would use it (Lemma 3 and Proposition 4). For example, a standard SD could allow unlimited borrowing against its balance, and also charge very high penalties for loans beyond a certain threshold.

To compare my policy implications with existing SDs, I consider the U.S. retirement market. In this market, federal laws regulate the special SDs, called 'tax-shielded' investment accounts (TSAs), as well as the standard SDs, called taxable investment accounts (TAs). As the names suggest, TSAs and TAs have different tax rules. Examples of TSAs include the individual retirement accounts (IRAs) and the 401(k) plans, which are important savings instruments in the U.S. and account for significant tax expenditures in the federal budget. More generally, savings taxation contributes to finance other redistributive goals. Therefore, budget considerations are likely to play a role in how the U.S. Congress regulates the TSAs and the TAs.

Comparing the TAs with the TSAs—in particular with the IRAs and the 401(k) plans—reveals some general features that are consistent with my implications. Both the IRAs and the 401(k) plans set maximal contribution limits; TAs don't. These limits are enforced with dear tax penalties for crossing them, and sometimes they bind: In 2007 (2008), about 59% (49%) of IRA-owners contributed at the limit (Holden et al. (2010b)), and roughly 11% of all 401(k) participants did so in 2004 (Munnell and Sundén (2006)). Also, both the IRAs and the 401(k) plans limit distributions; again, TAs don't. Except for a list of qualified cases (e.g., first-time home purchase for IRAs), any amount withdrawn before the age of $59\frac{1}{2}$ incurs a tax penalty, which seems to actually limit access to these TSAs: According to Holden and Schrass (2008-2010a), the vast majority of IRA withdrawals are retirement related, and only about 5% occurs before the owner turns $59\frac{1}{2}$. Moreover, any IRA-backed loan is *de facto* prohibited, and although for some 401(k) plans loans are allowed, they are capped and subject to quick repayments.

Finally, Amromin (2002, 2003) presents some evidence that is consistent with the principle that curtailing flexibility in the special SDs helps curb their appeal to less time-inconsistent agents. He shows that a share of U.S. savers does not take full advantage of the TSAs' tax benefits, and prefers to invest in the TAs because of the TSAs' liquidity constraints. These savers reveal that they care more about flexibility than about commitment, which may depend, among other things, on them being less time inconsistent.

6 Discussion and Literature Review

6.1 Discussion

Overconfidence. The empirical evidence suggests that people can underestimate their self-control problems, i.e., they can be overconfident (see, e.g., DellaVigna 2009 and the references therein). In my model, an agent is overconfident if in period 1 he believes that his type is $\hat{\phi}$, but learns in period 2 that it is $\phi > \hat{\phi}$. Overconfidence affects the analysis as follows. On the one hand, the characterization of the set of implementable outcomes does not change, because it only depends on the belief that each agent holds in period 1 about his type. On the other hand, the principal's problem of finding the optimal screening menus changes. Now the principal has to design each menu taking into account that both sophisticated and overconfident agents can choose it. Therefore, depending on her goal (i.e., on π), she may exploit or counteract the agents' overconfidence. This is especially relevant to the design of unused options. In the two-type model, for example, the principal may have to enlarge the C-menu to screen the sophisticated type I. However, if in period 1 some agent of type I wrongly believes that his type is C, the principal may also want to enlarge the C-menu to exploit his overconfidence in period 2. Thus, she may face a trade-off between optimally dealing with the overconfident type I and ensuring that the sophisticated type I does not choose the C-menu.[26]

Outside Option with Type-Dependent Values. Since only the principal can provide flexible commitment devices, we can interpret any period-1 outside option as the set of state-contingent choices that the agents would make in period 2 if left to their means. Without loss of generality, we can describe these choices through a menu $\{\mathbf{x}^0, k^0\}$, using the formalism of Section 4.1. For simplicity, consider the two-type model. By Proposition 1, $U^C(\mathbf{x}^0, k^0) \geq U^I(\mathbf{x}^0, k^0)$ with equality if and only if $\mathbf{x}^0$ is constant over Θ. In other words, C and I value the outside option equally—as I have assumed throughout the paper—if and only if the menu $\{\mathbf{x}^0, k^0\}$ is singleton. For example, the outside option may lead to the same default action in every state. Instead, if $\mathbf{x}^0$ is flexible, then C and I value the outside option differently. Similarly, in Mussa and Rosen (1978) a high-valuation buyer would value more an outside option featuring strictly positive consumption than would a low-valuation buyer. When C values $\{\mathbf{x}^0, k^0\}$ more than does I, my analysis changes to the extent that the constraint IR^C may bind, together with or even before the constraint IC^C. Obviously, if $\mathbf{x}^0 = \mathbf{x}^{I*}$, then the principal will never offer a distorted I-menu, because she can never extract from type C more than $U^C(\mathbf{x}^{I*}, k^0) = U^I(\mathbf{x}^{I*}, k^0) + R^C(\mathbf{x}^{I*})$. Similarly, if in Mussa and Rosen the outside option features the efficient level of trade with low-valuation buyers, then the monopolist will never offer them a distorted level of trade. Finally, if IC^C binds before IR^C, then the principal will distort the I-menu as discussed in Section 4.2.2, possibly stopping if IR^C starts to bind too.

Different Period-1 Heterogeneity. In this paper, time-inconsistent agents overweight ($\phi > 1$) the immediate benefits (or costs) of their actions when they have to choose one. If all time-inconsistent agents have $\phi < 1$, the substance of the paper doesn't change. Significant differences appear, instead, if some agents have $\phi < 1$ and others have $\phi > 1$. For example, with two types and $\phi^1 > 1 > \phi^2$, neither may have an information advantage (see Appendix A). However, the approach that I introduced in Section 4.3 should be useful to analyze this case as well.

6.2 Relation to the Previous Literature

As noted in the Introduction, this paper relates to the literature on behavioral contract theory and optimal paternalism. Several papers have already studied how to design incentive schemes to tackle people's self-control problems (e.g., O'Donoghue and Rabin (1999)). The novelty of the

[26] A more detailed analysis of the model with overconfident agents is available upon request.

present paper consists in studying, in a market environment, the role of asymmetric information about people's conflicting desires for commitment and for flexibility. A more detailed comparison of my work to the closest papers in the literature follows.

Amador, Werning, and Angeletos (AWA) (2006). In AWA, a decision-maker faces a consumption-savings problem and lacks self-control. Anticipating this situation, ex ante he desires commitment, but because he still doesn't know how much he will value consumption, he also desires flexibility. AWA focus on commitment strategies that only entail removing options from the decision-maker's future budget set; they purposefully rule out payments across states to eliminate any form of insurance. AWA's main contribution is to identify necessary and sufficient conditions for the optimal (paternalistic) strategy to coincide with minimum-savings plans.

In contrast, I look at markets in which, to commit to a desired plan, agents obtain incentive schemes from a profit- or welfare-maximizing supplier; I also allow for payments across states. AWA's model is more appropriate to study public pension plans which involve no payments, such as Social Security and defined-benefit plans. But to study defined-contribution plans and other forms of voluntary savings that receive tax incentives from the government, my model seems better suited.

Furthermore, AWA essentially assume that people's self-control is observable. However, with unobservable self-control, AWA's analysis does not change because their commitment policies raise no incentive compatibility issue. To see why, suppose AWA allowed for both time-inconsistent and time-consistent agents, type I and type C. AWA's optimal policies involve an effective minimum-savings constraint for I, but not for C. Given these policies, by definition, C doesn't want to mimic I ex ante, although if C did, ex post C would save at least as much as does I in every state and *strictly more* in some states. The flip side of AWA's policies is that I's ex-post choices are (almost) always inefficient, according to his ex-ante preference. By contrast, I show that allowing for payments would permit to fully solve the agent's self-control problems under symmetric information, but causes screening problems under asymmetric information.

DellaVigna and Malmendier (DM) (2004). DM's influential work studies how firms design two-part tariffs when selling to quasi-hyperbolic buyers goods that cause immediate costs (benefits) and deferred benefits (costs). DM's model helps explain why firms set per-usage prices below (above) their marginal costs, and introduce renewal or cancellation fees.

DM's model differs from mine as follows. First, in DM first- and second-period preferences differ for any given agent, but they coincide across agents within each period. Also, at the first-period contracting stage, some agents correctly forecast their preference at the second-period consumption stage, but others are partially naive in the sense of O'Donoghue and Rabin (2001). Second, DM restrict *a priori* the agent to a buy-or-not-buy decision—with no quantity variable—a restriction that precludes studying the trade-off between commitment and flexibility. Finally, DM assume that firms know the agents' degree of time inconsistency and of naivete. This symmetric information assumption is relaxed by Jianye (2011), who takes DM's model and studies whether their predictions about per-usage prices are robust to asymmetric information.

DM also find that firms will provide sophisticated agents with a two-part tariff that perfectly solves their self-control problems. I confirm and extend this result in Section 3. Nonetheless, I cast doubt on DM's optimistic message in two directions. First, the result is not robust to asymmetric information, which leads a monopolist to supply commitment inefficiently. Second, the result relies on the property, which holds also in DM, that self-1 and self-2 rank any two states in the same order. To see this, suppose the utility function of self-1 is $sa + d(a) - p$, but that of self-2 is $\frac{a}{s}+d\left(a\right)-p$—the inconsistency is now about which state makes a more valuable. In this case, to be period-2 incentive compatible, an allocation rule $\mathbf{x}$ must be non-increasing; so, the efficient outcome $\mathbf{a}^*$—which is increasing—cannot be sustained with self-2.

Esteban and Miyagawa (EM) (2005). EM study a standard model of non-linear pricing except that buyers have Gul-Pesendorfer (2001) preferences and the monopolist can offer them non-singleton menus. EM's find that the monopolist may fully extract buyers' surplus using "decorated" menus. EM have two types of buyers (H and L), which have different valuations of the monopolist's good, but also different self-control costs. Buyers dislike exerting self-control and desire commitment. So, intuitively, by adding to the singleton L-menu an item that makes a dishonest H incur high self-control costs, the monopolist makes the L-menu less appealing to H and can extract more surplus from H. In my model, the unused item added to the C-menu works similarly: By making I view the C-menu as offering even less commitment, it increases I's expected cost of mimicking C.

Except this similarity, EM's paper is quite different from mine. In EM, agents value commitment but not flexibility, so no trade-off can arise between the two. In EM, unused items help screen 'strong' from 'weak' buyers; instead, in my model, they help screen 'weak' from 'strong' agents. In EM, maximizing welfare never requires adding unused items to menus, but maximizing profits may; instead, in my model, maximizing welfare may require unused items while maximizing profits may be possible without them.

Eliaz and Spiegler (ES) (2006). ES analyze a model in which agents sign contracts with a monopolist in period 1, to get access to a set of actions in period 2. All agents have the same utility function u in period 1, which changes to v in period 2 with probability q. The monopolist knows q; instead, in period 1, each agent has his own private belief $\hat{q}$ about the chance of switching to v. Crucially, in ES the agents' private information affects what they are willing to pay for a given contract in period 1, but it doesn't affect what the monopolist expects them to choose in period 2. Instead, in my model, period-1 private information also matters in predicting the period-2 outcome of a (non-trivial) device—as in adverse-selection models. Furthermore, in ES, the agents have the same information about the environment in both periods, so they care about commitment (unless $\hat{q}=0$), but not about flexibility.

7 Conclusion

I study how a monopolist and a paternalistic planner supply flexible commitment devices to agents who privately know their preference for commitment and for flexibility. If information were symmetric, both the monopolist and the planner would help each agent fully solve his self-control problems. This is generally not true, however, with asymmetric information: Agents with superior self-control enjoy an information advantage, which causes a screening problem. I derive the optimal screening devices using non-standard techniques for solving dynamic mechanism-design problems. To screen a more time-inconsistent from a less time-inconsistent agent, the monopolist and (possibly) the planner inefficiently curtail the flexibility of the device tailored to the first agent, and include unused items in the device tailored to the second agent. Finally, in my model the familiar 'no distortion at the top' property fails.

My results are relevant to the problem of how governments should design special savings devices with tax incentives that help time-inconsistent people adequately save for retirement.

Although such incentives also involve penalties, the special devices will appeal to time-consistent agents. So, to ensure that time-consistent agents don't choose the special devices, the standard devices should receive a tax discount. Finally, in order to curb this discount, the liquidity of the special devices should be inefficiently curtailed, by setting limits on the ability both to tap and to contribute to their balance. These implications are useful for assessing, for instance, the U.S. regulation of individual retirement accounts and 401(k) plans.

The paper does not look at the effects of competition among many providers of flexible commitment devices. While a detailed study of this case is beyond the scope of the current paper, I expect that even perfectly competitive markets lead to inefficient outcomes, with distortions qualitatively similar to the monopolistic case. Intuitively, competition pushes firms to generate the largest possible surplus and payoff for each type of agent. However, the asymmetric information creates an adverse-selection problem for the firms, who have to screen the different types. Conditional on sustaining the efficient outcome with time-consistent agents, each firm has to balance the efficiency with time-inconsistent agents against the incentives of the time-consistent agents to pretend to be time-inconsistent. This trade-off is similar to that analyzed in the paper, and hence is likely to have similar implications.

8 Appendix A: Irrelevance of Asymmetric Information with Finitely Many States

In this appendix, I show that if the number of states is finite, then asymmetric information at the contracting stage may be immaterial as far as sustaining the efficient outcome is concerned.

To understand the role of the discreteness of the state space S, consider the case with two states, $s_2 > s_1$. If ϕ is observable, the principal sustains $a_2^* = \mathbf{a}^*(s_2) > \mathbf{a}^*(s_1) = a_1^*$ (Lemma 1), with payments $p_1 = \mathbf{p}(s_1)$ and $p_2 = \mathbf{p}(s_2)$ that satisfy the condition

$$u_2(a_2^*; s_2, \phi) - u_2(a_1^*; s_2, \phi) \geq p_2 - p_1 \geq u_2(a_2^*; s_1, \phi) - u_2(a_1^*; s_1, \phi), \tag{1}$$

which follows by combining the *IC* constraints. Since $u_2(a; s, \phi)$ has strictly increasing differences in (a, s), the discreteness of S creates some slack in the period-2 incentive constraints evaluated at (a_2^*, a_1^*). Thus, for any ϕ, condition (1) does not uniquely pin down p_1 and p_2.

Now suppose that type C and type I have about the same self-control, i.e., that ϕ^I is almost equal to ϕ^C. Intuitively, to implement the same outcome with both types, it should be possible to use incentive devices that are sufficiently alike. Furthermore, if the discreteness of S leaves some leeway in the choice of payments, it may be even possible to find *one* device that works for both types. On the other hand, if ϕ^I is very different from ϕ^C, then sustaining the efficient outcome with both type requires different devices. Since $\phi^I > \phi^C$, type I is more tempted to pick a_2^* also in state s_1 than is type C, and the more so, the higher ϕ^I is relative to ϕ^C. Thus, for I to choose a_2^* only in state s_2, a_2^* must be sufficiently more expensive than is a_1^*, and this price premium must increase as ϕ^I rises. Therefore, it must eventually exceeds C's willingness to pay for switching from a_1^* to a_2^* in s_2.

The following lemma formalizes this intuition. Since the result does not depend on having two types with $\phi^I > \phi^C = 1$, consider a finite set Φ of types, which may include both $\phi > 1$ and $\phi < 1$. Let $\overline{\phi} := \max \Phi$ and $\underline{\phi} := \min \Phi$.

Lemma 8 *Suppose S is finite with $s_N > s_{N-1} > \ldots > s_1$. There exists a single commitment device that sustains the efficient outcome $\mathbf{a}^*$ with each $\phi \in \Phi$ if and only if $\overline{\phi}/\underline{\phi} \leq \min_i s_{i+1}/s_i$.*

Proof. With N states the period-2 incentive constraints become

$$u_2(a_i; s_i, \phi) - p_i \geq u_2(a_j; s_i, \phi) - p_j \qquad (IC_{i,j})$$

for all $i, j = 1, \ldots, N$, where $a_i := \mathbf{a}(s_i)$ and $p_i := \mathbf{p}(s_i)$. By standard arguments, it is enough to focus on the adjacent constraints $IC_{i,i-1}$ and $IC_{i-1,i}$. For $i = 2, \ldots, N$, let $\Delta_i := p_i - p_{i-1}$. If $\mathbf{a}(s_i) = \mathbf{a}^*(s_i)$ for all i, then $a^*_N > a^*_{N-1} > \ldots > a^*_1$ (Assumption 1). To implement $\mathbf{a}^*$ with ϕ, we must have, for $i = 2, \ldots, N$,

$$u_2\left(a^*_i; s_i, \phi\right) - u_2(a^*_{i-1}; s_i, \phi) \geq \Delta_i \geq u_2\left(a^*_i; s_{i-1}, \phi\right) - u_2(a^*_{i-1}; s_{i-1}, \phi). \qquad (CIC^*_{i,i-1})$$

Recall that for any s and ϕ, $u_2\left(a'; s, \phi\right) - u_2\left(a; s, \phi\right) = \phi s\left(a' - a\right) + d\left(a'\right) - d\left(a\right)$. Let $s_k/s_{k-1} = \min_i s_i/s_{i-1}$ and suppose $\overline{\phi} s_{k-1} > s_k \underline{\phi}$. Then,

$$u_2(a^*_k; s_{k-1}, \overline{\phi}) - u_2(a^*_{k-1}; s_{k-1}, \overline{\phi}) > u_2(a^*_k; s_k, \underline{\phi}) - u_2(a^*_{k-1}; s_k, \underline{\phi}),$$

and no Δ_k satisfies $CIC^*_{k,k-1}$ for both $\underline{\phi}$ and $\overline{\phi}$. If instead $\underline{\phi} s_i \geq \overline{\phi} s_{i-1}$ for every $i = 2, \ldots, N$, then for each ϕ and i

$$\begin{aligned} u_2\left(a^*_i; s_i, \phi\right) - u_2(a^*_{i-1}; s_i, \phi) &\geq u_2(a^*_i; s_{i-1}, \overline{\phi}) - u_2(a^*_{i-1}; s_{i-1}, \overline{\phi}) \\ &\geq u_2\left(a^*_i; s_{i-1}, \phi\right) - u_2(a^*_{i-1}; s_{i-1}, \phi). \end{aligned}$$

Set $\Delta^*_i := u_2(a^*_i; s_{i-1}, \overline{\phi}) - u_2(a^*_{i-1}; s_{i-1}, \overline{\phi})$ for each i. Then $\{\Delta^*_i\}^N_{i=2}$ satisfies all the $CIC^*_{i,i-1}$ conditions for each ϕ. And the payment rule defined by $p^*_i = p^*_1 + \sum^i_{j=2} \Delta^*_j$—with p^*_1 small enough to ensure participation—sustains $\mathbf{a}^*$ with each ϕ. ■

Therefore, if the heterogeneity across agents—measured by $\overline{\phi}/\underline{\phi}$—is small enough, the principal can sustain the efficient outcome without worrying about period-1 incentive constraints.

The condition in Lemma 8, however, is not necessary for asymmetric information to be irrelevant when sustaining the efficient outcome $\mathbf{a}^*$. Even if $\overline{\phi}/\underline{\phi}$ is large, it may be possible to design different devices such that each device sustains $\mathbf{a}^*$ with one type of agent, and each agent optimally chooses the device designed for his type. To see why, consider a simple example with two types with $\phi^h > \phi^l$, and two states, $s_2 > s_1$. Suppose that $\phi^h > 1 > \phi^l$, $\phi^h s_1 > \phi^l s_2$, but $s_2 > s_1 \phi^h$ and $s_2 \phi^l > s_1$. Consider all payments (p_1, p_2) that satisfy (1) and the IR constraint with equality, i.e.,

$$(1 - f)\, p_2 + f p_1 = (1 - f)\, u_1\left(a^*_2; s_2\right) + f u_1\left(a^*_1; s_1\right) \qquad (2)$$

where $f := F(s_1)$. Finally, choose (p^h_1, p^h_2) so that h's self-1 strictly prefers a^*_2 in state s_2—i.e., $u_1(a^*_2; s_2) - p^h_2 > u_1(a^*_1; s_2) - p^h_1$—and (p^l_1, p^l_2) so that l's self-1 strictly prefers a^*_1 in state s_1—i.e., $u_1(a^*_1; s_1) - p^l_1 > u_1(a^*_2; s_1) - p^l_2$. Then, the device defined by (p^l_1, p^l_2) (respectively (p^h_1, p^h_2)) sustains $\mathbf{a}^*$ and generates zero expected payoffs to the agent if and only if type l (h) chooses it. Moreover, I claim that type l (h) strictly prefers the device with payments (p^l_1, p^l_2) (respectively (p^h_1, p^h_2)). Note that, if the *self-1* of either type had to choose in period 2, under either device he would strictly prefer to behave exactly as does his self-2. Therefore, by choosing the 'wrong' device, either type can only decrease his payoff below zero.

The next lemma provides a necessary condition for asymmetric information to be irrelevant when sustaining $\mathbf{a}^*$. Let $\Phi^u := \Phi \cap [0, 1]$ and $\Phi^o := \Phi \cap [1, +\infty)$. For $k = o, u$, let $\overline{\phi}^k := \max \Phi^k$ and $\underline{\phi}^k := \min \Phi^k$, and set $r^k := \overline{\phi}^k / \underline{\phi}^k$.

Lemma 9 *Suppose that S is finite with $s_N > s_{N-1} > \ldots > s_1$. If $\max\{r^u, r^o\} > \min_i s_{i+1}/s_i$, there is no collection of commitment devices, each tailored to a specific type $\phi \in \Phi$, such that (I) type ϕ chooses 'his' device, (II) each device sustains the efficient outcome $\mathbf{a}^*$ with the corresponding type ϕ, and (III) all types enjoy the same expected payoff.*

Proof. Suppose $\max\{r^u, r^o\} = r^u$; the other case follows similarly. Suppose that devices that satisfy (I)-(III) exist. Let $\{\mathbf{p}^\phi\}_{\phi \in \Phi}$ be the corresponding payment schedules, and normalize each type's expected payoff to zero. Consider $\mathbf{p}^{\underline{\phi}^u}$ and denote it by $\underline{\mathbf{p}}$. I claim that if $\overline{\phi}^u$ selects $\underline{\mathbf{p}}$ rather than $\mathbf{p}^{\overline{\phi}^u}$, he enjoys a positive expected payoff. Given $\underline{\mathbf{p}}$, let $\underline{a}_i(\phi)$ be an optimal choice of $\phi \in \Phi^u$ in s_i (this is well defined by assumption). For $\underline{\phi}^u$, $\underline{a}_i(\underline{\phi}^u) = a_i^*$ for every i. Let $S^- := \{i : s_{i+1}/ s_i < r^u\} \neq \varnothing$. We have: (a) for every i, $\overline{\phi}^u s_i > \underline{\phi}^u s_i$ and hence $\underline{a}_i(\overline{\phi}^u) \geq a_i^*$; (b) for $i \in S^-$, $\overline{\phi}^u s_i > \underline{\phi}^u s_{i+1}$, and hence $\underline{a}_i(\overline{\phi}^u) \geq a_{i+1}^* > a_i^*$. Observations (a) and (b), together with $\phi \leq 1$, imply that

$$\underline{\mathbf{p}}(\underline{a}_i(\overline{\phi}^u)) - \underline{\mathbf{p}}\,(a_i^*) \leq u_2(\underline{a}_i(\overline{\phi}^u); s_i, \overline{\phi}^u) - u_2(a_i^*; s_i, \overline{\phi}^u) \leq u_1(\underline{a}_i(\overline{\phi}^u); s_i) - u_1\,(a_i^*; s_i)\,,$$

where the first inequality is strict for $i \in S^-$. The expected payoff to $\overline{\phi}^u$ from $\underline{\mathbf{p}}$ is then

$$\textstyle\sum_{i=1}^N [u_1(\underline{a}_i(\overline{\phi}^u); s_i) - \underline{\mathbf{p}}(\underline{a}_i(\overline{\phi}^u))] f_i > \sum_{i=1}^N [u_1(a_i^*; s_i) - \underline{\mathbf{p}}(a_i^*)] f_i = 0,$$

where $f_i := F(s_i) - F(s_{i-1})$ for $i = 2, \ldots, N$ and $f_1 := F(s_1)$. ■

Note that, if either $\Phi^u \setminus \{1\} = \varnothing$ or $\Phi^o \setminus \{1\} = \varnothing$, then the sufficient condition in Lemma 8 becomes also necessary for sustaining the efficient outcome without worrying about the agent's private information.

9 Appendix B: Omitted Proofs from Sections 3-4.

All proofs allow for $\phi^C \geq 1$, unless the specific result assumes otherwise.

9.1 Proof of Lemma 1

The planner's problem is

$$\max_{\{\mathbf{a},\mathbf{p}\}} \left\{(1-\pi)\textstyle\int_S [u_1\left(\mathbf{a}\left(s\right); s\right) - c(\mathbf{a}(s))] dF + \pi \textstyle\int_S \left[\mathbf{p}\left(s\right) - c\left(\mathbf{a}\left(s\right)\right)\right] dF\right\} \text{ s.t. } IR \text{ and } IC.$$

If $\pi > 0$, *IR* must bind; if $\pi = 0$, assume w.l.o.g. that *IR* holds with equality. Thus, the problem reduces to

$$\max_{\mathbf{a}} \left\{\textstyle\int_S [u_1\left(\mathbf{a}\left(s\right); s\right) - c(\mathbf{a}(s))] dF\right\} \quad \text{s.t. } IC.$$

Ignoring *IC*, the resulting relaxed problem admits $\mathbf{a} \equiv \mathbf{a}^*$ as its unique solution (up to the set of zero measure $\{\underline{s}, \overline{s}\}$). It remains to show that there exist $\mathbf{p}$ that sustains $\mathbf{a}^*$ with *self-2*. For any ϕ, standard arguments imply that the necessary and sufficient condition for the existence of such a $\mathbf{p}$ is that $\mathbf{a}$ is non-decreasing in s, a property satisfied by $\mathbf{a}^*$.

9.2 Proof of Proposition 1

Using expression EU and changing variables, re-write $U^i(\mathbf{x}^j, k^j)$ as

$$U^i(\mathbf{x}^j, k^j) = \int_{\underline{s}}^{\overline{s}} [s\mathbf{x}^j(s\phi^i) - s\phi^i \mathbf{x}^j(s\phi^i) + \int_{\underline{\theta}}^{s\phi^i} \mathbf{x}^j(y)\, dy] dF - k^j. \tag{3}$$

Now define for each s

$$\begin{aligned} \Delta(s|\,\mathbf{x}^i) &= s(1-\phi^C)\mathbf{x}^i(s\phi^C) + \int_{\underline{\theta}}^{s\phi^C} \mathbf{x}^i(y)\, dy - s(1-\phi^I)\mathbf{x}^i(s\phi^I) - \int_{\underline{\theta}}^{s\phi^I} \mathbf{x}^i(y)\, dy \\ &= s(\phi^C - 1)(\mathbf{x}^i(s\phi^I) - \mathbf{x}^i(s\phi^C)) + \int_{s\phi^C}^{s\phi^I} (\mathbf{x}^i(s\phi^I) - \mathbf{x}^i(y))dy. \end{aligned} \tag{4}$$

Note that $\mathbf{x}^i \in \mathcal{M}$ and $\phi^I > \phi^C \geq 1$ imply that $\Delta(s|\,\mathbf{x}^i) \geq 0$ for every s. Also, if $\mathbf{x}^i(\theta) = a$ for every $\theta \in (\underline{\theta}, \overline{\theta})$, then $\Delta(s|\,\mathbf{x}^i) = 0$ for every $s \in (\underline{s}, \overline{s})$.

Consider first type C. Using (3) and (4), we have

$$R^C(\mathbf{x}^I) = U^C(\mathbf{x}^I, k^I) - U^I(\mathbf{x}^I, k^I) = \int_{\underline{s}}^{\overline{s}} \Delta(s|\,\mathbf{x}^I) dF \geq 0, \tag{5}$$

with equality if $\mathbf{x}^I(\theta) = a$ over $(\underline{\theta}, \overline{\theta})$. Now suppose $\mathbf{x}^I(\theta)$ is not constant over $(\underline{\theta}, \overline{\theta})$. Since $\mathbf{x}^I$ is non-decreasing, there exists $\tilde{\theta} \in (\underline{\theta}, \overline{\theta})$ such that $\mathbf{x}^I(\theta) < \mathbf{x}^I(\theta')$ for every θ, θ' such that $\theta < \tilde{\theta} < \theta'$. Let $\tilde{s}_1 := \tilde{\theta}/\phi^I$ and $\tilde{s}_2 := \tilde{\theta}/\phi^C$, and consider the interval $\mathcal{I} := (\tilde{s}_1, \tilde{s}_2) \cap [\underline{s}, \overline{s}] \neq \varnothing$. For $s \in \mathcal{I}$, $s\phi^C < \tilde{\theta} < s\phi^I$; hence $\mathbf{x}^I(s\phi^C) < \mathbf{x}^I(s\phi^I)$. To prove that $R^C(\mathbf{x}^I) > 0$, it is enough to show that

$$\int_{\mathcal{I}} \left[\int_{s\phi^C}^{s\phi^I} (\mathbf{x}^I(s\phi^I) - \mathbf{x}^I(y))dy \right] dF > 0. \tag{6}$$

For $s \in \mathcal{I}$, the integrand in (6) satisfies

$$\int_{\tilde{\theta}}^{s\phi^I} (\mathbf{x}^I(s\phi^I) - \mathbf{x}^I(y))dy + \int_{s\phi^C}^{\tilde{\theta}} (\mathbf{x}^I(s\phi^I) - \mathbf{x}^I(y))dy \geq \int_{s\phi^C}^{\tilde{\theta}} (\mathbf{x}^I(s\phi^I) - \mathbf{x}^I(y))dy > 0,$$

where the first inequality follows from MON^I and the last from $\mathbf{x}^I(y) < \mathbf{x}^I(s\phi^I)$ for every $y \in (s\phi^C, \tilde{\theta})$. Since $\mathcal{I}$ has positive measure, (6) follows.

Now consider type I. Using again (3) and (4), we have $-R^I(\mathbf{x}^C) = \int_{\underline{s}}^{\overline{s}} \Delta(s|\,\mathbf{x}^C)dF$, and the properties of $R^I(\mathbf{x}^C)$ follow from the same type of arguments used for $R^C(\mathbf{x}^I)$.

9.3 Proof of Lemma 2

Using (4) and splitting the integral at the state s such that $\theta^I = \overline{\theta}^C$, we have

$$-R^I(\mathbf{x}^{C*}) = \int_{\underline{s}}^{\overline{s}\frac{\phi^C}{\phi^I}} \Delta(s|\,\mathbf{x}^{C*})dF + \int_{\overline{s}\frac{\phi^C}{\phi^I}}^{\overline{s}} \Delta(s|\,\mathbf{x}^{C*})dF.$$

For every $s \geq \overline{s}\frac{\phi^C}{\phi^I}$, since $\mathbf{x}^{C*}(s\phi^I) = \mathbf{x}^{C*}(\overline{s}\phi^C)$,

$$\Delta(s|\,\mathbf{x}^{C*}) = \overline{s}\phi^C \mathbf{x}^{C*}(\overline{s}\phi^C) - s\phi^C \mathbf{x}^{C*}(s\phi^C) - \int_{s\phi^C}^{\overline{s}\phi^C} \mathbf{x}^{C*}(y)\, dy + s(\mathbf{x}^{C*}(s\phi^C) - \mathbf{x}^{C*}(\overline{s}\phi^C)),$$

and because $\mathbf{x}^{C*}$ is continuous, $\Delta(s|\,\mathbf{x}^{C*}) \to 0$ as $s \to \overline{s}$. Now consider $R^C(\mathbf{x}^{I*})$. Recall that since $s\phi^I > s\phi^C$, $\Delta(s|\,\mathbf{x}^{I*}) > 0$ for $s > \underline{s}$. For convenience, fix $s_0 := \frac{1}{2}(\overline{s} + \underline{s})$. By continuity,

$\min_{[s_0,\bar{s}]} \Delta(s|\,\mathbf{x}^{I*}) = \kappa > 0$. Choose $s_{\kappa/2} < \bar{s}$ so that $\Delta(s|\,\mathbf{x}^{C*}) \leq \kappa/2$ for $s \in [s_{\kappa/2}, \bar{s}]$. Finally, let $s_1 := \max\{s_0, s_{\kappa/2}\}$. We have:

$$R^C(\mathbf{x}^{I*}) \geq \int_{s_1}^{\bar{s}} \Delta(s|\,\mathbf{x}^{I*})dF \geq \kappa\,(1 - F\,(s_1))\,,$$

$$-R^I(\mathbf{x}^{C*}) \leq \sup_{s<s_1} \Delta(s|\,\mathbf{x}^{C*})F\,(s_1) + \frac{\kappa}{2}\,(1 - F\,(s_1))\,.$$

Therefore, $R^C(\mathbf{x}^{I*}) > -R^I(\mathbf{x}^{C*})$ if $(1 - F\,(s_1))/\,F\,(s_1) > \frac{2}{\kappa}\sup_{s<s_1} \Delta(s|\,\mathbf{x}^{C*})$.

9.4 Proof of Lemma 3

Using EU, we have

$$R^I(\mathbf{x}^C) = \int_{\underline{\theta}}^{\overline{\theta}} \left[\frac{1-\phi^I}{\phi^I}\theta\mathbf{x}^C\,(\theta) + \int_{\underline{\theta}}^{\theta}\mathbf{x}^C\,(y)\,dy\right] dF^I - \int_{\underline{\theta}}^{\overline{\theta}} \left[\frac{1-\phi^C}{\phi^C}\theta\mathbf{x}^C\,(\theta) + \int_{\underline{\theta}}^{\theta}\mathbf{x}^C\,(y)\,dy\right] dF^C.$$

Changing the order of integration and rearranging, we can express $R^I(\mathbf{x}^C)$ as

$$R^I(\mathbf{x}^C) = -\int_{\overline{\theta}^C}^{\overline{\theta}^I}\mathbf{x}^C\,(\theta)\,g^I\,(\theta)\,d\theta + \int_{\underline{\theta}^C}^{\overline{\theta}^C}\mathbf{x}^C\,(\theta)\,V^C\,(\theta)\,dF^C, \tag{7}$$

where the function $g^I : [\overline{\theta}^C, \overline{\theta}^I] \to \mathbb{R}$ is given by

$$g^I\,(\theta) := \frac{\phi^I - 1}{\phi^I}\theta f^I\,(\theta) - (1 - F^I\,(\theta)),$$

and the function $V^C : [\underline{\theta}^C, \overline{\theta}^C] \to \mathbb{R}$ is given by

$$V^C\,(\theta) := \frac{\phi^C - 1}{\phi^C}\theta + \frac{F^C\,(\theta)}{f^C\,(\theta)} - \frac{f^I\,(\theta)}{f^C\,(\theta)}\left[\frac{\phi^I - 1}{\phi^I}\theta + \frac{F^I\,(\theta)}{f^I\,(\theta)}\right].$$

Then, minimizing $R^I(\mathbf{x}^C)$ over $\mathcal{M}$ while sustaining $\boldsymbol{\chi}^C$ corresponds to maximizing $\int_{\overline{\theta}^C}^{\overline{\theta}^I}\mathbf{x}(\theta)g^I\,(\theta)\,d\theta$ by choosing a non-decreasing function $\mathbf{x} : (\overline{\theta}^C, \overline{\theta}^I] \to [\boldsymbol{\chi}^C(\overline{\theta}^C), \overline{a}]$. Although $g^I(\overline{\theta}^I) > 0$ and by continuity also for θ close to $\overline{\theta}^I$, nothing ensures that $g^I(\theta)$ is always positive, or at least non-decreasing. I therefore apply Toikka's (2011).

Let $\widetilde{F}$ be the uniform distribution on $[\overline{\theta}^C, \overline{\theta}^I]$ and $\widetilde{F}^{-1} : [0,1] \to [\overline{\theta}^C, \overline{\theta}^I]$ its continuous and increasing inverse function. For $q \in [0,1]$, define

$$z\,(q) := g^I(\widetilde{F}^{-1}\,(q)) \quad \text{and} \quad Z\,(q) := \int_0^q z\,(y)\,dy.$$

Let $\Omega := conv Z$ be the convex hull of Z on $[0,1]$ (see, e.g., Rockafellar (1970), p.36), i.e., the highest convex function on $[0,1]$ that satisfies $\Omega \leq Z$. By definition, Ω is continuously differentiable except at possibly countably many points, so define $\omega : [0,1] \to \mathbb{R}$ as

$$\omega\,(q) := \Omega'\,(q)\,,$$

whenever $\Omega'\,(q)$ exists. Also, w.l.o.g. extend $\omega\,(q)$ by right-continuity to all $[0,1)$ and by left-continuity at 1. For $\theta \in [\overline{\theta}^C, \overline{\theta}^I]$, let $\overline{g}^I\,(\theta) := \omega(\widetilde{F}\,(\theta))$; so $\overline{g}^I$ is non-decreasing.

Lemma 10 *If z is continuous at $q \in (0,1)$, then so is ω.*

Proof. Continuity of z at q implies that $Z'(q) = z(q)$. There are two cases to consider. First, suppose $\Omega(q) < Z(q)$. By definition, $\omega(\cdot)$ must be constant at $\omega(q)$ in a neighborhood of q. It follows that ω is continuous at q. Second, suppose $\Omega(q) = Z(q)$. Since Ω is convex and $Z(y) \geq \Omega(y)$ for every y, we have:

$$\Omega^{+}(q) = \lim_{y \downarrow q} \frac{\Omega(y) - \Omega(q)}{y - q} \leq \lim_{y \downarrow q} \frac{Z(y) - Z(q)}{y - q} = Z^{+}(q)$$

$$\Omega^{-}(q) = \lim_{y \uparrow q} \frac{\Omega(q) - \Omega(y)}{q - y} \geq \lim_{y \uparrow q} \frac{Z(q) - Z(y)}{q - y} = Z^{-}(q).$$

Since $\Omega^{-}(q) \leq \Omega^{+}(q)$ and Z is differentiable at q, we have that $\Omega^{-}(q) = \Omega^{+}(q)$, and ω is continuous at q. ■

Let $\theta^m := \max\{\overline{\theta}^C, \underline{\theta}^I\} < \overline{\theta}^I$ and $q^m := \widetilde{F}(\theta^m) < 1$. Since g^I is continuous over $[\underline{\theta}^I, \overline{\theta}^I]$, by Lemma 10 ω is continuous over $[q^m, 1]$, hence $\overline{g}^I$ is continuous over $[\theta^m, \overline{\theta}^I]$.

Lemma 11 $\overline{g}^I(\overline{\theta}^I) \geq g^I(\overline{\theta}^I)$.

Proof. If the opposite holds, then $\omega(1) < z(1)$. Since z is continuous over $[q^m, 1]$ and ω is non-decreasing, there exists $q < 1$ such that $\omega(y) < z(y)$ for every $y \geq q$. Since $Z(1) = \Omega(1)$, it follows that

$$Z(q) = Z(1) - \int_q^1 z(y)\,dy < \Omega(1) - \int_q^1 \omega(y)\,dy = \Omega(q).$$

A contradiction. ■

It follows that $\theta_{\mathrm{u}} := \inf\{\theta \in [\overline{\theta}^C, \overline{\theta}^I] \mid \overline{g}^I(\theta) > 0\} < \overline{\theta}^I$. Similarly, define $\theta_{\mathrm{d}} := \sup\{\theta \in [\overline{\theta}^C, \overline{\theta}^I] \mid \overline{g}^I(\theta) < 0\}$, if the set is non-empty, otherwise $\theta_{\mathrm{d}} := \overline{\theta}^C$. By Theorem 3.7 of Toikka (2011), any best extension $\mathbf{x}^C$ of $\boldsymbol{\chi}^C$ must satisfy $\mathbf{x}^C(\theta) = \overline{a}$ for $\theta \in (\theta_{\mathrm{u}}, \overline{\theta}^I)$ and $\mathbf{x}^C(\theta) = \boldsymbol{\chi}^C(\overline{\theta}^C)$ for $\theta \in (\overline{\theta}^C, \theta_{\mathrm{d}})$, if any. Letting $\mathbf{x}^C(\theta_{\mathrm{d}}) = \boldsymbol{\chi}^C(\overline{\theta}^C)$ is w.l.o.g.. Finally, over $(\theta_{\mathrm{d}}, \theta_{\mathrm{u}}]$, $\mathbf{x}^C$ can be any non-decreasing function mapping to $[\boldsymbol{\chi}^C(\overline{\theta}^C), \overline{a}]$, so long as it satisfies the necessary *pooling property* described by Toikka (see Definition 3.5). By Corollary 3.8 of Toikka (2011), it is then w.l.o.g. to set $\mathbf{x}^C(\theta) = \boldsymbol{\chi}^C(\overline{\theta}^C)$ for $\theta \in (\theta_{\mathrm{d}}, \theta_{\mathrm{u}}]$.

9.5 Proof of Corollary 1

Using *EU*, we have

$$R^C(\mathbf{x}^I) = \int_{\underline{\theta}}^{\overline{\theta}} \left[\frac{1 - \phi^C}{\phi^C} \theta \mathbf{x}^I(\theta) + \int_{\underline{\theta}}^{\theta} \mathbf{x}^I(y)\,dy \right] dF^C - \int_{\underline{\theta}}^{\overline{\theta}} \left[\frac{1 - \phi^I}{\phi^I} \theta \mathbf{x}^I(\theta) + \int_{\underline{\theta}}^{\theta} \mathbf{x}^I(y)\,dy \right] dF^I$$

The same steps as in the derivation of $R^I(\mathbf{x}^C)$ in (7) lead to

$$R^C(\mathbf{x}^I) = -\int_{\underline{\theta}^C}^{\underline{\theta}^I} \mathbf{x}^I(\theta)\, g^C(\theta)\, d\theta - \int_{\underline{\theta}^I}^{\overline{\theta}^I} \mathbf{x}^I(\theta)\, V^I(\theta)\, dF^I(\theta), \tag{8}$$

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

$$g^C(\theta) := \tfrac{\phi^C - 1}{\phi^C}\theta f^C(\theta) + F^C(\theta) \quad \text{and} \quad V^I(\theta) := v^I(\theta) - \tfrac{f^C(\theta)}{f^I(\theta)} v^C(\theta). \tag{9}$$

Observe that $g^C(\theta) > 0$ for $\theta \in (\underline{\theta}^C, \underline{\theta}^I)$. Therefore, given $\boldsymbol{\chi}^I$, the rule $\mathbf{x}^I \in \mathcal{M}$ that minimizes $R^C(\mathbf{x}^I)$ while sustaining $\boldsymbol{\chi}^I$ must satisfy $\mathbf{x}^I(\theta) = \boldsymbol{\chi}^I(\underline{\theta}^I)$ for $\theta \in (\underline{\theta}^C, \underline{\theta}^I)$. So, let $\hat{\mathbf{x}}^I(\theta) = \boldsymbol{\chi}^I(\underline{\theta}^I)$ for $\theta \in [\underline{\theta}^C, \underline{\theta}^I)$, and $\hat{\mathbf{x}}^I(\theta) = \boldsymbol{\chi}^I(\theta)$ otherwise. Using (7) and the best extension $\mathbf{x}^C_{\mathrm{u}}$ in Lemma 3 of $\boldsymbol{\chi}^C$, the RR condition becomes

$$R^C(\hat{\mathbf{x}}^I) + \int_{\underline{\theta}^C}^{\overline{\theta}^C} \boldsymbol{\chi}^C(\theta) V^C(\theta)\, dF^C - \boldsymbol{\chi}^C(\overline{\theta}^C) \int_{\overline{\theta}^C}^{\overline{\theta}^I} g^I(\theta)\, d\theta \leq (\overline{a} - \boldsymbol{\chi}^C(\overline{\theta}^C)) \int_{\theta_{\mathrm{u}}}^{\overline{\theta}^I} g^I(\theta)\, d\theta.$$

By assumption the left hand side is positive; by construction $R^C(\hat{\mathbf{x}}^I)$ has been minimized, and $\int_{\theta_{\mathrm{u}}}^{\overline{\theta}^I} g^I(\theta)\, d\theta > 0$. The result follows.

9.6 Proof of Lemma 4

The space $\mathcal{X} := \{(\mathbf{x}^C, \mathbf{x}^I) \mid \mathbf{x}^i : \Theta \to \mathbb{R}, i = C, I\}$ is linear and $\mathcal{Y} := \mathcal{M} \times \mathcal{M}$ is a convex subset of $\mathcal{X}$. The constraint functional $R^I(\cdot) + R^C(\cdot)$ maps $\mathcal{X}$ into $\mathbb{R}$, which has nonempty, positive, and closed cone $\mathbb{R}_+$. The objective is concave because $W^i(\mathbf{x})$ is so due to the concavity of d and $-c$. The constraint $R^I(\cdot) + R^C(\cdot)$ is linear, and there exists $(\mathbf{x}^C, \mathbf{x}^I) \in \mathcal{Y}$ such that $R^I(\mathbf{x}^C) + R^C(\mathbf{x}^I) < 0$: e.g., $\mathbf{x}^C = \mathbf{x}^{C*}$ and $\mathbf{x}^I$ constant.

Let $\mu \geq 0$ and define the Lagrangian

$$\begin{aligned} L(\mathbf{x}^C, \mathbf{x}^I, \mu) : &= \gamma W^C(\mathbf{x}^C) + (1-\gamma)\left[W^I(\mathbf{x}^I) - \tfrac{\pi\gamma}{1-\gamma} R^C(\mathbf{x}^I)\right] - \mu[R^I(\mathbf{x}^C) + R^C(\mathbf{x}^I)] \\ &= \gamma[W^C(\mathbf{x}^C) - \rho^I R^I(\mathbf{x}^C)] + (1-\gamma)\left[W^I(\mathbf{x}^I) - \rho^C R^C(\mathbf{x}^I)\right]. \end{aligned} \tag{10}$$

By Corollary 1, p. 219, and Theorem 2, p. 221, of Luenberger (1969), $(\mathbf{x}^C, \mathbf{x}^I)$ solve $\mathcal{P}'$ if and only if there exists $\mu \geq 0$ such that, for all $(\mathbf{x}^C_0, \mathbf{x}^I_0) \in \mathcal{Y}, \mu_0 \geq 0$,

$$L(\mathbf{x}^C, \mathbf{x}^I, \mu_0) \geq L(\mathbf{x}^C, \mathbf{x}^I, \mu) \geq L(\mathbf{x}^C_0, \mathbf{x}^I_0, \mu).$$

Given $\mu \geq 0$, $\mathbf{x}^C$ and $\mathbf{x}^I$ maximize the first and the second term in brackets in (10) within $\mathcal{M}$ if and only if $L(\mathbf{x}^C, \mathbf{x}^I, \mu) \geq L(\mathbf{x}^C_0, \mathbf{x}^I_0, \mu)$ for every $(\mathbf{x}^C_0, \mathbf{x}^I_0) \in \mathcal{Y}$. Moreover, if $(\mathbf{x}^C, \mathbf{x}^I, \mu)$ satisfies $R^C(\mathbf{x}^I) + R^I(\mathbf{x}^C) \leq 0$ and $\mu[R^C(\mathbf{x}^I) + R^I(\mathbf{x}^C)] = 0$, then $L(\mathbf{x}^C, \mathbf{x}^I, \mu_0) \geq L(\mathbf{x}^C, \mathbf{x}^I, \mu)$ for every $\mu_0 \geq 0$. Finally, if $(\mathbf{x}^C, \mathbf{x}^I)$ solves $\mathcal{P}'$, then $R^C(\mathbf{x}^I) + R^I(\mathbf{x}^C) \leq 0$. And if $R^C(\mathbf{x}^I) + R^I(\mathbf{x}^C) < 0$, then μ must equal zero: otherwise there exists $\mu_0 \in [0, \mu)$ such that

$$L(\mathbf{x}^C, \mathbf{x}^I, \mu_0) - L(\mathbf{x}^C, \mathbf{x}^I, \mu) = (\mu - \mu_0)\left[R^I(\mathbf{x}^C) + R^C(\mathbf{x}^I)\right] < 0.$$

9.7 Proof of Lemma 5

It follows from ES and (8) in the proof of Corollary 1.

9.8 Proof of Proposition 2

Part 1: Existence and Uniqueness.

Step 1: Construction of the generalized version of VS^I using Toikka's (2011) technique over $[\underline{\theta}^I, \overline{\theta}^I]$.

Since the density f is positive, the inverse function $(F^I)^{-1} : [0,1] \to [\underline{\theta}^I, \overline{\theta}^I]$ is well-defined, increasing, and continuous. Fix ρ^C and define, for $q \in [0,1]$,

$$z(q;\rho^C) := w^I((F^I)^{-1}(q);\rho^C) \quad \text{and} \quad Z(q;\rho^C) := \textstyle\int_0^q z(y;\rho^C)dy.$$

The function z is continuous in q, except possibly at $q^m := F^I(\theta^m) < 1$ (recall $\theta^m = \max\{\overline{\theta}^C, \underline{\theta}^I\}$): if $\phi^C > 1$ and $\underline{\theta}^I < \overline{\theta}^C$, so that $\theta^m = \overline{\theta}^C$, we have

$$\lim_{\theta \uparrow \overline{\theta}^C} w^I(\theta;\rho^C) = \lim_{\theta \downarrow \overline{\theta}^C} w^I(\theta;\rho^C) - \rho^C \frac{f^C(\overline{\theta}^C)}{f^I(\overline{\theta}^C)} \frac{\phi^C - 1}{\phi^C} \overline{\theta}^C. \tag{11}$$

Let $\Omega := convZ$ be the convex hull of Z on $[0,1]$. Let $\omega : [0,1] \to \mathbb{R}$ be defined as $\omega(q;\rho^C) := \Omega'(q;\rho^C)$, whenever $\Omega'(q;\rho^C)$ exists. Also, w.l.o.g., extend $\omega(q;\rho^C)$ by right-continuity to all $[0,1)$ and by left-continuity at 1.

Lemma 12 *The function ω is continuous in q and ρ^C.*

Proof. (Continuity in q). Throughout this part of the proof suppress ρ^C. Continuity at 0 and 1 holds by construction. Since z is continuous at every $q \in (0,1) \setminus \{q^m\}$, so is ω for the same argument as in Lemma 10. Consider now q^m. If $\theta^m = \underline{\theta}^I$, then $q^m = 0$ and we are done. To prove that ω is continuous if $q^m \in (0,1)$, it is enough to show that if z jumps at q^m, then $\Omega(q^m) < Z(q^m)$. Recall that

$$z(q^m-) := \lim\nolimits_{q \uparrow q^m} z(q) = \lim\nolimits_{\theta \uparrow \theta^m} w^I(\theta;\rho^C) \quad \text{and} \quad z(q^m+) := \lim\nolimits_{q \downarrow q^m} z(q) = \lim\nolimits_{\theta \downarrow \theta^m} w^I(\theta;\rho^C).$$

Hence z can at most jump down at q^m (see (11)), in which case $z(q^m-) > z(q^m+)$. Also recall that $z(q^m-) = z(q^m)$. Suppose $\Omega(q^m) = Z(q^m)$. By the same steps as in Lemma 10, $\Omega^-(q^m) \geq Z^-(q^m) = z(q^m)$. By convexity of Ω, $\omega(q) \geq \Omega^-(q^m)$ for all $q \geq q^m$. Therefore, for q close enough to q^m from the right

$$\Omega(q) = \textstyle\int_{q^m}^q \omega(s)\,ds + \Omega(q^m) > \int_{q^m}^q z(s)\,ds + Z(q^m) = Z(q).$$

A contradiction.

(Continuity in ρ^C). The function $Z(q;\rho^C)$ is continuous in ρ^C for every q. So Ω is continuous if $q \in \{0,1\}$, since $\Omega(0;\rho^C) = Z(0;\rho^C)$ and $\Omega(1;\rho^C) = Z(1;\rho^C)$. Consider $q \in (0,1)$. For every q and $\rho^C \geq 0$, Ω is defined as

$$\Omega(q;\rho^C) := \min\left\{\alpha Z(q_1;\rho^C) + (1-\alpha) Z(q_2;\rho^C) \mid (\alpha, q_1, q_2) \in [0,1], q = \alpha q_1 + (1-\alpha) q_2\right\}.$$

By continuity of $Z(q;\rho^C)$ and the Maximum Theorem, $\Omega(q,\cdot)$ is continuous in ρ^C for every q. Furthermore, $\Omega(\cdot;\rho^C)$ is differentiable in q with derivative $\omega(\cdot;\rho^C)$. Fix $q \in (0,1)$ and any sequence $\{\rho_n^C\}$ such that $\lim_{n\to\infty}\rho_n^C=\rho^C$. Since $\lim_{n\to\infty}\Omega(q;\rho_n^C)=\Omega(q;\rho^C)$, it follows from Theorem 25.7, p. 248, of Rockafellar (1970) that $\lim_{n\to\infty}\omega(q;\rho_n^C)=\omega(q;\rho^C)$. ■

Now, for $\theta\in\Theta^I$, define the generalized virtual valuation

$$\overline{w}^I(\theta;\rho^C):=\omega(F^I(\theta);\rho^C),$$

which is non-decreasing by construction and continuous by Lemma 12. Replace w^I with $\overline{w}^I$ in VS^I to get:

$$\overline{VS}^I(\mathbf{x};\rho^C):=\int_{\underline{\theta}^I}^{\overline{\theta}^I}\left[\mathbf{x}(\theta)\,\overline{w}^I(\theta;\rho^C)+d(\mathbf{x}(\theta))-c(\mathbf{x}(\theta))\right]dF^I+\rho^C\int_{\underline{\theta}^C}^{\underline{\theta}^I}\mathbf{x}(\theta)\,g^C(\theta)\,d\theta.$$

Step 2: Derivation of a candidate solution, which maximizes $\overline{VS}^I$, in two steps.

For $\theta\in\Theta^I$, define $\upsilon(a;\theta;\rho^C):=a\overline{w}^I(\theta;\rho^C)+d(a)-c(a)$ and let

$$\overline{\mathbf{x}}^I(\theta;\rho^C):=\arg\max_{a\in A}\upsilon(a;\theta;\rho^C),\tag{12}$$

and $\overline{\mathbf{x}}^I(\theta;\rho^C)=\overline{a}$ for $\theta\in[\underline{\theta}^C,\underline{\theta}^I)$; $\overline{\mathbf{x}}^I$ is the unique pointwise maximizer of $\overline{VS}^I$. Although $\overline{\mathbf{x}}^I(\rho^C)$ is non-decreasing on Θ^I, it may violate MON^I. The next lemma shows that any monotone maximizer of $\overline{VS}^I$, if it exists, must belong to a certain subclass of $\mathcal{M}$.

Lemma 13 *Suppose $\mathbf{x}^I\in\mathcal{M}$ and $\overline{VS}^I(\mathbf{x}^I;\rho^C)=\max_{\mathbf{x}\in\mathcal{M}}\overline{VS}^I(\mathbf{x};\rho^C)$. Then, $\mathbf{x}^I$ must satisfy*

$$\mathbf{x}^I(\theta;\rho^C)=\begin{cases}a_{\mathrm{b}}(\rho^C) & \text{if } \underline{\theta}^C<\theta\le\theta_{\mathrm{b}}\\ \overline{\mathbf{x}}^I(\theta;\rho^C) & \text{if } \theta_{\mathrm{b}}<\theta<\overline{\theta}^I\end{cases},$$

where $\theta_{\mathrm{b}}\in[\underline{\theta}^I,\overline{\theta}^I]$ and $a_{\mathrm{b}}(\rho^C)\ge\overline{\mathbf{x}}^I(\theta_{\mathrm{b}};\rho^C)$. Furthermore, if $\theta_{\mathrm{b}}<\overline{\theta}^I$, then $a_{\mathrm{b}}(\rho^C)=\overline{\mathbf{x}}^I(\theta_{\mathrm{b}};\rho^C)$.

Proof. Suppress ρ^C and suppose that $\mathbf{x}^I\in\mathcal{M}$ maximizes $\overline{VS}^I$. I claim that $\mathbf{x}^I(\theta)=\mathbf{x}^I(\underline{\theta}^I)$ for $\theta\in(\underline{\theta}^C,\underline{\theta}^I]$. Otherwise, there exists $\theta'\in(\underline{\theta}^C,\underline{\theta}^I)$ such that $\mathbf{x}^I(\theta)<\mathbf{x}^I(\underline{\theta}^I)$ for every $\theta<\theta'$. But then $\mathbf{x}^I$ can't be optimal within $\mathcal{M}$ because

$$\int_{\underline{\theta}^C}^{\underline{\theta}^I}\left[\mathbf{x}^I(\underline{\theta}^I)-\mathbf{x}^I(\theta)\right]g^C(\theta)\,d\theta\ge\int_{\underline{\theta}^C}^{\theta'}\left[\mathbf{x}^I(\underline{\theta}^I)-\mathbf{x}^I(\theta)\right]g^C(\theta)\,d\theta>0.$$

Now consider $\mathbf{x}^I(\theta)$ for $\theta\in\Theta^I$. Recall that $\upsilon(a,\theta)$ in (12) is strictly concave in a and continuous in θ. Because $\overline{\mathbf{x}}^I(\theta)$ is continuous and non-decreasing over Θ^I, only two cases can arise.

Case 1: $\mathbf{x}^I(\underline{\theta}^I)\ge\overline{\mathbf{x}}^I(\overline{\theta}^I)$. I claim that $\mathbf{x}^I(\theta)=\mathbf{x}^I(\underline{\theta}^I)$ for $\theta\in[\underline{\theta}^I,\overline{\theta}^I)$. If not, there exists $\theta'<\overline{\theta}^I$ such that $\mathbf{x}^I(\theta)>\mathbf{x}^I(\underline{\theta}^I)\ge\overline{\mathbf{x}}^I(\theta)$ for every $\theta\ge\theta'$. By strict concavity, for $\theta\in[\underline{\theta}^I,\overline{\theta}^I)$, $\upsilon(\mathbf{x}^I(\underline{\theta}^I),\theta)\ge\upsilon(\mathbf{x}^I(\theta),\theta)$, with strict inequality at least for every $\theta\ge\theta'$; hence $\int_{\Theta^I}\upsilon(\mathbf{x}^I(\underline{\theta}^I),\theta)dF^I>\int_{\Theta^I}\upsilon(\mathbf{x}^I(\theta),\theta)dF^I$, contradicting the optimality of $\mathbf{x}^I$.

Case 2: $\mathbf{x}^I(\underline{\theta}^I)=\overline{\mathbf{x}}^I(\theta_{\mathrm{b}})<\overline{\mathbf{x}}^I(\overline{\theta}^I)$ for some $\theta_{\mathrm{b}}\in[\underline{\theta}^I,\overline{\theta}^I)$. I claim that $\mathbf{x}^I(\theta)=\max\{\overline{\mathbf{x}}^I(\theta_{\mathrm{b}}),\overline{\mathbf{x}}^I(\theta)\}$ for $\theta\in[\underline{\theta}^I,\overline{\theta}^I)$. Suppose not. First, consider $[\underline{\theta}^I,\theta_{\mathrm{b}})$ and suppose that $\mathbf{x}^I(\theta)>\overline{\mathbf{x}}^I(\theta_{\mathrm{b}})$ for some $\theta<\theta_{\mathrm{b}}$. Then, by the same argument as in case 1, setting $\mathbf{x}^I(\theta)=\overline{\mathbf{x}}^I(\theta_{\mathrm{b}})$ for every $\theta\in[\underline{\theta}^I,\theta_{\mathrm{b}})$ is a strict improvement on $\mathbf{x}^I$: the resulting allocation rule is in $\mathcal{M}$ and $\int_{\underline{\theta}^I}^{\theta_{\mathrm{b}}}\upsilon(\overline{\mathbf{x}}^I(\theta_{\mathrm{b}}),\theta)dF^I>$

$\int_{\underline{\theta}^I}^{\theta_b} \upsilon\left(\mathbf{x}^I(\theta), \theta\right) dF^I$. Second, consider $[\theta_b, \overline{\theta}^I)$ and suppose $\mathbf{x}^I(\theta') \neq \overline{\mathbf{x}}^I(\theta')$ for some θ'. If $\mathbf{x}^I(\theta') > \overline{\mathbf{x}}^I(\theta')$, then by continuity of $\overline{\mathbf{x}}^I$ and monotonicity of $\mathbf{x}^I$ there exists a $\theta'' > \theta'$ such that $\mathbf{x}^I(\theta) > \overline{\mathbf{x}}^I(\theta)$ for every $\theta \in (\theta', \theta'')$. Similarly, if $\mathbf{x}^I(\theta') < \overline{\mathbf{x}}^I(\theta')$, then there exists $\theta''' < \theta'$ such that $\mathbf{x}^I(\theta) < \overline{\mathbf{x}}^I(\theta)$ for every $\theta \in (\theta''', \theta')$. Finally, since $\overline{\mathbf{x}}^I$ is the unique maximizer of $\upsilon(a, \theta)$, for $\theta \in [\theta_b, \overline{\theta}^I)$, $\upsilon(\overline{\mathbf{x}}^I(\theta), \theta) \geq \upsilon(\mathbf{x}^I(\theta), \theta)$, with strict inequality at least for every $\theta \in (\theta''', \theta') \cup (\theta', \theta'')$; hence $\int_{\theta_b}^{\overline{\theta}^I} \upsilon(\overline{\mathbf{x}}^I(\theta), \theta) dF^I > \int_{\theta_b}^{\overline{\theta}^I} \upsilon(\mathbf{x}^I(\theta), \theta) dF^I$, which contradicts the optimality of $\mathbf{x}^I$.
It remains to show that $\mathbf{x}^I(\underline{\theta}^I) < \overline{\mathbf{x}}^I(\underline{\theta}^I)$ is impossible. Suppose the opposite is true. By the same argument as in case 2, $\mathbf{x}^I(\theta) = \overline{\mathbf{x}}^I(\theta)$ for $\theta \in (\underline{\theta}^I, \overline{\theta}^I)$. Then setting $\mathbf{x}^I(\underline{\theta}^I) < \overline{\mathbf{x}}^I(\underline{\theta}^I)$ can't be optimal for the following reason: since $\mathbf{x}^I(\theta) = \mathbf{x}^I(\underline{\theta}^I)$ for $\theta \in (\underline{\theta}^C, \underline{\theta}^I)$, and $g^C(\theta)$ is positive, raising $\mathbf{x}^I(\underline{\theta}^I)$ up to $\overline{\mathbf{x}}^I(\underline{\theta}^I)$ satisfies monotonicity and strictly improves $\overline{VS}^I$. ■

By Lemma 13, $\mathbf{x}^I$ must be continuous on $(\underline{\theta}^C, \overline{\theta}^I)$. Although Lemma 13 doesn't pin down $\mathbf{x}^I$ at $\overline{\theta}^I$ and $\underline{\theta}^C$, it is w.l.o.g. to extend $\mathbf{x}^I$ at $\overline{\theta}^I$ and $\underline{\theta}^C$ by continuity. The next lemma proves that a maximizer of $\overline{VS}^I$ exists, and shows that it is unique over $(\underline{\theta}^C, \overline{\theta}^I)$ and, therefore, also over $[\underline{\theta}^C, \overline{\theta}^I]$ w.l.o.g..

Lemma 14 *There exists* $\mathbf{x}^I$ *such that* $\overline{VS}^I(\mathbf{x}^I; \rho^C) = \max_{\mathbf{x} \in \mathcal{M}} \overline{VS}^I(\mathbf{x}; \rho^C)$*; such* $\mathbf{x}^I$ *is unique.*

Proof. Suppress ρ^C. By Lemma 13, if a solution $\mathbf{x}^I$ exists, it can take only two forms: (1) $\mathbf{x}^I$ equals a constant $a \geq \overline{\mathbf{x}}^I(\overline{\theta}^I)$ over $[\underline{\theta}^C, \overline{\theta}^I]$, (2) $\mathbf{x}^I$ is constant at $\overline{\mathbf{x}}^I(\theta_b)$ over $[\underline{\theta}^C, \theta_b]$, with $\theta_b \geq \underline{\theta}^I$, and equals $\overline{\mathbf{x}}^I(\theta)$ for every $\theta \geq \theta_b$. So we only need to look for a solution within this subclass of $\mathcal{M}$.
Subclass 1: If $\mathbf{x}^I$ constant at a over $[\underline{\theta}^C, \overline{\theta}^I]$, then $\overline{VS}^I(\mathbf{x}^I) = VS^I(\mathbf{x}^I) = W^I(\mathbf{x}^I)$: the first equality follows from

$$\int_{\underline{\theta}^I}^{\overline{\theta}^I} [w^I(\theta) - \overline{w}^I(\theta)] dF^I = \int_0^1 [z(q) - \omega(q)]\, dq = 0;$$

the second from Proposition 1. Moreover, if $\mathbf{x}^I$ is constant at a over $[\underline{\theta}^C, \overline{\theta}^I]$

$$W^I(\mathbf{x}^I) = a \int_{\underline{\theta}^I}^{\overline{\theta}^I} (\theta/\phi^I) dF^I + d(a) - c(a). \tag{13}$$

Since $d(\cdot) - c(\cdot)$ is continuous and strictly concave, there exists a unique constant function that maximizes $\overline{VS}^I(\mathbf{x}^I)$. Call it $\mathbf{x}_1^I$.
Subclass 2: Using $\upsilon(a, \theta)$ in (12), $\overline{VS}^I$ equals to the following function of θ_b:

$$\Upsilon(\theta_b) = \int_{\underline{\theta}^I}^{\theta_b} \upsilon(\overline{\mathbf{x}}^I(\theta_b), \theta) dF^I + \int_{\theta_b}^{\overline{\theta}^I} \upsilon(\overline{\mathbf{x}}^I(\theta), \theta) dF^I + \overline{\mathbf{x}}^I(\theta_b) K, \tag{14}$$

where $K := \rho^C \int_{\underline{\theta}^C}^{\underline{\theta}^I} g^C(\theta)\, d\theta$. Continuity of $\overline{\mathbf{x}}^I$ implies that $\Upsilon(\theta_b)$ is continuous, therefore there exists an optimal $\theta_b \in [\underline{\theta}^I, \overline{\theta}^I]$ that completely identifies a maximizer within the second subclass of functions. Since $\overline{\mathbf{x}}^I$ can be locally flat, there can be multiple solutions θ_b. Nonetheless, I claim that there can't be two solutions θ_b^1 and θ_b^2 such that $\overline{\mathbf{x}}^I(\theta_b^1) \neq \overline{\mathbf{x}}^I(\theta_b^2)$. Suppose to the contrary that $\theta_b^1 < \theta_b^2$ both maximize $\Upsilon(\theta_b)$, and $\overline{\mathbf{x}}^I(\theta_b^1) < \overline{\mathbf{x}}^I(\theta_b^2)$. W.l.o.g. assume that θ_b^2 is the smallest θ such that $\overline{\mathbf{x}}^I(\theta) = \overline{\mathbf{x}}^I(\theta_{\rm b}^2)$. Let $\mathbf{x}^1$ and $\mathbf{x}^2$ be the allocation rules corresponding to $\theta_{\rm b}^1$ and $\theta_{\rm b}^2$, and for $\alpha \in (0,1)$ let $\widetilde{\mathbf{x}} := \alpha\mathbf{x}^1 + (1-\alpha)\mathbf{x}^2 \in \mathcal{M}$. For $\theta \in [\underline{\theta}^C, \theta_{\rm b}^2]$, $\mathbf{x}^2(\theta) \neq \mathbf{x}^1(\theta)$, whereas for $\theta \in [\theta_{\rm b}^2, \overline{\theta}^I]$, $\mathbf{x}^2(\theta) = \mathbf{x}^1(\theta) = \overline{\mathbf{x}}^I(\theta)$. By the strict concavity of $\upsilon(a,\theta)$, we have

$$\int_{\underline{\theta}^I}^{\theta_{\rm b}^2} \upsilon(\widetilde{\mathbf{x}}(\theta),\theta)dF^I + \int_{\theta_{\rm b}^2}^{\overline{\theta}^I} \upsilon(\widetilde{\mathbf{x}}(\theta),\theta)dF^I + \widetilde{\mathbf{x}}(\theta)K > \alpha\Upsilon(\theta_{\rm b}^1) + (1-\alpha)\Upsilon(\theta_{\rm b}^2).$$

Note that $\widetilde{\mathbf{x}}$ is constant on $\left[\underline{\theta}^C, \theta_{\rm b}^1\right]$ at some $\overline{\mathbf{x}}^I(\tilde{\theta}_{\rm b})$, with $\tilde{\theta}_{\rm b} \in (\theta_{\rm b}^1, \theta_{\rm b}^2)$. Therefore, the rule $\mathbf{x}^I(\theta) = \max\{\overline{\mathbf{x}}^I(\tilde{\theta}_{\rm b}), \overline{\mathbf{x}}^I(\theta)\}$ satisfies property (2) and, by the same argument as in Lemma 13 (Case 2),

$$\Upsilon(\tilde{\theta}_{\rm b}) \geq \int_{\underline{\theta}^I}^{\theta_{\rm b}^2} \upsilon(\widetilde{\mathbf{x}}(\theta),\theta)dF^I(\theta) + \int_{\theta_{\rm b}^2}^{\overline{\theta}^I} \upsilon(\widetilde{\mathbf{x}}(\theta),\theta)dF^I(\theta) + \widetilde{\mathbf{x}}(\theta)K > \Upsilon(\theta_{\rm b}^1).$$

The claim follows. Hence any maximizer of $\Upsilon(\theta_{\rm b})$ yields a unique $\mathbf{x}^I$ that satisfies property (2). Call it $\mathbf{x}_2^I$.

By an argument similar to that for the uniqueness of $\mathbf{x}_2^I$, $\overline{VS}^I(\mathbf{x}_2^I) = \overline{VS}^I(\mathbf{x}_1^I)$ if and only if $\mathbf{x}_2^I \equiv \mathbf{x}_1^I$ (over $(\underline{\theta}^C, \overline{\theta}^I)$). Therefore, the overall maximizer of $\overline{VS}^I$ is unique, and equals $\mathbf{x}_1^I$ if $\overline{VS}^I(\mathbf{x}_1^I) \geq \overline{VS}^I(\mathbf{x}_2^I)$, and $\mathbf{x}_2^I$ otherwise. ■

Step 3: The unique maximizer of $\overline{VS}^I$, denoted $\mathbf{x}^I(\rho^C)$, is also the unique maximizer of VS^I. The argument modifies Toikka's (2011) proof of Theorem 3.7 and Corollary 3.9 to account for $[\underline{\theta}^C, \underline{\theta}^I)$.

Lemma 15 *The allocation $\mathbf{x}^I(\rho^C)$ is the unique maximizer of VS^I.*

Proof. Suppress ρ^C. Integrating by parts and recalling that $\mathbf{x} \in \mathcal{M}$, we have

$$\begin{aligned}
\int_{\underline{\theta}^I}^{\overline{\theta}^I} \mathbf{x}(\theta)[w^I(\theta) - \overline{w}^I(\theta)]dF^I &= \int_{\underline{\theta}^I}^{\overline{\theta}^I} \mathbf{x}(\theta)[z(F^I(\theta)) - \omega(F^I(\theta))]dF^I \\
&= \mathbf{x}(\theta)[Z(F^I(\theta)) - \Omega(F^I(\theta))]\Big|_{\underline{\theta}^I}^{\overline{\theta}^I} \\
&\quad -\int_{\underline{\theta}^I}^{\overline{\theta}^I} [Z(F^I(\theta)) - \Omega(F^I(\theta))]d\mathbf{x}(\theta) \\
&= \int_{\underline{\theta}^I}^{\overline{\theta}^I} [\Omega(F^I(\theta)) - Z(F^I(\theta))]d\mathbf{x}(\theta) \leq 0.
\end{aligned}$$

The last equality follows from $Z(0) = \Omega(0)$ and $Z(1) = \Omega(1)$; the inequality follows from $\mathbf{x} \in \mathcal{M}$ and $\Omega(q) \leq Z(q)$ for $q \in [0,1]$. Re-writing VS^I, we have

$$\sup_{\mathbf{x}\in\mathcal{M}} VS^I(\mathbf{x}) = \sup_{\mathbf{x}\in\mathcal{M}} \{\overline{VS}^I(\mathbf{x}) + \int_{\underline{\theta}^I}^{\overline{\theta}^I} [\Omega(F^I(\theta)) - Z(F^I(\theta))]d\mathbf{x}(\theta)\}.$$

We know that $\mathbf{x}^I \in \mathcal{M}$ and achieves the supremum of $\overline{VS}^I(\mathbf{x})$. We only have to show that

$$\int_{\underline{\theta}^I}^{\overline{\theta}^I} [\Omega(F^I(\theta)) - Z(F^I(\theta))]d\mathbf{x}^I(\theta) = 0. \tag{15}$$

If $\mathbf{x}^I$ is constant over $[\underline{\theta}^C, \overline{\theta}^I]$, then $d\mathbf{x}^I \equiv 0$ and we are done. Otherwise, consider the pointwise solution $\overline{\mathbf{x}}^I$ over $[\underline{\theta}^I, \overline{\theta}^I]$ as defined in (12), and a θ such that $\Omega(F^I(\theta)) < Z(F^I(\theta))$. For some open interval N around θ, $\overline{w}^I(\cdot) = \omega(F^I(\theta))$, and $\overline{\mathbf{x}}^I$ is constant over N. Therefore, $d\overline{\mathbf{x}}^I(\cdot)$ assigns zero measure to any such N, and satisfies (15). I claim that also $d\mathbf{x}^I$ does so to any such N. Consider $[\underline{\theta}^C, \theta_{\mathrm{b}}]$, over which $\mathbf{x}^I$ is constant at $\overline{\mathbf{x}}^I(\theta_{\mathrm{b}})$. If $N \subset [\underline{\theta}^C, \theta_{\mathrm{b}}]$, the claim is immediate. The same holds if $N \cap [\underline{\theta}^C, \theta_{\mathrm{b}}] = \varnothing$, because then $\mathbf{x}^I(\theta) = \overline{\mathbf{x}}^I(\theta)$ for $\theta \in N$. Finally, if both $N \cap [\underline{\theta}^C, \theta_{\mathrm{b}}] \neq \varnothing$ and $N \cap (\theta_{\mathrm{b}}, \overline{\theta}^I] \neq \varnothing$ (therefore, $\theta_{\mathrm{b}} < \overline{\theta}^I$), then $\mathbf{x}^I$ equals $\overline{\mathbf{x}}^I(\theta_{\mathrm{b}})$ for every $\theta \in [\underline{\theta}^C, \theta_{\mathrm{b}}] \cup N$, which implies the claim. Hence (15) holds also for a non-constant $\mathbf{x}^I$. By Lemma 14, for any $\widetilde{\mathbf{x}} \in \mathcal{M}$ that differs from $\mathbf{x}^I$ over $(\underline{\theta}^C, \overline{\theta}^I)$, $\overline{VS}^I(\widetilde{\mathbf{x}}) < \overline{VS}^I(\mathbf{x}^I)$. Uniqueness follows on $(\underline{\theta}^C, \overline{\theta}^I)$, and extending it to $[\underline{\theta}^C, \overline{\theta}^I]$ is w.l.o.g.. ■

Part 2: Continuity and Limit Behavior of $\mathbf{x}^I$

I proved continuity of $\mathbf{x}^I$ in θ in Part 1; I now prove continuity in ρ^C. By the definition of $\overline{\mathbf{x}}^I(\cdot,\cdot)$ in (12) and the Maximum Theorem, $\overline{\mathbf{x}}^I(\theta,\cdot)$ is continuous in ρ^C for $\theta \in [\underline{\theta}^I, \overline{\theta}^I]$. Now consider $\Upsilon(\theta_{\mathrm{b}};\rho^C)$ in (14). Pointwise continuity of $\overline{w}^I(\theta;\rho^C)$ and $\overline{\mathbf{x}}^I(\theta;\rho^C)$ implies that $\Upsilon(\theta_{\mathrm{b}};\rho^C)$ is continuous in ρ^C, hence $\Theta_{\mathrm{b}}(\rho^C) := \arg\max_{\theta\in\Theta^I} \Upsilon(\theta;\rho^C)$ is u.h.c.. Furthermore, recall that for every $\theta, \theta' \in \Theta_{\mathrm{b}}(\rho^C)$, $\overline{\mathbf{x}}^I(\theta;\rho^C) = \overline{\mathbf{x}}^I(\theta';\rho^C)$. Take any sequence $\{\rho_n^C\}$ with $\lim_{n\to\infty}\rho_n^C = \rho^C$. Then, $\lim_{n\to\infty}\theta_{\mathrm{b}}(\rho_n^C) = \theta_{\mathrm{b}} \in \Theta_{\mathrm{b}}(\rho^C)$. Recall that the candidate $\mathbf{x}_2^I(\rho^C)$ maximizing $\Upsilon(\theta_{\mathrm{b}};\rho^C)$ satisfies $\mathbf{x}_2^I(\theta;\rho^C) = \max\{\overline{\mathbf{x}}^I(\theta_{\mathrm{b}}(\rho^C);\rho^C), \overline{\mathbf{x}}^I(\theta;\rho^C)\}$ for $\theta \in [\underline{\theta}^I, \overline{\theta}^I]$, and $\mathbf{x}_2^I(\theta;\rho^C) = \overline{\mathbf{x}}^I(\theta_{\mathrm{b}}(\rho^C);\rho^C)$ for $\theta < \underline{\theta}^I$. Hence, by continuity of $\overline{\mathbf{x}}^I$, $\lim_{n\to\infty}\mathbf{x}_2^I(\theta;\rho_n^C) = \mathbf{x}_2^I(\theta;\rho^C)$ for every $\theta \in [\underline{\theta}^C, \overline{\theta}^I]$. Finally, recall that the constant solution $\mathbf{x}_1^I$ in the proof of Lemma 14, as well as (13), is independent of ρ^C. It remains to show that the actual solution $\mathbf{x}^I(\rho_n^C)$ converges pointwise to the actual solution $\mathbf{x}^I(\rho^C)$. Suppose first that $VS^I(\mathbf{x}_1^I) > VS^I(\mathbf{x}_2^I(\rho^C)) = \Upsilon(\theta_{\mathrm{b}}(\rho^C);\rho^C)$. By continuity of Υ, there exists N such that $n \geq N$ implies $VS^I(\mathbf{x}_1^I) > VS^I(\mathbf{x}_2^I(\rho_n^C))$. Therefore, for $n \geq N$, $\mathbf{x}^I(\theta;\rho_n^C) = \mathbf{x}_1^I$ for every $\theta \in [\underline{\theta}^C, \overline{\theta}^I]$. Now, suppose that $VS^I(\mathbf{x}_1^I) < VS^I(\mathbf{x}_2^I(\rho^C))$. Similarly, for n large enough $\mathbf{x}^I(\rho_n^C) = \mathbf{x}_2^I(\rho_n^C)$ which converges pointwise to $\mathbf{x}^I(\rho^C)$. Finally, if $VS^I(\mathbf{x}_1^I) = VS^I(\mathbf{x}_2^I(\rho^C))$, then $\mathbf{x}_1^I \equiv \mathbf{x}_2^I(\rho^C)$. Hence $\left|\mathbf{x}_1^I - \mathbf{x}^I(\theta;\rho_n^C)\right| \leq \max\{0, \left|\mathbf{x}_1^I - \mathbf{x}_2^I(\theta;\rho_n^C)\right|\} \to 0$ as $n \to \infty$.

To prove that $\mathbf{x}^I(\rho^C) \to \mathbf{x}^{I*}$ pointwise as $\rho^C \to 0$, it is enough to observe that $VS^I(\mathbf{x}, 0) = W^I(\mathbf{x})$. Therefore, $\mathbf{x}^I(\theta, 0) = \mathbf{x}^{I*}(\theta)$ for $\theta \in (\underline{\theta}^I, \overline{\theta}^I)$, which can be extended to $[\underline{\theta}^C, \overline{\theta}^I]$ by letting $\mathbf{x}^I(\overline{\theta}^I, 0) = \mathbf{x}^{I*}(\overline{\theta}^I)$ and $\mathbf{x}^I(\theta, 0) = \mathbf{x}^{I*}(\underline{\theta}^I)$ for $\theta \leq \underline{\theta}^I$. Finally, I prove that $\max_\Theta \left|\mathbf{x}^I(\theta;\rho^C) - a^{\mathrm{nf}}\right| \to 0$ as $\rho^C \to +\infty$ by first showing pointwise convergence. Recall that $\mathbf{x}^I(\rho^C)$ maximizes $VS^I(\mathbf{x};\rho^C) = W^I(\mathbf{x}) - \rho^C R^C(\mathbf{x})$ and that, by Proposition 1, $R^C(\mathbf{x}) > 0$ for any $\mathbf{x} \in \mathcal{M}$ that is not constant on $(\underline{\theta}^C, \overline{\theta}^I)$. Clearly, $\mathbf{x}^I(\rho^C)$ cannot converge to a constant function that takes value $a_0 \neq a^{\mathrm{nf}}$, because a^{nf} is the unique maximizer of $W^I(\cdot)$ in (13). Now, suppose that $\mathbf{x}^I(\rho^C)$ converges pointwise to a function, denoted $\mathbf{x}_\infty^I$, that is not constant on $(\underline{\theta}^C, \overline{\theta}^I)$. Then, there exists $\hat{\rho}^C$ large enough so that for every $\rho^C > \hat{\rho}^C$, $\mathbf{x}_\infty^I$ is strictly dominated by a^{nf}. This is because $W^I(\mathbf{x}_\infty^I)$ is bounded and $R^C(\mathbf{x}_\infty^I) > 0$; hence there exists $\hat{\rho}^C \geq 0$ so that $W^I(\mathbf{x}_\infty^I) - \hat{\rho}^C R^C(\mathbf{x}_\infty^I) \leq W^I(a^{\mathrm{nf}})$. Now, consider the unique extension of $\mathbf{x}^I(\rho^C)$ by continuity. By monotonicity, we have

$$\max_\Theta \left|\mathbf{x}^I(\theta;\rho^C) - a^{\mathrm{nf}}\right| = \max\{\left|\mathbf{x}^I(\underline{\theta};\rho^C) - a^{\mathrm{nf}}\right|, \left|\mathbf{x}^I(\overline{\theta};\rho^C) - a^{\mathrm{nf}}\right|\},$$

which completes the argument.

9.9 Proof of Proposition 3

Property (b): Suppress ρ^C. Recall: (1) $\mathbf{x}^I$ satisfies Lemma 13; (2) for $\theta \in \Theta^I$, $\overline{\mathbf{x}}^I$ is defined by (12) and is continuous; (3) $\Upsilon(\cdot)$ as in (14). Let $\xi(\cdot) := c(\cdot) - d(\cdot)$. Suppose $\theta_{\mathrm{b}} < \overline{\theta}^I$. For $\theta > \theta_{\mathrm{b}}$, $\mathbf{x}^I(\theta) = \overline{\mathbf{x}}^I(\theta) > \overline{\mathbf{x}}^I(\theta_{\mathrm{b}}) = \mathbf{x}^I(\theta_{\mathrm{b}})$ by Lemma 13, and $\Upsilon(\theta_{\mathrm{b}}) \geq \Upsilon(\theta)$ by construction; hence

$$\begin{aligned}\frac{\Upsilon(\theta) - \Upsilon(\theta_{\mathrm{b}})}{\overline{\mathbf{x}}^I(\theta) - \overline{\mathbf{x}}^I(\theta_{\mathrm{b}})} &= \int_{\underline{\theta}^I}^{\theta} \overline{w}^I(y)\, dF^I - \frac{\xi(\overline{\mathbf{x}}^I(\theta)) - \xi(\overline{\mathbf{x}}^I(\theta_{\mathrm{b}}))}{\overline{\mathbf{x}}^I(\theta) - \overline{\mathbf{x}}^I(\theta_{\mathrm{b}})} F^I(\theta) + K \\ &\quad - \int_{\theta_{\mathrm{b}}}^{\theta} \frac{\overline{\mathbf{x}}^I(y) - \overline{\mathbf{x}}^I(\theta_{\mathrm{b}})}{\overline{\mathbf{x}}^I(\theta) - \overline{\mathbf{x}}^I(\theta_{\mathrm{b}})} \overline{w}^I(y)\, dF^I + \int_{\theta_{\mathrm{b}}}^{\theta} \frac{\xi(\overline{\mathbf{x}}^I(y)) - \xi(\overline{\mathbf{x}}^I(\theta_{\mathrm{b}}))}{\overline{\mathbf{x}}^I(\theta) - \overline{\mathbf{x}}^I(\theta_{\mathrm{b}})} dF^I \leq 0.\end{aligned}$$

Since $\overline{\mathbf{x}}^I$ and $\overline{w}^I$ are non-decreasing,

$$0 \leq \int_{\theta_{\mathrm{b}}}^{\theta} \frac{\overline{\mathbf{x}}^I(y) - \overline{\mathbf{x}}^I(\theta_{\mathrm{b}})}{\overline{\mathbf{x}}^I(\theta) - \overline{\mathbf{x}}^I(\theta_{\mathrm{b}})} \overline{w}^I(y)\, dF^I \leq \overline{w}^I(\overline{\theta}^I)(F^I(\theta) - F^I(\theta_{\mathrm{b}})).$$

Since $\overline{\mathbf{x}}^I$ is continuous, there exists $\tilde{\theta} \in (\theta_{\mathrm{b}}, \overline{\theta}^I]$ such that $\overline{\mathbf{x}}^I(\theta)$ is interior for $\theta \in (\theta_{\mathrm{b}}, \tilde{\theta})$. By the Mean Value Theorem, for every $\theta \in (\theta_{\mathrm{b}}, \tilde{\theta})$, there exists $y \in (\theta_{\mathrm{b}}, \theta)$ such that

$$\frac{\xi(\overline{\mathbf{x}}^I(\theta)) - \xi(\overline{\mathbf{x}}^I(\theta_{\mathrm{b}}))}{\overline{\mathbf{x}}^I(\theta) - \overline{\mathbf{x}}^I(\theta_{\mathrm{b}})} = \xi'(\overline{\mathbf{x}}(y)) = \overline{w}^I(y),$$

hence for $\theta \in (\theta_{\mathrm{b}}, \tilde{\theta})$

$$-\left|\overline{w}^I(\theta_{\mathrm{b}})\right| (F^I(\theta) - F^I(\theta_{\mathrm{b}})) \leq \int_{\theta_{\mathrm{b}}}^{\theta} \frac{\xi(\overline{\mathbf{x}}^I(y)) - \xi(\overline{\mathbf{x}}^I(\theta_{\mathrm{b}}))}{\overline{\mathbf{x}}^I(\theta) - \overline{\mathbf{x}}^I(\theta_{\mathrm{b}})} dF^I \leq \overline{w}^I(\overline{\theta}^I)(F^I(\theta) - F^I(\theta_{\mathrm{b}})).$$

Therefore,

$$\lim_{\theta \downarrow \theta_{\mathrm{b}}} \frac{\Upsilon(\theta) - \Upsilon(\theta_{\mathrm{b}})}{\overline{\mathbf{x}}^I(\theta) - \overline{\mathbf{x}}^I(\theta_{\mathrm{b}})} = \int_{\underline{\theta}^I}^{\theta_{\mathrm{b}}} \overline{w}^I(y)\, dF^I(y) - \xi'(\overline{\mathbf{x}}^I(\theta_{\mathrm{b}}))F^I(\theta_{\mathrm{b}}) + K \leq 0. \tag{16}$$

It follows that $\theta_{\mathrm{b}} > \underline{\theta}^I$ because $K > 0$ and $\xi'(\overline{\mathbf{x}}^I(\theta_{\mathrm{b}})) \in [\overline{w}^I(\underline{\theta}^I), \overline{w}^I(\overline{\theta}^I)]$.

I claim that there exists $\theta \in [\underline{\theta}^I, \theta_{\mathrm{b}})$ such that $\overline{\mathbf{x}}^I(\theta) < \overline{\mathbf{x}}^I(\theta_{\mathrm{b}})$. Suppose not. If $\overline{\mathbf{x}}^I(\theta_{\mathrm{b}})$ is interior, $\overline{w}^I(\theta) = \xi'(\overline{\mathbf{x}}^I(\theta))$ for every $\theta \leq \theta_{\mathrm{b}}$, and (16) is violated. If $\overline{\mathbf{x}}^I(\theta_{\mathrm{b}}) = \underline{a}$ and the set $\Theta_{\underline{a}} := \left\{\theta \in \Theta^I \mid \overline{w}^I(\theta) < \xi'(\underline{a})\right\}$ is non-empty—if $\Theta_{\underline{a}} = \varnothing$, we are back to the previous case—then (16) is violated again. To see this, note that since θ_{b} is the largest θ for which $\overline{w}^I(\theta) = \xi'(\underline{a})$, $Z(F^I(\theta_{\mathrm{b}})) = \Omega(F^I(\theta_{\mathrm{b}}))$, i.e., $\int_{\underline{\theta}^I}^{\theta_{\mathrm{b}}} w^I(y)\, dF^I = \int_{\underline{\theta}^I}^{\theta_{\mathrm{b}}} \overline{w}^I(y)\, dF^I$; hence

$$\int_{\underline{\theta}^I}^{\theta_{\mathrm{b}}} [\overline{w}^I(y) - \xi'(\underline{a})] dF^I + K = \int_{\underline{\theta}^I}^{\theta_{\mathrm{b}}} (y/\phi^I - \xi'(\underline{a})) dF^I + \rho^C \left[\int_{\underline{\theta}^C}^{\underline{\theta}^I} g^C(y)\, dy + \int_{\underline{\theta}^I}^{\theta_{\mathrm{b}}} V^I(y)\, dF^I\right].$$

By Assumption 1 and convexity of ξ, $\underline{\theta}^I/\phi^I = \xi'(\mathbf{a}^*(\underline{\theta}^I/\phi^I)) \geq \xi'(\underline{a})$, so the first integral is positive. The second term is too, contradicting 16. To see this, integrate by parts:

$$\int_{\underline{\theta}^C}^{\underline{\theta}^I} g^C(\theta)\, d\theta = -\int_{\underline{\theta}^C}^{\underline{\theta}^I} \left(\theta/\phi^C\right) dF^C + \underline{\theta}^I F^C(\underline{\theta}^I),$$

$$\begin{aligned}\textstyle\int_{\underline{\theta}^I}^{\theta_b} v^i(\theta)\,dF^i &= \textstyle\int_{\underline{\theta}^I}^{\theta_b}(\theta/\phi^i)dF^i + \theta^b(1-F^i(\theta^b)) - \underline{\theta}^I(1-F^i(\underline{\theta}^I)) \\ &= \textstyle\int_{\underline{\theta}^I}^{\overline{\theta}^I}(\theta/\phi^i)dF^i - \int_{\theta_b}^{\overline{\theta}^I}(\theta/\phi^i - \theta_b)dF^i - \underline{\theta}^I(1-F^i(\underline{\theta}^I)).\end{aligned}$$

Thus, we have

$$\begin{aligned}\textstyle\int_{\underline{\theta}^C}^{\underline{\theta}^I} g^C(\theta)\,d\theta + \int_{\underline{\theta}^I}^{\theta_b} V^I(\theta)\,dF^I &= \textstyle\int_{\underline{\theta}^C}^{\underline{\theta}^I} g^C(\theta)\,d\theta + \int_{\underline{\theta}^I}^{\theta_b} v^I(\theta)\,dF^I - \int_{\underline{\theta}^I}^{\theta_b} v^C(\theta)\,dF^C \\ &= \textstyle\int_{\theta_b}^{\overline{\theta}^I}(\theta/\phi^C - \theta_b)dF^C - \int_{\theta_b}^{\overline{\theta}^I}(\theta/\phi^I - \theta_b)dF^I \\ &= \textstyle\int_{\theta_b/\phi^I}^{\theta_b/\phi^C}(\theta_b - s)dF > 0, \qquad (17)\end{aligned}$$

where the second equality uses $\int_{\underline{\theta}}^{\overline{\theta}} \theta dF^i = \phi^i E(s)$, the third is a change of variables, and the inequality follows from $\overline{\theta}^I > \theta_b > \underline{\theta}^I > 0$ and $\phi^I > \phi^C \geq 1$.
Now define $\theta^1 := \min\{\theta \mid \overline{\mathbf{x}}^I(\theta) = \overline{\mathbf{x}}^I(\theta_b)\} > \underline{\theta}^I$. For $\theta < \theta^1$, $\overline{\mathbf{x}}^I(\theta) < \overline{\mathbf{x}}^I(\theta^1)$ and $\Upsilon(\theta) \leq \Upsilon(\theta^1) = \Upsilon(\theta_b)$. The same steps that led to (16) yield

$$\lim_{\theta\uparrow\theta^1}\frac{\Upsilon(\theta^1)-\Upsilon(\theta)}{\overline{\mathbf{x}}^I(\theta^1)-\overline{\mathbf{x}}^I(\theta)} = \textstyle\int_{\underline{\theta}^I}^{\theta^1}\overline{w}^I(y)\,dF^I - \xi'(\overline{\mathbf{x}}^I(\theta^1))F^I(\theta^1) + K \geq 0.$$

As $\overline{\mathbf{x}}^I(y)$ is interior and constant over $[\theta^1, \theta_b]$, $\xi'(\overline{\mathbf{x}}^I(y)) = \overline{w}^I(y) = \overline{w}^I(\theta_b) = \xi'(\overline{\mathbf{x}}^I(\theta_b))$, and

$$0 \leq \textstyle\int_{\underline{\theta}^I}^{\theta^1}[\overline{w}^I(y) - \overline{w}^I(\theta^1)]dF^I + K = \int_{\underline{\theta}^I}^{\theta_b}[\overline{w}^I(y) - \overline{w}^I(\theta_b)]dF^I + K \leq 0.$$

Finally, since $Z(F^I(\theta_b)) = \Omega(F^I(\theta_b))$, the same argument as in Lemma 10 yields $w^I(\theta_b) = \overline{w}^I(\theta_b)$. Therefore,

$$\textstyle\int_{\underline{\theta}^I}^{\theta_b}[\overline{w}^I(y) - \overline{w}^I(\theta_b)]dF^I + K = \int_{\underline{\theta}^I}^{\theta_b}[w^I(y) - w^I(\theta_b)]dF^I + K,$$

which gives *NFB* by rearranging $w^I(y)$.

Property (a): Since both $\mathbf{x}^I(\rho^C)$ and $\mathbf{x}^{I*}$ are continuous and non-decreasing, it is enough to prove that $\mathbf{x}^I(\underline{\theta}^I;\rho^C) > \mathbf{x}^{I*}(\underline{\theta}^I)$ and $\mathbf{x}^I(\overline{\theta}^I;\rho^C) < \mathbf{x}^{I*}(\overline{\theta}^I)$.

Case 1: $\mathbf{x}^I$ not constant. This implies that $\theta_b < \overline{\theta}^I$, and by Lemma 13 $\mathbf{x}^I(\underline{\theta}^I;\rho^C) = \mathbf{x}^I(\theta_b;\rho^C) = \overline{\mathbf{x}}^I(\theta_b;\rho^C)$. To show that $\mathbf{x}^I(\underline{\theta}^I;\rho^C) > \mathbf{x}^{I*}(\underline{\theta}^I)$ for $\rho^C > 0$, it is enough to prove that $\overline{w}^I(\theta_b;\rho^C) > \underline{\theta}^I/\phi^I$. This inequality holds because θ_b and $\overline{w}^I(\theta_b;\rho^C)$ must satisfy (see the proof of Property (b))

$$\textstyle\int_{\underline{\theta}^I}^{\theta_b}(\overline{w}^I(\theta_b;\rho^C) - y/\phi^I)dF^I = \rho^C\left[\int_{\underline{\theta}^C}^{\underline{\theta}^I} g^C(y)\,dy + \int_{\underline{\theta}^I}^{\theta_b} V^I(y)\,dF^I\right] > 0.$$

I now show that $\mathbf{x}^I(\overline{\theta}^I;\rho^C) < \mathbf{x}^{I*}(\overline{\theta}^I)$. Given $\rho^C > 0$, let $\theta^b := \min\{\theta \mid \overline{w}^I(\theta;\rho^C) = \overline{w}^I(\overline{\theta}^I;\rho^C)\}$.

Lemma 16 *$\overline{w}^I(\overline{\theta}^I;\rho^C) \geq w^I(\overline{\theta}^I;\rho^C)$. Moreover, if the inequality is strict, then $\theta^b < \overline{\theta}^I$.*

Proof. Suppress ρ^C and recall that $\overline{w}^I(\overline{\theta}^I) = \omega(1)$ and $w^I(\overline{\theta}^I) = z(1)$. If $\omega(1) < z(1)$, the same argument as in Lemma 11 leads to a contradiction. Suppose that $\omega(1) > z(1)$ and let $\hat{q} := \inf\{q \mid \forall q' > q, \omega(q') > z(q')\}$; $\hat{q} < 1$ by continuity. Then, for every $1 > q > \hat{q}$,

$$Z(q) = Z(1) - \textstyle\int_q^1 z(y)\,dy > \Omega(1) - \int_q^1 \omega(y)\,dy = \Omega(q),$$

It follows that $\theta^{\rm b} \le (F^I)^{-1}(\hat{q}) < \overline{\theta}^I$. ■

So if $\overline{w}^I(\overline{\theta}^I;\rho^C) = w^I(\overline{\theta}^I;\rho^C)$, it equals $(\overline{\theta}^I/\phi^I)(1-\rho^C(\phi^I-1)) < \overline{\theta}^I/\phi^I$. If instead $\overline{w}^I(\overline{\theta}^I;\rho^C) > w^I(\overline{\theta}^I;\rho^C)$, then it is constant over $[\theta^{\rm b},\overline{\theta}^I]$ at $\overline{w}^I(\theta^{\rm b};\rho^C)$. Since at $\theta^{\rm b}$ it must be that $Z(F^I(\theta^{\rm b});\rho^C) = \Omega(F^I(\theta^{\rm b});\rho^C)$, and $Z(1;\rho^C) = \Omega(1;\rho^C)$, $\theta^{\rm b}$ must satisfy

$$\int_{\theta^{\rm b}}^{\overline{\theta}^I} [w^I(y;\rho^C) - \overline{w}^I(\theta^{\rm b};\rho^C)]dF^I = 0 \tag{18}$$

or equivalently,

$$\int_{\theta^{\rm b}}^{\overline{\theta}^I} (y/\phi^I - \overline{w}^I(\theta^{\rm b};\rho^C))dF^I = -\rho^C \int_{\theta^{\rm b}}^{\overline{\theta}^I} V^I(y)\, dF^I. \tag{19}$$

Integrating by parts, we have

$$\begin{aligned}\int_{\theta^{\rm b}}^{\overline{\theta}^I} V^I(y)\, dF^I &= \int_{\theta^{\rm b}}^{\overline{\theta}^I} v^I(y)\, dF^I - \int_{\theta^{\rm b}}^{\overline{\theta}^I} v^C(y)\, dF^C \\ &= \int_{\theta^{\rm b}}^{\overline{\theta}^I} (y/\phi^I - \theta^{\rm b})dF^I - \int_{\theta^{\rm b}}^{\overline{\theta}^I} (y/\phi^C - \theta^{\rm b})dF^C = -\int_{\theta^{\rm b}/\phi^I}^{\theta^{\rm b}/\phi^C} (\theta^{\rm b} - s)dF < 0,\end{aligned}$$

where the last equality follows from a change of variables and the inequality from $\overline{\theta}^I > \theta^{\rm b} > 0$ and $\phi^I > \phi^C \ge 1$. Therefore the left hand side of (19) must be positive, and $\overline{w}^I(\theta^{\rm b};\rho^C) < \overline{\theta}^I/\phi^I$. In either case, $\mathbf{x}^I(\overline{\theta}^I;\rho^C)$ must be interior and strictly smaller than $\mathbf{x}^{I*}(\overline{\theta}^I)$.

Case 2: $\mathbf{x}^I$ constant. From the proof of Lemma 14, $\mathbf{x}^I(\theta;\rho^C)$ equals $a^{\rm nf}$ for every $\theta \in \Theta$. Since $\underline{\theta}^I/\phi^I < E(s) < \overline{\theta}^I/\phi^I$, Assumption 1 implies $\mathbf{x}^{I*}(\underline{\theta}^I) < a^{\rm nf} < \mathbf{x}^{I*}(\overline{\theta}^I)$.

Property (c): By Assumption 2, $H^I(\theta) = \phi^I H(\theta/\phi^I)$ is non-increasing over $[\theta^\dagger, \overline{\theta}^I]$ where $\theta^\dagger = \phi^I s^\dagger$. For any $\theta, \theta' \in [\max\{\theta^\dagger, \theta^m\}, \overline{\theta}^I]$ (recall $\theta^m = \max\{\overline{\theta}^C, \underline{\theta}^I\}$) with $\theta' > \theta$,

$$w^I(\theta';\rho^C) - w^I(\theta;\rho^C) = \frac{\theta' - \theta}{\phi^I}(1 - \rho^C(\phi^I - 1)) + \rho^C(H^I(\theta') - H^I(\theta)).$$

Hence, there exists $0 < \overline{\rho}^C \le (\phi^I - 1)^{-1} + \frac{1}{2}$ with the following property: if $\rho^C > \overline{\rho}^C$, then $w^I(\cdot;\rho^C)$ is decreasing over $[\max\{\theta^\dagger, \theta^m\}, \overline{\theta}^I]$, which implies that $\overline{w}^I(\theta;\rho^C)$ and $\mathbf{x}^I(\rho^C)$ must be constant over $[\theta^{\rm b}, \overline{\theta}^I] \ne \varnothing$.

9.10 Proof of Lemma 7

Suppose $\rho^I > 0$. Using $R^I(\mathbf{x})$ in (7), rewrite $W^C(\mathbf{x}) - \rho^I R^I(\mathbf{x})$ as

$$VS^C(\mathbf{x}, \rho^I) = \int_{\underline{\theta}^C}^{\overline{\theta}^C} [\mathbf{x}(\theta)\, w^C(\theta, \rho^I) - \xi(\mathbf{x}(\theta))]dF^C + \rho^I \int_{\overline{\theta}^C}^{\overline{\theta}^I} \mathbf{x}(\theta)\, g^I(\theta)\, d\theta,$$

where $\xi(\cdot) = c(\cdot) - d(\cdot)$ and $w^C(\theta, \rho^I) = \theta/\phi^C - \rho^I V^C(\theta)$. Let $\mathbf{x}_{\rm u}^C$ be the best enlargement of $\boldsymbol{\chi}^{C*}$ as in Lemma 3 and $\mathcal{M}^*$ be the set of $\mathbf{x}^C \in \mathcal{M}$ that sustain $\boldsymbol{\chi}^{C*}$. By construction, $VS^C(\mathbf{x}_{\rm u}^C, \rho^I) = \max_{\mathbf{x}\in\mathcal{M}^*} VS^C(\mathbf{x}, \rho^I)$. I claim that there exists $\hat{\mathbf{x}}^C \in \mathcal{M}\backslash\mathcal{M}^*$ such that $VS^C(\hat{\mathbf{x}}^C, \rho^I) > VS^C(\mathbf{x}_{\rm u}^C, \rho^I)$. Focus on $[\underline{\theta}^C, \theta_m]$ with $\theta_m := \min\{\overline{\theta}^C, \underline{\theta}^I\}$, and let $\hat{w}^C$ be the generalized version of w^C over this interval obtained with the same techniques as in the proof of Proposition 2. Since w^C is continuous on $[\underline{\theta}^C, \theta_m]$, so is $\hat{w}^C$ (Lemma 10). Because $\rho^I > 0$, V^C implies $w^C(\theta, \rho^I) < \theta/\phi^C$ for $\theta \in (\underline{\theta}^C, \theta_m]$. I claim that $\hat{w}^C(\theta_m, \rho^I) < \theta_m/\phi^C$. By the same argument as in Lemma 16, $\hat{w}^C(\theta_m,\rho^I)\geq w^C(\theta_m,\rho^I)$ and, if the inequality is strict, then there exists $\theta_0<\theta_m$ such that $\hat{w}^C(\theta,\rho^I)=w^C(\theta_0,\rho^I)$ for $\theta\in[\theta_0,\theta_m]$. If $\hat{w}^C(\theta_m,\rho^I)=w^C(\theta_m,\rho^I)$, then the claim follows. If $\hat{w}^C(\theta_m,\rho^I)=w^C(\theta_0,\rho^I)$, then $\hat{w}^C(\theta_m,\rho^I)\leq\theta_0/\phi^C<\theta_m/\phi^C$. Finally, because $\hat{w}^C$ is continuous and non-decreasing, in either case there exists $\theta_1<\theta_m$ such that $\hat{w}^C(\theta,\rho^I)<\theta/\phi^C$ for $\theta\in[\theta_1,\theta_m]$. Construct $\hat{\mathbf{x}}^C$ by letting $\hat{\mathbf{x}}^C(\theta)=\arg\max_{a\in A}a\hat{w}^C(\theta,\rho^I)-\xi(a)$ if $\theta\in[\underline{\theta}^C,\theta_m]$, and $\mathbf{x}^C_{\mathrm{u}}(\theta)$ if $\theta\in(\theta_m,\overline{\theta}^I]$. Then, $\hat{\mathbf{x}}^C\in\mathcal{M}$, but $\hat{\mathbf{x}}^C(\theta)<\boldsymbol{\chi}^{C*}(\theta)$ for every $\theta\in[\theta_1,\theta_m]$; so $\hat{\mathbf{x}}^C\notin\mathcal{M}^*$. Finally,

$$\begin{aligned}VS^C(\hat{\mathbf{x}}^C,\rho^I)-VS^C(\mathbf{x}^C_{\mathrm{u}},\rho^I) &= \int_{\underline{\theta}^C}^{\theta_m}[\hat{\mathbf{x}}^C(\theta)\,w^C(\theta,\rho^I)-\xi\left(\hat{\mathbf{x}}^C(\theta)\right)]dF^C\\ &\quad-\int_{\underline{\theta}^C}^{\theta_m}[\mathbf{x}^C_{\mathrm{u}}(\theta)\,w^C(\theta,\rho^I)-\xi(\mathbf{x}^C_{\mathrm{u}}(\theta))]dF^C>0.\end{aligned}$$

9.11 Proof of Proposition 4

Recall that $\gamma\in(0,1)$ and let $\hat{\rho}^C:=\frac{\hat{\pi}\gamma+\mu}{1-\gamma}$ and $\rho^C:=\frac{\pi\gamma+\mu}{1-\gamma}$. By revealed optimality, $\hat{\rho}^C>\rho^C$ implies that $R^C(\mathbf{x}^I(\hat{\rho}^C))\leq R^C(\mathbf{x}^I(\rho^C))$. Uniqueness of $\mathbf{x}^I(\rho^C)$ (Proposition 2) implies $R^C(\mathbf{x}^I(\hat{\rho}^C))=R^C(\mathbf{x}^I(\rho^C))$ if $\hat{\rho}^C=\rho^C$.

For the first part, I claim that $R^C(\mathbf{x}^I(\hat{\rho}^C))+R^I(\mathbf{x}^{C*})>0$. By Lemma 4, $\mathbf{x}^{C*}$ is part of a solution to $\mathcal{P}'$ if and only if $\mathbf{x}^{C*}\in\arg\max_{\mathbf{x}\in\mathcal{M}}W^C(\mathbf{x})-\rho^IR^I(\mathbf{x})$. By Lemma 7 this occurs if and only if $\rho^I:=\frac{\mu}{\gamma}$ equals zero, which implies $\hat{\rho}^C=\frac{\hat{\pi}\gamma}{1-\gamma}$. By assumption $(\mathbf{x}^I(\frac{\hat{\pi}\gamma}{1-\gamma}),\mathbf{x}^{C*})$ is not a solution to $\mathcal{P}'$, and the claim follows by Lemma 4. Suppose $(\mathbf{x}^I(\rho^C),\mathbf{x}^{C*})$ solves $\mathcal{P}'$. Repeating the same argument implies $\rho^C=\frac{\pi\gamma}{1-\gamma}<\frac{\hat{\pi}\gamma}{1-\gamma}$. But then, $R^C(\mathbf{x}^I(\rho^C))+R^I(\mathbf{x}^{C*})>0$. A contradiction.

For the second part, suppose $(\mathbf{x}^I(\rho^C),\mathbf{x}^C)$ solves $\mathcal{P}'$ with $\mathbf{x}^C\in\mathcal{M}$ that sustains $\boldsymbol{\chi}^{C*}$. Lemma 4 and Lemma 7 imply that $\rho^C=\frac{\pi\gamma}{1-\gamma}$. By construction, the enlargement $\mathbf{x}^C_{\mathrm{u}}$ of $\boldsymbol{\chi}^{C*}$ (Lemma 3) implies $R^C(\mathbf{x}^I(\rho^C))+R^I(\mathbf{x}^C)\geq R^C(\mathbf{x}^I(\rho^C))+R^I(\mathbf{x}^C_{\mathrm{u}})$. By definition of $D(\boldsymbol{\chi}^{C*},\boldsymbol{\chi}^I(\frac{\hat{\pi}\gamma}{1-\gamma}))$ (Corollary 1), $\overline{a}<D(\boldsymbol{\chi}^{C*},\boldsymbol{\chi}^I(\frac{\hat{\pi}\gamma}{1-\gamma}))+\mathbf{a}^*(\overline{s})$ implies $R^C(\mathbf{x}^I(\frac{\hat{\pi}\gamma}{1-\gamma}))+R^I(\mathbf{x}^C_{\mathrm{u}})>0$. Since $R^C(\mathbf{x}^I(\frac{\pi\gamma}{1-\gamma}))\geq R^C(\mathbf{x}^I(\frac{\hat{\pi}\gamma}{1-\gamma}))$, $D(\boldsymbol{\chi}^{C*},\boldsymbol{\chi}^I(\frac{\pi\gamma}{1-\gamma}))\geq D(\boldsymbol{\chi}^{C*},\boldsymbol{\chi}^I(\frac{\hat{\pi}\gamma}{1-\gamma}))$.

9.12 Proof of Lemma 6

The inverse hazard rate of the uniform distribution is $H^i(\theta)=\overline{\theta}^i-\theta$. By Lemma 5, we have

$$w^I(\theta;\rho^C)=\begin{cases}(\theta/\phi^I)(1-\rho^C(2\phi^I-1))+\rho^C\overline{\theta}^I & \text{if } \theta\in(\overline{\theta}^C,\overline{\theta}^I]\\ (\theta/\phi^I)(1+\rho^C(\phi^I-1)^2) & \text{if } \theta\in[\underline{\theta}^I,\overline{\theta}^C]\end{cases}.$$

The function w^I is continuous at $\overline{\theta}^C$. It is increasing and greater than θ/ϕ^I over $[\underline{\theta}^I,\overline{\theta}^C]$, because $\rho^C>0$ and $\phi^I>1$; w^I is increasing over $(\overline{\theta}^C,\overline{\theta}^I]$ if and only if $\rho^C<(2\phi^I-1)^{-1}$. So, the threshold $\overline{\rho}^C$ in Proposition 3 equals $(2\phi^I-1)^{-1}$.

Consider first the behavior of θ_{b} and θ^{b}, when $\theta_{\mathrm{b}}<\theta^{\mathrm{b}}$ and $\mathbf{x}^I$ is not constant. If $\rho^C<(2\phi^I-1)^{-1}$, then w^I is increasing and coincides with $\overline{w}^I$ (see the proof of Proposition 2); hence $\overline{\mathbf{x}}^I$ (see (12)) is increasing over $[\underline{\theta}^I,\overline{\theta}^I]$, which implies $\theta^{\mathrm{b}}=\overline{\theta}^I$. Otherwise, $\theta^{\mathrm{b}}\leq\overline{\theta}^C<\overline{\theta}^I$ and θ^{b}

is characterized by (18), which boils down to

$$(\overline{\theta}^I - \theta^{\mathrm{b}})^2 = -\frac{1-\rho^C(\phi^I)^2}{1+\rho^C(\phi^I-1)^2}(\overline{\theta}^I - \overline{\theta}^C)^2 \tag{20}$$

Since w^I is always increasing over $[\underline{\theta}^I, \theta^{\mathrm{b}}]$, it coincides with $\overline{w}^I$. Using *NFB*, θ_{b} must satisfy

$$\int_{\underline{\theta}^I}^{\theta_{\mathrm{b}}}[w^I(y;\rho^C) - w^I(\theta_{\mathrm{b}};\rho^C)]dy = -(\overline{\theta}^I - \underline{\theta}^I)\rho^C\int_{\underline{\theta}^C}^{\underline{\theta}^I} g^C(y)\,dy. \tag{21}$$

Since

$$\frac{\partial}{\partial\theta_{\mathrm{b}}}\int_{\underline{\theta}^I}^{\theta_{\mathrm{b}}}[w^I(y;\rho^C) - w^I(\theta_{\mathrm{b}};\rho^C)]dy = -w^I_\theta(\theta_{\mathrm{b}};\rho^C)(\theta_{\mathrm{b}} - \underline{\theta}^I) < 0.$$

for $\rho^C > 0$, there is a unique $\theta_{\mathrm{b}} < \theta^{\mathrm{b}}$ that satisfies (21). Letting $K = \int_{\underline{\theta}^C}^{\underline{\theta}^I} g^C(y)\,dy > 0$, condition (21) becomes

$$\rho^C[2\phi^I(\overline{\theta}^I - \underline{\theta}^I)K] = \begin{cases} ((1+\rho^C(\phi^I-1)^2)(\theta_{\mathrm{b}} - \underline{\theta}^I)^2 & \text{if } \theta_{\mathrm{b}} \le \overline{\theta}^C \\ \rho^C(\phi^I)^2(\overline{\theta}^C - \underline{\theta}^I)^2 + (1-\rho^C(2\phi^I-1))(\theta_{\mathrm{b}} - \underline{\theta}^I)^2 & \text{if } \theta_{\mathrm{b}} > \overline{\theta}^C \end{cases}.$$

The function $\theta^{\mathrm{b}}(\rho^C)$ is constant at $\overline{\theta}^I$ for $\rho^C < (2\phi^I-1)^{-1}$ and at $(2\phi^I-1)^{-1}$ it jumps from $\overline{\theta}^I$ to $\overline{\theta}^C$. Monotonicity for $\rho^C > (2\phi^I-1)^{-1}$ follows by applying the Implicit Function Theorem to (20):

$$\frac{d\theta^{\mathrm{b}}}{d\rho^C} = -\frac{1}{2}\left[\frac{\phi^I}{1+\rho^C(\phi^I-1)^2}\right]^2 \frac{(\overline{\theta}^I - \overline{\theta}^C)^2}{(\overline{\theta}^I - \theta^{\mathrm{b}})} < 0.$$

Similarly, for $\theta_{\mathrm{b}}(\rho^C)$ we have:

$$\frac{d\theta_{\mathrm{b}}}{d\rho^C} = \begin{cases} \frac{\theta^{\mathrm{b}} - \underline{\theta}^I}{2\rho^C(1+\rho^C(\phi^I-1)^2)} > 0 & \text{if } \theta_{\mathrm{b}} \le \overline{\theta}^C \\ \frac{\theta^{\mathrm{b}} - \underline{\theta}^I}{2\rho^C(1-\rho^C(2\phi^I-1))} > 0 & \text{if } \theta_{\mathrm{b}} > \overline{\theta}^C \end{cases};$$

for the second inequality, recall that $\theta^{\mathrm{b}} > \theta_{\mathrm{b}} > \overline{\theta}^C$ if and only if $\rho^C < (2\phi^I-1)^{-1}$.

Consider now the behavior of $\mathbf{x}^I(\rho^C)$. By Proposition 3 and Assumption 1, $\mathbf{x}^I(\theta;\rho^C)$ is always interior. Also, $\mathbf{x}^I(\theta;\rho^C) = \arg\max_{a\in A} a\overline{w}^I(\theta;\rho^C) - \xi(a)$. By strict convexity of ξ, it is enough to consider the properties of $\overline{w}^I(\rho^C)$ in relation with the function θ/ϕ^I. The function $\overline{w}^I(\cdot;\rho^C)$ crosses θ/ϕ^I only once at some $\underline{\theta}^I < \theta^* < \overline{\theta}^I$. Furthermore, for $\theta \in [\theta_{\mathrm{b}}, \theta^{\mathrm{b}}]$, $\overline{w}^I(\theta;\rho^C) = w^I(\theta;\rho^C)$. Hence, it is enough to show that, as ρ^C rises, $w^I(\theta_{\mathrm{b}}(\rho^C);\rho^C)$ rises and $w^I(\theta^{\mathrm{b}}(\rho^C);\rho^C)$ falls.

Lemma 17 *Suppose θ_{b} and θ^{b} are characterized by NFB and (18). If $w^I_\theta(\theta_{\mathrm{b}};\rho^C) > 0$ and $w^I_\theta(\theta^{\mathrm{b}};\rho^C) > 0$, then $\frac{d}{d\rho^C}w^I(\theta_{\mathrm{b}}(\rho^C);\rho^C) > 0$ and $\frac{d}{d\rho^C}w^I(\theta^{\mathrm{b}}(\rho^C);\rho^C) < 0$.*

Proof. It follows by applying the Implicit Function Theorem to *NFB* and (18). ■

Consider $w^I(\theta^{\mathrm{b}}(\rho^C);\rho^C)$. If $\rho^C < \overline{\rho}^C$, then $\theta^{\mathrm{b}}(\rho^C) = \overline{\theta}^I$ and $w^I_\rho(\overline{\theta}^I;\rho^C) = (1-\phi^I)(\overline{\theta}^I/\phi^I) < 0$. Therefore, as $\rho^C \uparrow \overline{\rho}^C$, $w^I(\overline{\theta}^I;\rho^C) \downarrow w^I(\overline{\theta}^I, \overline{\rho}^C) = w^I(\overline{\theta}^C, \overline{\rho}^C)$. By Lemma 17 $w^I(\theta^{\mathrm{b}}(\rho^C);\rho^C)$ decreases in ρ^C, for $\rho^C > \overline{\rho}^C$, because here $w^I_\theta(\theta^{\mathrm{b}}(\rho^C);\rho^C)$ is positive when $\theta^{\mathrm{b}} < \overline{\theta}^C$. Hence, $\rho^C \downarrow \overline{\rho}^C$ implies $w^I(\theta^{\mathrm{b}}(\rho^C);\rho^C) \uparrow w^I(\overline{\theta}^C, \overline{\rho}^C)$. By the same argument $w^I(\theta_{\mathrm{b}}(\rho^C);\rho^C)$ increases in ρ^C again because here $w^I_\theta(\theta_{\mathrm{b}}(\rho^C);\rho^C)$ is positive when $\theta_{\mathrm{b}} < \theta^{\mathrm{b}}$.

9.13 Proof of Proposition 5

Step 1: There exists $\overline{a}$ large enough so that unused items suffice to satisfy *R-IC*^{i,j} for $i > j$.

Recall that *R-IC*^{i,j} is $U^i(\mathbf{x}^i, k^i) \geq U^j(\mathbf{x}^j, k^j) + R^i(\mathbf{x}^j)$. If $i < j$, $\underline{\theta}^i < \underline{\theta}^j$ and

$$R^i\left(\mathbf{x}^j\right) = -\int_{\underline{\theta}^i}^{\underline{\theta}^j} \mathbf{x}^j(\theta)\, G^i(\theta)\, d\theta - \int_{\underline{\theta}^j}^{\overline{\theta}^j} \mathbf{x}^j(\theta)\, V^{j,i}(\theta)\, dF^j,$$

where

$$G^i(\theta) = \frac{\phi^i - 1}{\phi^i}\theta f^i(\theta) + F^i(\theta) \quad \text{and} \quad V^{j,i}(\theta) = v^j(\theta) - \frac{f^i(\theta)}{f^j(\theta)} v^i(\theta);$$

if $i > j$, $\overline{\theta}^j < \overline{\theta}^i$ and

$$R^i\left(\mathbf{x}^j\right) = -\int_{\overline{\theta}^j}^{\overline{\theta}^i} \mathbf{x}^j(\theta)\, \overline{G}^i(\theta)\, d\theta - \int_{\underline{\theta}^j}^{\overline{\theta}^j} \mathbf{x}^j(\theta)\, \overline{V}^{j,i}(\theta)\, dF^j,$$

where

$$\overline{G}^i(\theta) = \frac{\phi^i - 1}{\phi^i}\theta f^i(\theta) - \left(1 - F^i(\theta)\right)$$

$$\overline{V}^{j,i}(\theta) = \frac{\phi^j - 1}{\phi^j}\theta + \frac{F^j(\theta)}{f^j(\theta)} - \frac{f^i(\theta)}{f^j(\theta)}\left[\frac{\phi^i - 1}{\phi^i}\theta + \frac{F^i(\theta)}{f^i(\theta)}\right].$$

Take $j > i$. Suppose that *R-IC*^{j,i} is binding (and all the other constraints are satisfied), i.e., $U^j(\mathbf{x}^j, k^j) < U^i(\mathbf{x}^i, k^i) + R^j(\mathbf{x}^i)$. Fix $\mathbf{x}^i$ for every $\theta \leq \overline{\theta}^i$, and let $\mathbf{x}^i(\theta) = \overline{a}$ for every $\theta > \overline{\theta}^i$. Then,

$$R^j\left(\mathbf{x}^i\right) = -\overline{a}\int_{\overline{\theta}^i}^{\overline{\theta}^j} \overline{G}^j(\theta)\, d\theta - \int_{\underline{\theta}^i}^{\overline{\theta}^i} \mathbf{x}^i(\theta)\, \overline{V}^{i,j}(\theta)\, dF^i.$$

Lemma 18 $\int_{\overline{\theta}^i}^{\overline{\theta}^j} \overline{G}^j(\theta)\, d\theta > 0$.

Proof. Rewrite the integral as

$$\int_{\overline{\theta}^i}^{\overline{\theta}^j} (\phi^j - 1)(\theta/\phi^j) f^j(\theta)\, d\theta - \int_{\overline{\theta}^i}^{\overline{\theta}^j} \left(1 - F^j(\theta)\right) d\theta.$$

Integrate by parts the second term, to get

$$-\int_{\overline{\theta}^i}^{\overline{\theta}^j} (\theta/\phi^j) f^j(\theta)\, d\theta + (1 - F^j(\overline{\theta}^i))\overline{\theta}^i = \int_{\overline{\theta}^i}^{\overline{\theta}^j} (\overline{\theta}^i - (\theta/\phi^j)) f^j(\theta)\, d\theta.$$

Finally, note that $\overline{\theta}^i \geq \overline{s} \geq \theta/\phi^j$, with strict inequality for $\theta \in (\overline{\theta}^j, \overline{\theta}^i)$. ∎

It follows that there exists $\overline{a}$ large enough so that the $\widetilde{\mathbf{x}}^i$ so constructed satisfies $U^j(\mathbf{x}^j, k^j) \geq U^i(\mathbf{x}^i, k^i) + R^j(\widetilde{\mathbf{x}}^i)$. Now, we need to check the other constraints. For all $j' < i$, the values that $\mathbf{x}^i$ takes for $\theta > \overline{\theta}^i$ are irrelevant; so, all *R-IC*$^{j',i}$ remain unchanged. For $\hat{\jmath} > i$ and $\hat{\jmath} \neq j$, it could be that $\widetilde{\mathbf{x}}^i$ is such that $R^{\hat{\jmath}}(\widetilde{\mathbf{x}}^i) > R^{\hat{\jmath}}(\mathbf{x}^i)$, and $\widetilde{\mathbf{x}}^i$ may violate *R-IC*$^{j',i}$ while $\mathbf{x}^i$ did not. However, since Lemma 18 holds for every $j > i$ and N is finite, clearly we can find $\overline{a}$ large enough so that no $j > i$ wants to mimic i. Thus, for the rest of the proof I will assume that *R-IC*^{j,i} never binds for $j > i$.

Step 2: Reformulation of the principal's problem.

Recall that, given $\mathbf{x}^j$, there is a one-to-one relationship between $U^j(\mathbf{x}^j, k^j)$ and k^j. Therefore define $u^j := U^j(\mathbf{x}^j, k^j)$, and write

$$R\text{-}IC^{i,j}\text{:}\quad u^i \geq u^j + R^i(\mathbf{x}^j) \quad \text{and} \quad IR^i\text{:}\quad u^i \geq 0.$$

As usual, IR^N and $R\text{-}IC^{i,N}$ imply IR^i for all $i < N$.

Now, let $\mathcal{Y} := (\mathcal{M} \times \mathbb{R})^N$ be the the subspace of $(\mathcal{X}\times\mathbb{R})^N$, where $\mathcal{X}=\{\mathbf{x}|\, \mathbf{x} : \Theta \to \mathbb{R}\}$. Let $\Pi(\mathbf{X}, \mathbf{u}) = \sum_{i=1}^N \gamma_i [W^i(\mathbf{x}^i) - u^i]$ and rewrite $\mathcal{P}^N$ as

$$\overline{\mathcal{P}}^N := \begin{cases} \max_{\{\mathbf{X},\mathbf{u}\}\in\mathcal{Y}} (1-\pi)W(\mathbf{X}) + \pi\Pi(\mathbf{X}, \mathbf{u}) \\ \text{s.t. } \Gamma(\mathbf{X}, \mathbf{u}) \leq \mathbf{0} \end{cases},$$

where the functional $\Gamma\colon (\mathcal{X}\times\mathbb{R})^N \to \mathbb{R}^r$ $(r = 1 + \frac{N(N-1)}{2})$ is given by $\Gamma^1(\mathbf{X}, \mathbf{u}) = -u^N$ and, for all $n = 2, \ldots, r$, $\Gamma^n(\mathbf{X}, \mathbf{u}) = R^i(\mathbf{x}^j) + u^j - u^i$ for some $i \in N$, $j > i$.

Step 3: Existence of interior points.

Lemma 19 *In $\overline{\mathcal{P}}^N$, there exists $\{\mathbf{X}, \mathbf{u}\} \in \mathcal{Y}$ such that $\Gamma(\mathbf{X}, \mathbf{u}) < 0$.*

Proof. The condition $\Gamma(\mathbf{X}, \mathbf{u}) < 0$ is equivalent to $u^N > 0$, and $u^i > u^j + R^i(\mathbf{x}^j)$ for every $i \in N$, $j > i$. For every $i \in N$, let $\mathbf{x}^i = \mathbf{x}^{i*}$ over $[\underline{\theta}, \overline{\theta}^i]$ and possibly extend it over $(\overline{\theta}^i, \overline{\theta}]$ to include appropriate unused items. Note that these extensions are irrelevant for $R^j(\mathbf{x}^i)$ if $j < i$. Recall that $R^j(\mathbf{x}^i) \geq 0$ for all $j < i$, and it can be easily shown that $R^1(\mathbf{x}^i) \geq R^j(\mathbf{x}^i)$ for all $1 < j < i$. Thus, let $u^N = 1$, and for every $i < N$, let $u^i = u^{i+1} + R^1(\mathbf{x}^{i+1}) + 1$. Now, fix $i < N$ and consider any $j > i$. We have

$$u^i = u^j + \textstyle\sum_{n=1}^{j-i} R^1\left(\mathbf{x}^{i+n}\right) + (j-i) \geq u^j + R^i\left(\mathbf{x}^j\right) + (j-i) > u^j + R^i\left(\mathbf{x}^j\right).$$

Note that since $R^i(\mathbf{x}^j)$ are bounded and N is finite, the constructed vector $\mathbf{u}$ is well defined. ■

Step 4: Lagrangian characterization of the solution to $\overline{\mathcal{P}}^N$.

To characterize the solutions to $\overline{\mathcal{P}}^N$, I use Corollary 1, p. 219, and Theorem 2, p. 221, of Luenberger (1969). Note that $(\mathcal{X}\times\mathbb{R})^N$ is a linear vector space, and $\mathbb{R}^r$ is normed spaces with the usual Euclidean norm, with positive closed cone containing an interior point. $\mathcal{Y}$ is a convex subset of $(\mathcal{X}\times\mathbb{R})^N$. By Lemma 19, Γ admits interior points. Finally, Π and W are concave by the assumptions on $d(\cdot)$ and $c(\cdot)$. Hence, the objective is concave and $\Gamma(\mathbf{X}, \mathbf{u})$ is convex.

Now, for $\boldsymbol{\lambda} \in\mathbb{R}_+^r$, define the Lagrangian

$$\begin{aligned} L(\mathbf{X}, \mathbf{u}; \boldsymbol{\lambda}) &= (1-\pi)W(\mathbf{X}) + \pi\Pi(\mathbf{X}, \mathbf{u}) + \lambda^N u^N - \textstyle\sum_{i=1}^N \sum_{j<i} \lambda^{j,i} \left(R^j(\mathbf{x}^i) + u^i - u^j\right) \\ &= \textstyle\sum_{i=1}^N \gamma_i \left[W^i(\mathbf{x}^i) - \sum_{j<i} \frac{\lambda^{j,i}}{\gamma_i} R^j(\mathbf{x}^i)\right] + \sum_{i=1}^N u^i \mu^i(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \pi), \end{aligned}$$

where

$$\mu^i(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \pi) = \begin{cases} \sum_{j>i} \lambda^{i,j} - \sum_{j<i} \lambda^{j,i} - \pi\gamma_i & \text{if } i < N \\ \lambda^N - \sum_{j<N} \lambda^{j,N} - \pi\gamma_N & \text{if } i = N \end{cases}.$$

Then $\{\mathbf{X}, \mathbf{u}\}$ solve $\overline{\mathcal{P}}^N$ if and only if there exists $\boldsymbol{\lambda} \geq \mathbf{0}$ such that

$$L(\mathbf{X}, \mathbf{u}; \boldsymbol{\lambda}') \geq L(\mathbf{X}, \mathbf{u}; \boldsymbol{\lambda}) \geq L(\mathbf{X}', \mathbf{u}'; \boldsymbol{\lambda})$$

for all $\{\mathbf{X}', \mathbf{u}'\} \in \mathcal{Y}$, $\boldsymbol{\lambda}' \geq \mathbf{0}$. The second inequality is equivalent to

$$\mathbf{x}^i \in \arg\max_{\mathbf{x}\in\mathcal{M}} W^i(\mathbf{x}) - \sum_{j<i} \frac{\lambda^{j,i}}{\gamma_i} R^j(\mathbf{x}) \tag{22}$$

and

$$u^i \in \arg\max_{u\in\mathbb{R}} \mu^i(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \pi)\, u. \tag{23}$$

The first inequality is equivalent to the complementary slackness conditions

$$-u^N \leq 0 \quad \text{and} \quad \lambda^N u^N = 0 \ , \tag{24}$$

and for every $i \in N$, $j > i$

$$R^i(\mathbf{x}^j) + u^j - u^i \leq 0 \quad \text{and} \quad \lambda^{i,j}\left[R^i(\mathbf{x}^j) + u^j - u^i\right] = 0. \tag{25}$$

Lemma 20 *If $(\mathbf{X}, \mathbf{u}, \boldsymbol{\lambda})$ satisfies the conditions (22)-(25), then $\mu^i(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \pi) = 0$ for all $i \in N$.*

Proof. Recall that $u^i \geq 0$ for all $i \in N$, by combining *IR*^{N} and *R-IC*^{i,N}. Therefore, $\mu^i(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \pi) \geq 0$ for all $i \in N$. On the other hand, $(1-\pi)W(\mathbf{X}) + \pi\Pi(\mathbf{X}, \mathbf{u})$ is bounded below by $a^{\text{nf}}E(s) + d(a^{\text{nf}}) - c\left(a^{\text{nf}}\right) > 0$, hence $\mu^i(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \pi) \leq 0$. ■

This lemma has two immediate implications about the binding constraints.

Corollary 2 *If $\pi = 0$, then $\boldsymbol{\lambda} = \mathbf{0}$. If $\pi > 0$, IR^N binds and, for every $i < N$, there is $j > i$ such that R-IC^{i,j} binds.*

Proof. The second part follows from the last Lemma. For the first part, note that, since $\mu^i(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \pi) = 0$ for all $i = 1, \ldots, N$,

$$0 = \sum_{i=1}^{N} \mu^i(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \pi) = \sum_{i=1}^{N-1}\left[\sum_{j>i}\lambda^{i,j} - \sum_{j<i}\lambda^{j,i}\right] + \lambda^N - \sum_{j<N}\lambda^{j,N} - \pi = \lambda^N - \pi,$$

where the last equality follows from re-arraging $\sum_{i=1}^{N-1}\sum_{j>i}\lambda^{i,j}$. So, if $\pi = 0 = \lambda^N$, then $\mu^N(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \pi) = 0$ implies $\sum_{j<N}\lambda^{j,N} = 0$. Hence, $\lambda^{j,N} = 0$ for all $j < N$. Suppose for all $j \geq i+1$, $\lambda^{n,j} = 0$ for all $n < j$. Then, by $\mu^i(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \pi) = 0$, we have $\sum_{j<i}\lambda^{j,i} = \sum_{j>i}\lambda^{i,j} = 0$. Hence, $\lambda^{j,i} = 0$ for all $j < i$. ■

Therefore, although $\mu^i(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \pi) = 0$ makes any $u^i \in \mathbb{R}$ solve (23), we can use the upward binding constraints to pin down $\mathbf{u}$, once $\mathbf{X}$ has been chosen.

Thus, a solution to the $\overline{\mathcal{P}}^N$ exists if we can find $(\mathbf{X}, \boldsymbol{\lambda})$ so that, for every $i = 1, \ldots, N$, $\mathbf{x}^i$ solves (22), $\mu^i(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \pi) = 0$, and (24) and (25) hold. By replicating the arguments in the proof of Proposition 2 (see Step 5 below), we have that, for every $\boldsymbol{\lambda} \geq \mathbf{0}$, a solution $\mathbf{x}^i$ to (22) always exists unique over $(\underline{\theta}^i, \overline{\theta}^i)$ and is pointwise continuous in $\boldsymbol{\lambda}$. Furthermore, if $\lambda^{j,i} \to +\infty$ for some $j < i$, then $\mathbf{x}^i \to a^{\text{nf}}$ over $(\underline{\theta}^j, \overline{\theta}^i)$, i.e., $R^j(\mathbf{x}^i) \to 0$. Moreover, since $\mu^i(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \pi) = 0$, $\lambda^{i,j'} \to +\infty$ for some $j' > i$, so that $R^i\left(\mathbf{x}^{j'}\right) \to 0$ and $u^i \to 0$ (using the binding *R-IC*$^{i,j'}$). Therefore, there exists $\lambda^{j,i}$ large enough to make (25) hold. Finally, (24) always hold with $u^N = 0$.

Step 5: Characterization of expected virtual surplus and existence of a solution.

Fix $i > 1$. Then, by (22), $\mathbf{x}^i$ must solve

$$\max_{\mathbf{x}\in\mathcal{M}} W^i(\mathbf{x}) - \sum_{n=1}^{i-1} \lambda^{n,i} R^n(\mathbf{x}),$$

where $\boldsymbol{\lambda}^i \in \mathbb{R}_+^{i-1}$. Now, using the expression of $R^n(\mathbf{x}^i)$, and letting $\xi(\cdot) = c(\cdot) - d(\cdot)$, we have

$$\begin{aligned} W^i\left(\mathbf{x}^i\right) - \sum_{n=1}^{i-1} \lambda^{n,i} R^n\left(\mathbf{x}^i\right) &= \sum_{n=1}^{i-1} \lambda^{n,i} \int_{\underline{\theta}^n}^{\theta^i} \mathbf{x}^i(\theta) G^n(\theta)\, d\theta + \int_{\underline{\theta}^i}^{\overline{\theta}^i} [\mathbf{x}^i(\theta) w^i\left(\theta, \boldsymbol{\lambda}^i\right) - \xi(\mathbf{x}^i(\theta))] dF^i \\ &: = VS^i\left(\mathbf{x}^i; \boldsymbol{\lambda}^i\right), \end{aligned}$$

where

$$w^i\left(\theta; \boldsymbol{\lambda}^i\right) := \frac{\theta}{\phi^i} + \sum_{n=1}^{i-1} \lambda^{n,i} V^{i,n}(\theta) = \frac{\theta}{\phi^i} + \sum_{n=1}^{i-1} \lambda^{n,i} v^i(\theta) - \sum_{n=1}^{i-1} \lambda^{n,i} \frac{f^n(\theta)}{f^i(\theta)} v^n(\theta).$$

We can now apply to $VS^i\left(\mathbf{x}^i; \boldsymbol{\lambda}^i\right)$ the technique used in the two-type case to characterize $\mathbf{x}^I$ (Proposition 3). Clearly, if $\boldsymbol{\lambda}^i = \mathbf{0}$, $VS^i(\mathbf{x}^i; \mathbf{0}) = W^i(\mathbf{x}^i)$ and $\mathbf{x}^i = \mathbf{x}^{i*}$ on Θ^i. For $\theta < \underline{\theta}^i$, we can let $\mathbf{x}^i(\theta) = \mathbf{x}^i(\underline{\theta}^i)$. For $\theta > \overline{\theta}^i$, $\mathbf{x}^i(\theta)$ may be strictly larger than $\mathbf{x}^i(\overline{\theta}^i)$ to ensure that the downward constraints involving i as the mimicked type are satisfied.

Now, suppose $\lambda^{n,i} > 0$ for some $n < i$. Apply the Myerson-Toikka ironing technique on Θ^i, by letting

$$z^i\left(q; \boldsymbol{\lambda}^i\right) = w^i\left((F^i)^{-1}(q); \boldsymbol{\lambda}^i\right) \text{ and } Z^i\left(q; \boldsymbol{\lambda}^i\right) = \int_0^q z^i\left(y; \boldsymbol{\lambda}^i\right) dy.$$

Let $\Omega^i\left(q; \boldsymbol{\lambda}^i\right) = conv Z^i\left(q; \boldsymbol{\lambda}^i\right)$, and $\omega^i\left(q; \boldsymbol{\lambda}^i\right) = \Omega_q^i\left(q; \boldsymbol{\lambda}^i\right)$ wherever defined. Extend ω^i by right continuity, and at 1 by left continuity. For ω^i to be continuous, it is sufficient to show that, if z^i is discontinuous of q, then z^i jumps down at q. To see this, note that w^i can be discontinuous only at points like $\overline{\theta}^j$ for $j < i$ and such that $\overline{\theta}^j \in (\underline{\theta}^i, \overline{\theta}^i)$. At such a point, we have

$$w^i(\overline{\theta}^j -; \boldsymbol{\lambda}^i) = \lim_{\theta \uparrow \overline{\theta}^j} \frac{\theta}{\phi^i} + \sum_{n=1}^{i-1} \lambda^{n,i} v^i(\theta) - \sum_{n=1}^{i-1} \lambda^{n,i} \frac{f^n(\theta)}{f^i(\theta)} v^n(\theta).$$

Note that (1) for $n < j$, $\overline{\theta}^n < \overline{\theta}^j$, which implies that $f^n(\overline{\theta}^j) = 0$, and (2)

$$v^j(\overline{\theta}^j) = (\overline{\theta}^j/\phi^j) - \overline{\theta}^j + H^j(\overline{\theta}^j) = -(\phi^j - 1)(\overline{\theta}^j/\phi^j) < 0.$$

So,

$$w^i(\overline{\theta}^j -; \boldsymbol{\lambda}^i) = \frac{\overline{\theta}^j}{\phi^i} + \sum_{n=1}^{i-1} \lambda^{n,i} v^i(\overline{\theta}^j) - \sum_{n=j}^{i-1} \lambda^{n,i} \frac{f^n(\overline{\theta}^j)}{f^i(\overline{\theta}^j)} v^n(\overline{\theta}^j).$$

On the other hand,

$$\begin{aligned} w^i(\overline{\theta}^j +; \boldsymbol{\lambda}^i) &= \lim_{\theta \downarrow \overline{\theta}^j} \frac{\theta}{\phi^i} + \sum_{n=1}^{i-1} \lambda^{n,i} v^i(\theta) - \sum_{n=1}^{i-1} \lambda^{n,i} \frac{f^n(\theta)}{f^i(\theta)} v^n(\theta) \\ &= \frac{\overline{\theta}^j}{\phi^i} + \sum_{n=1}^{i-1} \lambda^{n,i} v^i(\overline{\theta}^j) - \sum_{n=j+1}^{i-1} \lambda^{n,i} \frac{f^n(\overline{\theta}^j)}{f^i(\overline{\theta}^j)} v^n(\overline{\theta}^j). \end{aligned}$$

Therefore,

$$w^i(\overline{\theta}^j-;\boldsymbol{\lambda}^i)-w^i(\overline{\theta}^j+;\boldsymbol{\lambda}^i)=-\lambda^{j,i}\frac{f^j(\overline{\theta}^j)}{f^i(\overline{\theta}^j)}v^j(\overline{\theta}^j)=\lambda^{j,i}\frac{f^j(\overline{\theta}^j)}{f^i(\overline{\theta}^j)}\frac{\phi^j-1}{\phi^j}\overline{\theta}^j\geq 0.$$

This proves that z^i can at most jump down, and, hence, ω^i is continuous. Letting $\overline{w}^i(\theta;\boldsymbol{\lambda}^i)=\omega^i(F^i(\theta);\boldsymbol{\lambda}^i)$ for $\theta\in\Theta^i$, construct the generalized virtual surplus $\overline{VS}^i$, as we did in the proof of Proposition 2.

Now, note that $G^n(\theta)>0$ for $\theta\in\left(\underline{\theta}^n,\underline{\theta}^i\right)$. Therefore, since $\lambda^{n,i}>0$ for some $n<i$, the first term in VS^i is strictly positive. Let $\underline{n}=\min\left\{n:\lambda^{n,i}>0\right\}$. Then, over $(\underline{\theta}^{\underline{n}},\overline{\theta}^i)$, the characterization of Lemma 13 of any maximizer to $\overline{VS}^I$ extends to $\overline{VS}^i$. In particular, $\mathbf{x}^i$ must be constant at a^i_{b} over $(\underline{\theta}^{\underline{n}},\theta^i_{\mathrm{b}})$, where $\theta^i_{\mathrm{b}}\geq\underline{\theta}^i$ and $a^i_{\mathrm{b}}\geq\overline{\mathbf{x}}^i(\underline{\theta}^i)$. Furthermore, $a^i_{\mathrm{b}}=\overline{\mathbf{x}}^i(\theta^i_{\mathrm{b}})$, if $\theta^i_{\mathrm{b}}<\overline{\theta}^i$; and $\mathbf{x}^i(\theta)=\overline{\mathbf{x}}^i(\theta)$ for $\theta\in[\theta^i_{\mathrm{b}},\overline{\theta}^i]$. Next, the same argument as in Lemma 14 for $\overline{VS}^I$ yields that a (unique) maximizer of $\overline{VS}^i$ exists as well. Finally, the same argument as in Lemma 15 implies that the (unique) maximizer of $\overline{VS}^i$ is also the (unique) maximizer of VS^i.

Step 6: Properties of the solutions to (22).

Suppose that $\lambda^{n,i}>0$ for some $n<i$ and let $\underline{n}$ be defined as before. Letting

$$K^i=\textstyle\sum_{n=\underline{n}}^{i-1}\lambda^{n,i}\int_{\underline{\theta}^n}^{\underline{\theta}^i}G^n(\theta)\,d\theta>0,$$

the analog of the ironing condition for θ_{b} applies to θ^i_{b}:

$$\textstyle\int_{\underline{\theta}^i}^{\theta^i_{\mathrm{b}}}\left[w^i\left(y;\boldsymbol{\lambda}^i\right)-w^i\left(\theta^i_{\mathrm{b}};\boldsymbol{\lambda}^i\right)\right]dF^i+K^i=0,$$

which implies $\theta^i_{\mathrm{b}}>\underline{\theta}^i$, and can be written as

$$\textstyle\int_{\underline{\theta}^i}^{\theta^i_{\mathrm{b}}}[w^i\left(\theta^i_{\mathrm{b}};\boldsymbol{\lambda}^i\right)-(\theta/\phi^i)]dF^i=\left[\int_{\underline{\theta}^i}^{\theta^i_{\mathrm{b}}}\sum_{n=1}^{i-1}\lambda^{n,i}V^{i,n}(\theta)\,dF^i+K^i\right].$$

To prove that $w^i\left(\theta^i_{\mathrm{b}};\boldsymbol{\lambda}^i\right)>\underline{\theta}^i/\phi^i$, it is enough to show that the right hand side is positive:

$$\int_{\underline{\theta}^i}^{\theta^i_{\mathrm{b}}}\sum_{n=1}^{i-1}\lambda^{n,i}V^{i,n}(\theta)\,dF^i+\sum_{n=\underline{n}}^{i-1}\lambda^{n,i}\int_{\underline{\theta}^n}^{\underline{\theta}^i}G^n(\theta)\,d\theta=\sum_{n=\underline{n}}^{i-1}\lambda^{n,i}\left[\int_{\underline{\theta}^i}^{\theta^i_{\mathrm{b}}}V^{i,n}(\theta)\,dF^i+\int_{\underline{\theta}^n}^{\underline{\theta}^i}G^n(\theta)\,d\theta\right]>0$$

which follows by (17). Therefore, $\mathbf{x}^i$ exhibits bunching over $[\underline{\theta}^{\underline{n}},\theta^i_{\mathrm{b}}]$ with $\theta^i_{\mathrm{b}}>\underline{\theta}^i$, and at value $a^i_{\mathrm{b}}>\mathbf{x}^{i*}(\underline{\theta}^i)$.

Now, consider the top of $[\underline{\theta}^i,\overline{\theta}^i]$. By the same argument as in Lemma 16, $\overline{w}^i(\overline{\theta}^i;\boldsymbol{\lambda}^i)\geq w^i(\overline{\theta}^i;\boldsymbol{\lambda}^i)$ with strict inequality if $\theta^{i\mathrm{b}}<\overline{\theta}^i$. Furthermore, for $\theta>\overline{\theta}^{i-1}$, $w^i\left(\theta,\boldsymbol{\lambda}^i\right)=\theta/\phi^i+\sum_{n=1}^{i-1}\lambda^{n,i}v^i(\theta)$ and $w^i(\overline{\theta}^i;\boldsymbol{\lambda}^i)=(\overline{\theta}^i/\phi^i)[1-\left(\phi^i-1\right)\sum_{n=1}^{i-1}\lambda^{n,i}]<(\overline{\theta}^i/\phi^i)$. So if $\overline{w}^i(\overline{\theta}^i;\boldsymbol{\lambda}^i)=w^i(\overline{\theta}^i;\boldsymbol{\lambda}^i)$, then $\mathbf{x}^i(\overline{\theta}^i;\boldsymbol{\lambda}^i)<\mathbf{x}^{i*}(\overline{\theta}^i)$. Otherwise, there is ironing at the top over $[\theta^{i\mathrm{b}},\overline{\theta}^i]\neq\varnothing$ and the following must hold

$$\textstyle\int_{\theta^{i\mathrm{b}}}^{\overline{\theta}^i}\left[w^i(y;\boldsymbol{\lambda}^i)-\overline{w}^i(\theta^{i\mathrm{b}};\boldsymbol{\lambda}^i)\right]dF^i=0,$$

which corresponds to

$$\int_{\theta^{i\mathrm{b}}}^{\overline{\theta}^i}\left[y/\phi^i-\overline{w}^i(\theta^{i\mathrm{b}};\boldsymbol{\lambda}^i)\right]dF^i=-\sum_{n=1}^{i-1}\lambda^{n,i}\int_{\theta^{i\mathrm{b}}}^{\overline{\theta}^i}V^{i,n}(y)\,dF^i.$$

Now, for each $n < i$,

$$\begin{aligned}
\textstyle\int_{\theta^{\text{ib}}}^{\overline{\theta}^i} V^{i,n}(y)\, dF^i &= \textstyle\int_{\theta^{\text{ib}}}^{\overline{\theta}^i} v^i(y)\, dF^i - \int_{\theta^{\text{ib}}}^{\overline{\theta}^i} v^n(y)\, dF^n \\
&= \textstyle\int_{\theta^{\text{ib}}}^{\overline{\theta}^i} \left(y/\phi^i - \theta^{\text{ib}}\right) dF^i - \int_{\theta^{\text{ib}}}^{\overline{\theta}^i} \left(y/\phi^n - \theta^{\text{ib}}\right) dF^n \\
&= \textstyle\int_{\theta^{\text{ib}}/\phi^i}^{\overline{s}} \left(s - \theta^{\text{ib}}\right) dF - \int_{\theta^{\text{ib}}/\phi^n}^{\overline{s}} \left(s - \theta^{\text{ib}}\right) dF = -\int_{\theta^{\text{ib}}/\phi^i}^{\theta^{\text{ib}}/\phi^n} \left(\theta^{\text{ib}} - s\right) dF < 0.
\end{aligned}$$

Therefore, $\overline{w}^i(\theta^{i\text{b}}; \boldsymbol{\lambda}^i) < \overline{\theta}^i/\phi^i$, and $\mathbf{x}^i(\overline{\theta}^i; \boldsymbol{\lambda}^i) < \mathbf{x}^{i*}(\overline{\theta}^i)$.

Finally, for bunching at the top, note that for $\theta > \overline{\theta}^{i-1}$,

$$w^i\left(\theta; \boldsymbol{\lambda}^i\right) = \theta/\phi^i + \textstyle\sum_{n=1}^{i-1} \lambda^{n,i}[(1-\phi^i)(\theta/\phi^i) + H^i(\theta)].$$

Hence, for $\theta' > \theta > \overline{\theta}^{i-1}$

$$w^i\left(\theta'; \boldsymbol{\lambda}^i\right) - w^i\left(\theta; \boldsymbol{\lambda}^i\right) = \frac{\theta' - \theta}{\phi^i}[1 - \sum_{n=1}^{i-1} \lambda^{n,i}\left(\phi^i - 1\right)] + \sum_{n=1}^{i-1} \lambda^{n,i}\left(H^i(\theta') - H^i(\theta)\right).$$

Under the Assumption 2, for $\theta' > \theta \geq \max\{\theta^{i\dagger}, \overline{\theta}^{i-1}\}$, we have

$$w^i\left(\theta'; \boldsymbol{\lambda}^i\right) - w^i\left(\theta; \boldsymbol{\lambda}^i\right) \leq \frac{\theta' - \theta}{\phi^i}[1 - \left(\phi^i - 1\right) \textstyle\sum_{n=1}^{i-1} \lambda^{n,i}].$$

Hence, bunching at the top may occur if $\sum_{n=1}^{i-1} \lambda^{n,i}$ is large enough, i.e., if the principal assigns high enough shadow value to *not* increasing the rents of the types below type i.

10 Appendix C: The Planner's Constrained Problem

I consider the alternative specification of the planner's problem as that of maximizing the expected ex-ante social surplus subject to a budget constraint. For simplicity, I do so in the two-type model. I suggest looking at Section 4 before reading this part.

Let B be the planner's budget. Assume that B is less than the maximal expected surplus: $B \leq W^i(\mathbf{x}^{i*})$. Using *ES* and $\Pi(\mathbf{X}, \mathbf{k})$, the planner's constrained problem is

$$\mathcal{PB} := \begin{cases} \max_{\{\mathbf{X},\mathbf{k}\}} \gamma W^C(\mathbf{x}^C) + (1-\gamma) W^I(\mathbf{x}^I) \\ \text{s.t. } \mathbf{x}^i \in \mathcal{M},\ IR^i, \text{ and } R\text{-}IC^i \text{ for } i = I, C, \text{ and} \\ \Pi(\mathbf{X}, \mathbf{k}) \geq B. \qquad\qquad (IR^B) \end{cases}$$

In $\mathcal{PB}$, given $\mathbf{X}$, multiple vectors $\mathbf{k}$ may be optimal. However, we can safely focus on DMs that make $R\text{-}IC^C$ and IR^I hold with equality. Then, let

$$\overline{\Pi}(\mathbf{x}^C, \mathbf{x}^I; B) := \gamma W^C(\mathbf{x}^C) + (1-\gamma)\left[W^I(\mathbf{x}^I) - \frac{\gamma}{1-\gamma} R^C(\mathbf{x}^I)\right] - B,$$

to reduce $\mathcal{PB}$ to

$$\mathcal{PB}' = \begin{cases} \max_{\mathbf{x}} \gamma W^C(\mathbf{x}^C) + (1-\gamma)\, W^I(\mathbf{x}^I) \\ \text{s.t. } \mathbf{x}^C, \mathbf{x}^I \in \mathcal{M},\ RR \text{ and} \\ \overline{\Pi}(\mathbf{x}^C, \mathbf{x}^I; B) \geq 0. \qquad\qquad (\overline{IR}^{\,B}) \end{cases}$$

Let $\mathcal{X}$ and $\mathcal{Y}$ be as in the proof of Lemma 4. By concavity of d and $-c$, $\gamma W^C(\mathbf{x}^C) + (1-\gamma)\, W^I(\mathbf{x}^I)$ and $\overline{\Pi}(\mathbf{x}^C, \mathbf{x}^I; B)$ are concave; $R^C(\mathbf{x}^I) + R^I(\mathbf{x}^C)$ is linear. The constraint functional $\Gamma(\mathbf{x}^C, \mathbf{x}^I) := (R^C(\mathbf{x}^I) + R^I(\mathbf{x}^C), -\overline{\Pi}(\mathbf{x}^C, \mathbf{x}^I; B))$ maps to $\mathbb{R}^2$, which has non-empty, positive, and closed cone $\mathbb{R}^2_+$. Let $\mu \geq 0$ and $\eta \geq 0$ and define the Lagrangian

$$\begin{aligned} L(\mathbf{x}^C, \mathbf{x}^I, \mu, \eta) : \quad &= \gamma W^C(\mathbf{x}^C) + (1-\gamma)\, W^I(\mathbf{x}^I) + \eta \overline{\Pi}(\mathbf{x}^C, \mathbf{x}^I; B) - \mu[R^C(\mathbf{x}^I) + R^I(\mathbf{x}^C)] \\ &= (1+\eta) \Big\{ \gamma [W^C(\mathbf{x}^C) - \tfrac{\mu}{\gamma(1+\eta)} R^I(\mathbf{x}^C)] + \\ &\quad + (1-\gamma)\, [W^I(\mathbf{x}^I) - \tfrac{\gamma\eta+\mu}{(1-\gamma)(1+\eta)} R^C(\mathbf{x}^I)] \Big\} - \eta B. \end{aligned} \tag{26}$$

Now, comparing (26) with (10), we see that here the weights on profits and therefore on C's rents depend endogenously on the multiplier $\eta \geq 0$, rather than on π. However, the same fundamental trade-offs arise in finding the optimal $\mathbf{x}^C$ and $\mathbf{x}^I$.

Specifically, by Theorem 2, p. 221, of Luenberger (1969), if $(\mathbf{x}^C, \mathbf{x}^I, \mu, \eta)$ is such that, for all $(\mathbf{x}_0^C, \mathbf{x}_0^I) \in \mathcal{Y}, \eta_0 \geq 0, \mu_0 \geq 0$,

$$L(\mathbf{x}^C, \mathbf{x}^I, \mu_0, \eta_0) \geq L(\mathbf{x}^C, \mathbf{x}^I, \mu, \eta) \geq L(\mathbf{x}_0^C, \mathbf{x}_0^I, \mu, \eta), \tag{27}$$

then $(\mathbf{x}^C, \mathbf{x}^I)$ solves $\mathcal{PB}'$. Condition (27) essentially implies the same optimality conditions as in Lemma 4. Also, if there exists $(\hat{\mathbf{x}}^C, \hat{\mathbf{x}}^I) \in \mathcal{Y}$ such that $\Gamma(\hat{\mathbf{x}}^C, \hat{\mathbf{x}}^I) < 0$, then by Corollary 1, p. 219, of Luenberger (1969), the condition (27) is also necessary for $(\mathbf{x}^C, \mathbf{x}^I)$ to solve $\mathcal{PB}'$. In particular, let B_1 be the profits that the optimal DM—derived in Section 4.2—yields in the monopolist's case ($\pi = 1$).

Lemma 21 *There exists $\mathbf{x}^C, \mathbf{x}^I \in \mathcal{M}$ such that $\Gamma(\mathbf{x}^C, \mathbf{x}^I) < 0$ if and only if $B < B_1$.*

Proof. ($\Rightarrow$) By definition of B_1, if $\mathbf{x}^C, \mathbf{x}^I \in \mathcal{M}$ and RR holds, then $\overline{\Pi}(\mathbf{x}^C, \mathbf{x}^I; B_1) \leq 0$.
($\Leftarrow$) Consider a solution $(\mathbf{x}_1^C, \mathbf{x}_1^I)$ to the problem $\mathcal{P}$ with $\pi = 1$. Since $B < B_1$, $\overline{\Pi}(\mathbf{x}_1^C, \mathbf{x}_1^I; B) > 0$. If $R^C(\mathbf{x}_1^I) + R^I(\mathbf{x}_1^C) < 0$, we are done. Suppose $R^C(\mathbf{x}_1^I) + R^I(\mathbf{x}_1^C) = 0$. Recall that $\mathbf{x}_1^I = \mathbf{x}^I(\hat{\rho}^C) = \arg\max_{\mathbf{x} \in \mathcal{M}} W^I(\mathbf{x}) - \hat{\rho}^C R^C(\mathbf{x})$ with $\hat{\rho}^C > 0$ (Lemma 4). By revealed optimality $R^C(\mathbf{x}^I(\hat{\rho}^C)) \geq R^C(\mathbf{x}^I(\rho^C))$ for every $\rho^C > \hat{\rho}^C$. Because RR holds with equality, $R^C(\mathbf{x}^I(\hat{\rho}^C)) > 0$ and $\mathbf{x}^I(\hat{\rho}^C)$ is not constant (Proposition 1). I claim that there exists $\tilde{\rho}^C > \hat{\rho}^C$ such that $R^C(\mathbf{x}^I(\hat{\rho}^C)) > R^C(\mathbf{x}^I(\tilde{\rho}^C))$: by Proposition 2 $R^C(\mathbf{x}^I(\rho^C))$ is continuous in ρ^C, and $\lim_{\rho^C \to +\infty} R^C(\mathbf{x}^I(\rho^C)) = 0$ (Proposition 1). Finally, since also $\overline{\Pi}$ is continuous in ρ^C, it remains to choose $\rho_\varepsilon^C > \hat{\rho}^C$ so that $R^C(\mathbf{x}^I(\rho_\varepsilon^C)) = R^C(\mathbf{x}^I(\hat{\rho}^C)) - \varepsilon$ and $\overline{\Pi}(\mathbf{x}^I(\rho_\varepsilon^C), \mathbf{x}_1^C; B) > 0$. ■

References

[1] Akerlof, G. A. (1991): "Procrastination and Obedience", *American Economic Review (Papers and Proceedings)*, 81, 2, 1-19.

[2] Amador, M., Werning, I. and Angeletos, G.M. (2006): "Commitment vs. Flexibility", *Econometrica*, 74, 2, 365-396.

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

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

[5] Ashraf, N., N. Gons, D. S. Karlan and W. Yin, (2003): "A Review of Commitment Savings Products in Developing Countries", ERD Working Paper No. 45.

[6] 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.

[7] Bernheim, D. B. (2002): "Taxation and Savings", in Handbook of Public Economics, Volume 3, Edited by A. J. Auerbach and M. Feldstein.

[8] Bryan, G. and Karlan, D. and Nelson, S. (2010): "Commitment Devices", *Annual Review of Economics*.

[9] Courty, P. and Li, H. (2000): "Sequential Screening", *Review of Economic Studies*, 67, 4, 697–717.

[10] DellaVigna, S. (2009): "Psychology and economics: Evidence from the field", *Journal of Economic Literature,* 47,2, 315-372.

[11] DellaVigna, S. and Malmendier, U., (2004): "Contract Design And Self-Control: Theory And Evidence", *Quarterly Journal of Economics,* 119, 2, 353-402.

[12] DellaVigna, S. and Malmendier, U, (2006): "Paying not to go to the gym", *American Economic Review*, 96, 3, 694-719.

[13] Diamond, P. A. (1977): "A Framework for Social Security Analysis", *Journal of Public Economics*, 8, 275-298.

[14] Eliaz, K. and Spiegler, R., (2006): "Contracting with diversely naive agents", *Review of Economic Studies*, 73, 3, 689-714.

[15] Eliaz, K. and Spiegler, R., (2008): "Consumer optimism and price discrimination", *Theoretical Economics*, 3, 4, 459–497.

[16] Esteban, S. and Miyagawa, E., (2006): "Temptation, self-control, and competitive nonlinear pricing", *Economics Letters*, 90, 3, 348-355.

[17] Esteban, S. and Miyagawa, E., (2005): "Optimal menu of menus with self-control preferences", Mimeo, Universitat Autònoma de Barcelona.

[18] Esteban, S., Miyagawa, E. and Shum, M. (2007): "Nonlinear pricing with self-control preferences", *Journal of Economic Theory*, 135, 1, 306-338.

[19] Gul, F, and W. Pesendorfer (2001): "Temptation and Self-Control," *Econometrica*, 69, 1403-1435.

[20] Heidhues, P. and Koszegi, B. (2010): "Exploiting naivete about self-control in the credit market", *American Economic Review,* Forthcoming.

[21] Holden, S., Ireland, K., Leonard-Chambers, V., and Bodgan, M. (2005): "The Individual Retirement Account at Age 30: A Retrospective", Investment Company Institute, 11, 1.

[22] Holden, S., Schrass, D. (2008): "The Role of IRAs in U.S. Households' Saving for Retirement, 2008", Investment Company Institute, 18, 1.

[23] Holden, S., Schrass, D. (2009): "The Role of IRAs in U.S. Households' Saving for Retirement, 2009", Investment Company Institute, 19, 1.

[24] Holden, S., Schrass, D. (2010a): "The Role of IRAs in U.S. Households' Saving for Retirement, 2010", Investment Company Institute, 19, 8.

[25] Holden, S., Sabelhaus, J., and Bass, S. (2010b): "The IRA Investor Profile. Traditional IRA Investors' Contribution Activity, 2007 and 2008", Investment Company Institute.

[26] Investment Company Institute (2012): "The U.S. Retirement Market, Second Quarter 2012" (September). http://www.ici.org/info/ret_12_q2_data.xls

[27] Jianye, Yan (2011): "Contracting with Heterogeneous Quasi-Hyperbolic Agents: Investment Good Pricing under Asymmetric Information", Mimeo, Toulouse School of Economics.

[28] Kreps, D. (1979): "A Representation Theorem for Preference for Flexibility", *Econometrica*, 47, 565-577.

[29] Laibson, D. (1998): "Life-Cycle Consumption and Hyperbolic Discount Functions", *European Economic Review*, 42, 861-871.

[30] Luenberger, D. G. (1969): "Optimization by Vector Space Methods". New York: Wiley.

[31] Munnell, A. H., Sundén, A. E. (2006): "401(k) Plans Are Still Coming Up Short", Brookings Institute Press.

[32] Mussa, M., and Rosen, S. (1978): "Monopoly and Product Quality", *Journal of Economic Theory*, 18, 2, 301-317.

[33] Myerson, R. B. (1981): "Optimal Auction Design", *Mathematics of Operation Research*, 6, 1, 58-73.

[34] Myerson, R. B. (1986): "Multi Stage Games with Communication", *Econometrica*, 54, 2, 323-358.

[35] O'Donoghue, T. and Rabin, M. (1999): "Doing It Now or Later", *American Economic Review*, 89, 1, 103-124.

[36] O'Donoghue, T. and Rabin, M. (1999): "Incentives for Procrastinators", *Quarterly Journal of Economics*, 114, 3, 769-816.

[37] O'Donoghue, T. and Rabin, M. (2001): "Choice and Procrastination", *Quarterly Journal of Economics*, 116, 1, 121-160.

[38] O'Donoghue, T. and Rabin, M. (2003): "Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes", *American Economic Review*, 93, 2, 186-191.

[39] Pavan, A. (2007): "Long-Term Contracting in a Changing World", Mimeo. Northwestern University.

[40] Pavan, A., Segal, I., and J. Toikka: "Dynamic Mechanism Design: Incentive Compatibility, Profit Maximization, and Information Disclosure." Mimeo, Northwestern University.

[41] Rockafellar, R. T. (1970): "Convex Analysis", Princeton University Press.

[42] Royden, H. L. (1988): "Real Analysis", Third Edition, Prentice Hall.

[43] Sandroni, A. and F. Squintani (2007): "Overconfidence, Insurance, and Paternalism", *American Economic Review*, 97, 5, 1994-2004.

[44] Spiegler, R. (2011): "Bounded Rationality and Industrial Organization," Oxford University Press.

[45] Strotz, R. H. (1956): "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies*, 23, 165-180.

[46] Thaler, R. H. (1994): "Psychology and Savings Policies", *American Economic Review*, 84, 2, 186-192.

[47] Toikka, J. (2011): "Ironing Without Control", *Journal of Economic Theory*, 146, 6, 2510-2526.

[48] U.S. government, (2010): "Estimates of The Federal Tax Expenditures for the Fiscal Years 2010-2014", Washington.