Commitment contracts*

Philip Bond[†] Gustav Sigurdsson[‡]

April 14, 2011

Abstract

We analyze a consumption-saving problem with unverifiable savings in which time-inconsistent preferences generate demand for commitment, but uncertainty about future consumption needs generates demand for flexibility. We characterize in a standard contracting framework the circumstances under which this combination is possible, in the sense that a *commitment contract* exists that implements the desired state-contingent consumption plan, thus offering both commitment and flexibility. Although we do not impose any restrictions on the set of possible contracts, we show that commitment contracts—when they exist—take the familiar form of a long-term savings account that permits some early withdrawals without a penalty, but imposes a penalty on any further withdrawals. With such an account, an individual is deterred from making premature withdrawals by the prospect of future selves withdrawing even more at the penalty rate. The key insight is that time-inconsistent preferences turn a single individual into a collection of selves with different preferences but the same information, effectively turning a single-agent contracting problem into a multi-agent mechanism design problem.

*We thank Andres Almazan, Greg Fischer, David Laibson, Dilip Mookherjee, Andrew Postlewaite, Debraj Ray, and Andrzej Skrzypacz for their helpful comments. We also thank seminar and conference audiences at UCLA, El Colegio de Mexico, the London School of Economics, the University of North Carolina, the University of Pennsylvania, the Stanford Institute for Theoretical Economics, the second Theory Workshop on Corporate Finance and Financial Markets at New York University, the 2010 Annual Meeting of the Western Finance Association, and the Econometric Society's 2010 Latin American Workshop in Economic Theory.

[†]University of Minnesota.

[‡]University of Pennsylvania.

1 Introduction

Preferences with hyperbolic time discounting, introduced by Strotz (1956),[1] are increasingly used to model individual behavior in a wide variety of settings such as consumer finance (e.g., Laibson 1996 on savings behavior in general; Laibson, Repetto, and Tobacman 1998 on retirement planning; DellaVigna and Malmendier 2004 and Shui and Ausubel 2004 on credit card usage; Skiba and Tobacman 2008 on payday lending; and Jackson 1986 on bankruptcy law), asset pricing (e.g., Luttmer and Mariotti 2003), and procrastination (e.g., O'Donoghue and Rabin 1999a, 1999b, 2001). In his original article, Strotz observed that hyperbolic discounting generates demand for commitment.[2] While some subsequent papers have analyzed the effectiveness of particular commitment devices (e.g., Laibson 1997 on illiquid assets such as housing wealth), very few papers have analyzed the extent to which commitment is possible without imposing exogenous restrictions on the particular form of commitment device.[3]

In this paper, we derive a necessary and sufficient condition under which commitment is possible in a consumption-saving problem with hyperbolic discounting, uncertainty about future consumption needs, and unverifiable savings. In our setting, an individual would like to commit at date 0 to a consumption plan that may depend on an unverifiable shock that is realized at date 1. To this end, the individual can enter into a *commitment contract* with a counterparty such as a bank that implements self 0's[4] desired consumption plan. The key contracting difficulty is that the shock is realized only at date 1, after the contract is signed, and since it is unverifiable, the contract cannot directly condition the individual's consumption on its realization. Rather, a commitment contract must provide the individual both with flexibility to respond to the shock, and with incentives to adhere to self 0's desired consumption plan.

Our results establish that commitment is often possible. Although we do not impose any restrictions on the set of possible contracts, our results also establish that commitment contracts take the familiar form of a long-term savings account that permits some early withdrawals without a penalty, but imposes a penalty on any further withdrawals. With such an account, an individual is deterred from making premature withdrawals by the prospect of future selves withdrawing even

[1] See Frederick, Loewenstein and O'Donoghue (2002) for a review of models of time discounting.

[2] See Ariely and Wertenbroch (2002) for direct evidence of demand for commitment.

[3] See O'Donoghue and Rabin (1999b), Della Vigna and Malmendier (2004), and Amador, Werning, and Angeletos (2006) for models of contracting with a hyperbolic individual. We discuss our relation to these papers in detail below.

[4] We follow the literature and refer to the individual at date t as *self* t.

more at the penalty rate.

The key insight of our paper is that time-inconsistent preferences are not only the source of the individual's commitment problem, but also crucial to its possible solution. With time-inconsistent preferences, the individual's different selves have different preferences but still share knowledge of the shock's realization. This opens up the possibility of later selves punishing prior selves for deviating from self 0's desired consumption plan, which would be impossible if their preferences were the same. In essence, time-inconsistent preferences turn a single-agent contracting problem into a multi-agent mechanism design problem. As is well known from the implementation theory literature,[5] this can dramatically expand the set of outcomes that are attainable in equilibrium.

1.1 An illustrative example

At date 0, a retired individual with wealth of $3\frac{1}{2}$ knows that he has three periods to live. He also knows that his child will get married either at date 1 or at date 2, at which time he will incur an expense of $\frac{1}{2}$. The individual has log preferences over consumption at each date and his time preferences are quasi-hyperbolic, with a hyperbolic discount factor of $\beta = \frac{1}{2}$ and no regular time discounting.[6] He would like to commit to self 0's most preferred consumption plan: consume exactly 1 at each future date, which requires that $1\frac{1}{2}$ be available to him at date 1, in case the wedding is then. The problem, of course, is that if $1\frac{1}{2}$ is available at date 1, then hyperbolic discounting will lead self 1 to consume $1\frac{1}{2}$, even if the wedding is not at date 1.[7]

Suppose, however, that the individual arranges his financial affairs as follows. At date 0 he deposits 1 each in one- and two-period savings accounts and $1\frac{1}{2}$ in a three-period savings account. Early withdrawals from the one- and two-period savings accounts are not permitted, but $\frac{1}{2}$ can be withdrawn without penalty from the three-period account at either date 1 or date 2. Furthermore, an additional $\frac{1}{4}$ can be withdrawn at date 2, but this second withdrawal carries a penalty of $\frac{1}{4}$.

This arrangement allows the individual to commit to self 0's desired consumption plan of consuming 1 at each date and also paying for the wedding. To see why, first consider self 2's incentives. If self 1 withdraws early, then the penalty ensures that self 2 will withdraw early if and only if he needs the money, i.e., if the wedding is in fact at date 2.[8] As a result, selves 2 and 3 consume

[5] See Maskin and Sjöström (2002), Palfrey (2002), and Serrano (2004) for surveys of implementation theory.

[6] Formally, his utility at date 0 is $\ln(c_1) + \ln(c_2) + \ln(c_3)$, at date 1 it is $\ln(c_1) + \beta \ln(c_2) + \beta \ln(c_3)$, and at date 2 it is $\ln(c_2) + \beta \ln(c_3)$, where $\beta = \frac{1}{2}$.

[7] Formally, $\ln(1\frac{1}{2}) + \frac{1}{2}\ln(1 - \frac{1}{2}) + \frac{1}{2}\ln(1) = \ln(\frac{3}{2}\sqrt{\frac{1}{2}}) > 0 = \ln(1) + \frac{1}{2}\ln(1) + \frac{1}{2}\ln(1)$.

[8] Formally, $\ln(1 + \frac{1}{4} - \frac{1}{2}) + \frac{1}{2}\ln(1\frac{1}{2} - \frac{1}{2} - \frac{1}{2}) = \ln(\frac{3}{4}\sqrt{\frac{1}{2}}) > \ln(1 - \frac{1}{2}) + \frac{1}{2}\ln(1\frac{1}{2} - \frac{1}{2})$ (wedding is at date 2) and

1 each if self 1 withdraws early to finance the wedding, but $\frac{3}{4}$ and $\frac{1}{2}$ if self 1 withdraws early to overconsume. If instead self 1 does not withdraw early, it is easy to verify that self 2 will, regardless of when the wedding actually occurs.[9]

Next consider self 1's incentives. If the wedding is not at date 1, then self 1 understands that if he were to withdraw early, then self 2 would also withdraw early—this time at a penalty—to finance the wedding at date 2, leaving self 3 with very little consumption. This outcome is unattractive enough to deter self 1 from withdrawing early if the wedding is not at date 1.[10] If, however, the wedding is in fact at date 1, then self 1 will withdraw early the $\frac{1}{2}$ he needs, secure in the knowledge that self 2 does not need to finance the wedding and that the penalty deters self 2 from withdrawing early for extra consumption.[11]

1.2 Costly excess flexibility and preference reversal

The example illustrates a commitment contract's most important features. First, the contract gives the individual *excess flexibility* to consume early, in the form of the second withdrawal option. In a world with full commitment, this flexibility would not be needed, as the first withdrawal option gives the individual sufficient flexibility to respond to the shock. In our setting, however, the second withdrawal option enables self 2 to "punish" self 1 for overconsuming. Loosely speaking, the contract puts the individual in a position where he realizes that if he "slips" at date 1 and overconsumes, then he will "fall off the wagon" at date 2 and overconsume even more.

Second, the excess flexibility is *costly*, in the sense that the second withdrawal option carries a penalty. We elaborate on this point in the next subsection.

The example also illustrates how the possibility of commitment is determined by the nature of the shock. To ensure that self 2 withdraws early (thereby punishing self 1) if and only if self 1's early withdrawal is a deviation, self 2's preferences cannot be the same in both states. Specifically, self 2's desire to impose the punishment must be greater when the wedding actually is at date 2 (and therefore self 1's early withdrawal is a deviation) than when it is at date 1. This is indeed the case in our example: because of the wedding expense, self 2's marginal utility of withdrawing

$\ln(1) + \frac{1}{2}\ln(1\frac{1}{2} - \frac{1}{2}) = 0 > \ln(\frac{5}{4}\sqrt{\frac{1}{2}}) = \ln(1 + \frac{1}{4}) + \frac{1}{2}\ln(1\frac{1}{2} - \frac{1}{2} - \frac{1}{2})$ (wedding is at date 1).

[9] Formally, $\ln(1 + \frac{1}{2}) + \frac{1}{2}\ln(1\frac{1}{2} - \frac{1}{2}) = \ln(\frac{3}{2}) > \ln(\sqrt{\frac{3}{2}}) = \ln(1) + \frac{1}{2}\ln(1\frac{1}{2})$ (wedding is at date 1) and $\ln(1 + \frac{1}{2} - \frac{1}{2}) + \frac{1}{2}\ln(1\frac{1}{2} - \frac{1}{2}) = 0 > \ln(\frac{1}{2}\sqrt{\frac{3}{2}}) = \ln(1 - \frac{1}{2}) + \frac{1}{2}\ln(1\frac{1}{2})$ (wedding is at date 2). Moreover, self 2 never wants to simultaneously make both the penalty-free and the penalty withdrawal. Formally, $\ln\left(1 + \frac{1}{2}\right) + \frac{1}{2}\ln\left(1\frac{1}{2} - \frac{1}{2}\right) > \ln\left(1 + \frac{1}{2} + \frac{1}{4}\right) + \frac{1}{2}\ln\left(1\frac{1}{2} - \frac{1}{2} - \frac{1}{2}\right)$ and $\ln\left(1 + \frac{1}{2} - \frac{1}{2}\right) + \frac{1}{2}\ln\left(1\frac{1}{2} - \frac{1}{2}\right) > \ln\left(1 + \frac{1}{2} - \frac{1}{2} + \frac{1}{4}\right) + \frac{1}{2}\ln\left(1\frac{1}{2} - \frac{1}{2} - \frac{1}{2}\right)$.

[10] Formally, $\ln(1) + \frac{1}{2}\ln(1 + \frac{1}{2} - \frac{1}{2}) + \frac{1}{2}\ln(1\frac{1}{2} - \frac{1}{2}) = 0 > \ln(\frac{3}{2}\sqrt{(\frac{3}{4}\frac{1}{2})}) = \ln(1 + \frac{1}{2}) + \frac{1}{2}\ln(\frac{3}{4}) + \frac{1}{2}\ln(\frac{1}{2})$.

[11] Formally, $\ln(1 + \frac{1}{2} - \frac{1}{2}) + \frac{1}{2}\ln(1) + \frac{1}{2}\ln(1) = 0 > \ln(\frac{1}{2}\sqrt{\frac{3}{2}}) = \ln(1 - \frac{1}{2}) + \frac{1}{2}\ln(1 + \frac{1}{2}) + \frac{1}{2}\ln(1\frac{1}{2} - \frac{1}{2})$.

early is strictly greater when the wedding is at date 2 than when it is at date 1. More generally, commitment is possible only under the following *preference reversal* condition: the state in which the individual's desire for date-1 withdrawals is higher must also be the state in which his desire for date-2 withdrawals is lower.

For an example in which our preference reversal condition is not satisfied, suppose instead that the wedding either takes place at date 1 or not at all. Moreover, suppose that the wedding is a lavish affair, requiring an expense of 1/2 at date 1 (for the ceremony) and another 1/2 at date 2 (for the banquet). Here, the individual would like to withdraw an additional 1/2 both at date 1 and at date 2 if the wedding takes place. Since the individual's desire for higher date-1 withdrawals coincides with his desire for higher date-2 withdrawals—both occur when the wedding takes place—our preference reversal condition is not satisfied and so commitment is not possible.

1.3 Unverifiable savings

In the examples above, the individual could not save from one date to the next. Most of our analysis, however, allows for the important possibility of unverifiable savings.[12] As we will show, a hyperbolic individual's ability to save actually makes it more difficult for him to commit, even though present-bias generally leads individuals to save too little.[13]

At the most basic level, self 2's ability to save restricts significantly the range of punishments that can be imposed on self 1 for deviating. Specifically, self 2 cannot be induced to impose a punishment that entails very high date-2 consumption relative to date-3 consumption, as he would always prefer to partially smooth his consumption across the two dates by saving. As a result, self 1 cannot be punished by imposing on him a combination of extremely low date-3 consumption and commensurately high date-2 consumption. One consequence of this restriction, and a key implication of our analysis, is that self 1 must instead be punished with a reduction in his total consumption. Consequently, the property of the commitment contract in the example that excess flexibility is costly is in fact a key property of *all* commitment contracts.

The individual's ability to commit is further limited by self 1's ability to save. Since the punishment must increase self 2's current consumption at the expense of his future consumption, he grows less willing to impose it as he inherits a higher level of savings from self 1. In essence,

[12] See the growing literature on contracting with hidden savings, e.g., Kocherlakota (2004), Doepke and Townsend (2006), and He (2009).

[13] See references on page 1.

self 1's ability to save undercuts commitment because it enables him to overconsume and then "bribe" self 2 to not impose the punishment. Consequently, a stronger form of preference reversal is required. Specifically, self 2's preference for imposing the punishment when self 1 deviates must be sufficiently strong to ensure that self 2 imposes it even when he inherits savings from self 1. At the same time, self 2 must refrain from imposing the punishment when self 1 does not deviate. We give a formal definition of this condition—*strong preference reversal*—in the text below. In the important special case of commitment to self 0's most preferred consumption plan, strong preference reversal is satisfied as long as the tempting deviation at date 1 does not increase total consumption over dates 1 and 2.[14] Finally, strong preference reversal is often sufficient as well as necessary for commitment.

1.4 Discussion

Our results show that in some settings, an individual can contract to completely overcome his commitment problem, even in the face of uncertainty about his future consumption needs. In these cases, hyperbolic discounting ceases to affect the individual's behavior. Moreover, the contracts that enable an individual to commit are often easy to interpret, as the first example above suggests: at date 0 the individual arranges access to a savings account with limited penalty-free early withdrawal rights, coupled with additional withdrawal rights that carry a penalty.[15] However, since there also exist important cases in which commitment is not attainable, our results should not be interpreted as an argument against the importance of hyperbolic discounting in general.

In this paper, we focus on one particular form of time-inconsistent preferences, namely the present-bias generated by hyperbolic discounting. However, our key insight—that time-inconsistent preferences turn a single-agent contracting problem into a multi-agent mechanism design problem—is more widely applicable. In particular, consider any source of time-inconsistent preferences that an individual is self-aware enough to anticipate. For example, an individual may understand today that, in the future, he will misinterpret the relevance of a small number of data points. Just as in the current setting, he can potentially commit to a course of action that avoids this bias, while at the same time maintaining flexibility to respond to shocks.

[14] See Section 6.
[15] See Section 7 for a detailed discussion.

2 Related literature

The closest antecedent to our paper is Amador, Werning, and Angeletos (2006). Like us, they study a hyperbolic individual who is hit by unverifiable taste shocks, but consider only a two-period version of the problem. This restriction immediately rules out the possibility of self 2 imposing a state-contingent punishment on self 1 for deviating—a key feature of our setting—because with two periods self 1 is effectively the only strategic agent.[16] Consequently, the only way to deter self 1 from deviating is to distort consumption in at least some states; the authors characterize the least costly way to do so.

Like Amador, Werning, and Angeletos, DellaVigna and Malmendier (2004) restrict attention to two periods, again ruling out the possibility of self 2 punishing self 1. Moreover, in their setting self 1 faces a binary choice (e.g., whether or not to go to the gym) and consequently a contract exists under which self 1 acts exactly as self 0 desires. The authors characterize the contract that maximizes the profits of a monopolist counterparty facing a partially naïve agent (see Section 8 for a discussion of partial naïveté). In particular, they characterize the combination of flat upfront fees and per-usage fees in the profit-maximizing contract.[17]

O'Donoghue and Rabin (1999b) analyze optimal contracts for procrastinators in a multi-period environment, where the socially efficient date at which a task should be performed is random. They explicitly rule out the use of contracts that induce an agent to reveal his type, which are the focus of our paper. As they observe, this restriction is without loss of generality in the main case they study, that of agents who are completely naïve about their future preferences. By contrast, we study sophisticated agents (again, see Section 8 for a discussion of partial naïveté).

While we examine the use of *external* commitment devices, such as contracts, other research considers what might be termed *internal* commitment devices. Bernheim, Ray, and Yeltekin (1999) and Krusell and Smith (2003) consider deterministic models in which an individual is infinitely lived, and show that Markov-perfect equilibria exist in which he gains some commitment ability from the fact that deviations will cause future selves to punish him. Carrillo and Mariotti (2000) and Benabou and Tirole (e.g., 2002, 2004) consider models in which an individual can commit his

[16] Amador, Werning, and Angeletos (2003) extend the analysis to three or more periods. They assume that shocks are independent across dates and affect only contemporaneous utility. Together, these assumptions rule out the possibility of a later self imposing a state-contingent punishment on a prior self for deviating.

[17] Similarly, Eliaz and Spiegler (2006) analyze profit maximization by a monopolist who deals with a population of time-inconsistent individuals who differ in their degree of sophistication (see Section 8 below).

future selves to some action by manipulating their beliefs, respectively, through the extent of his own information acquisition, through direct distortion of beliefs, or through self-signalling.

3 Model

A single agent consumes at three dates $t = 1, 2, 3$. His preferences are time-inconsistent, with self t's preferences represented by the quasi-hyperbolic separable utility function

$$U^t(\tilde{c}; \phi) = u_t(\tilde{c}_t; \phi) + \beta \sum_{s=t+1}^{3} u_s(\tilde{c}_s; \phi)$$

for $t = 0, 1, 2, 3$ where $\tilde{c} = (\tilde{c}_1, \tilde{c}_2, \tilde{c}_3) \in \mathbb{R}^3$ is the agent's *consumption* ($u_0 \equiv 0$ so date-0 consumption does not matter to the agent); $\phi \in \Phi \equiv \{\theta, \theta'\}$ is one of two[18] *states* that determine the agent's preferences; u_t is strictly increasing and strictly concave in $\tilde{c}_t$ for every t and every ϕ; and $\beta \in (0, 1)$ is the hyperbolic discount factor. There is no regular time discounting and the risk-free rate in the economy is zero. Finally, the agent is self-aware (i.e., sophisticated), in the sense that at each date, he correctly anticipates his preferences at future dates. In Section 8 we discuss how our analysis is affected if this assumption is relaxed.

At the center of our analysis is an arbitrary consumption plan $\{c(\phi)\}_\phi$, to which the agent would like to commit at date 0. In Section 6 we analyze further the special case where $\{c(\phi)\}_\phi$ is self 0's most preferred consumption plan. Since our main focus is on the effect of hyperbolic discounting on intertemporal efficiency, not its effect on insurance across states, we rule out transfers across states, so that $\sum_t c_t(\phi) \le W$ for $\phi = \theta, \theta'$, where W is the sum of the agent's initial endowment and verifiable future income. This assumption also facilitates comparison with the existing literature, which like us focuses on intertemporal efficiency.[19] Moreover, it would be hard—and sometimes impossible—to insure the agent if self 0 had private information about the relative probability of the two states.[20]

The state is revealed to the agent at date 1 and is unverifiable in the sense that no contract between the agent and a counterparty can be made directly contingent on the state. By contrast,

[18] Section 9 extends our model to three or more states.

[19] Amador, Werning, and Angeletos (2006) rule out transfers across states. O'Donoghue and Rabin (1999b) and DellaVigna and Malmendier (2004) study risk-neutral agents, and so insurance across states is not a concern.

[20] Note that private information about the relative probability of the two states would not affect our analysis, which characterizes when intertemporal efficiency is possible.

if instead the state were either revealed at date 0 or verifiable, then commitment could be easily attained by entering into a contract under which the agent gives a counterparty his entire endowment with instructions to return the endowment over time according to self 0's desired consumption plan.

A further constraint on the contracting environment is that the agent can privately save from one date to the next and his saving decisions are unverifiable. We also analyze the simpler case of verifiable saving decisions in Section 4.

Throughout, we assume that self 2's preferences satisfy the following standard single crossing property (see Milgrom and Shannon 1994).

Assumption (Single crossing) *Fix* ϕ, $\phi' \neq \phi$, $\tilde{c}$, $\tilde{c}^a$, *and* $\tilde{c}^b$ *such that* $\tilde{c}_2 < \tilde{c}_2^a \leq \tilde{c}_2^b$, $U^2(\tilde{c}^a;\phi) \geq U^2(\tilde{c};\phi) \geq U^2(\tilde{c}^b;\phi)$, *and* $U^2(\tilde{c};\phi') \geq U^2(\tilde{c}^a;\phi')$. Then $U^2(\tilde{c};\phi') \geq U^2(\tilde{c}^b;\phi')$.

In words, single crossing says that self 2's indifference curves in the two states can cross at most once. Two examples of functional forms for utility that satisfy single crossing are *multiplicative shocks*—$\phi \in \mathbb{R}_+^3$ and $u_t(\tilde{c}_t;\phi) = \phi_t u_t(\tilde{c}_t)$—and *additive shocks*—$\phi \in \mathbb{R}^3$ and $u_t(\tilde{c}_t;\phi) = u_t(\tilde{c}_t + \phi_t)$, as long as $\phi_3 = \phi_3'$. In addition, we make the mild assumption that if, at any point, self 2's indifference curves have the same slope in both states, then they coincide everywhere.[21] While both the multiplicative and additive parameterizations can be interpreted in a large number of ways, several interpretations deserve particular discussion:

1. Under the widely used parameterization of utility functions in which consumption and leisure enter multiplicatively, multiplicative shocks can be interpreted (among other ways) as shocks to time endowments. For example, if date t is a vacation day for the agent in state ϕ, then his marginal utility of consumption is high.

2. Additive shocks where $\phi_t \leq 0$ can be interpreted as essential expenditures. For example, if at date t an individual is sick in state ϕ but not in state ϕ', and must pay \$100 for treatment in state ϕ, then his utility from spending a total of $\tilde{c}_t$ in state ϕ is the same as from spending $\tilde{c}_t - 100$ in state ϕ'. The wedding example of the introduction entails shocks of this type.

3. Symmetrically, additive shocks where $\phi_t \geq 0$ can be interpreted as (unverifiable) increases

[21] We use this assumption only to establish equivalence of the two conditions stated in Proposition 1. It is satisfied for the two classes of preferences mentioned above.

in income. With slight abuse of language, we continue to refer to $\tilde{c}_t$ as consumption, even though the agent's true consumption in this case is $\tilde{c}_t + \phi_t$.

Finally, we assume throughout that the period utility functions satisfy an Inada condition in all states and at all dates. To reflect the two interpretations of additive shocks above, we allow the Inada condition to depend on the state: there exists $\underline{\tilde{c}}_t(\phi)$ such that $u_t'(\tilde{c}_t; \phi) \to \infty$ as $\tilde{c}_t \to \underline{\tilde{c}}_t(\phi)$,[22] and $u_t'(\tilde{c}_t; \phi) \to 0$ as $\tilde{c}_t \to \infty$.

Our model can be interpreted more broadly than a consumption-saving problem. In one example, the individual faces a *procrastination* problem, where W is his total endowment of leisure (e.g., the time he has left after completing a referee report) and $\tilde{c}_t$ is his leisure at date t. Shocks may be either multiplicative (in a favorite O'Donoghue and Rabin example, a Johnny Depp film is showing) or additive (the individual must take his child to the doctor). In another example, the individual is a *myopic manager*, where $\tilde{c}_t$ is investment at date t and u is a production function. In this case, hyperbolic discounting captures the manager's present-bias.

Our model has one shock realization and three periods. Adding more periods would only make it easier to commit to self 0's desired consumption plan: as the opening example demonstrates, self 2 must be induced to punish self 1 for deviating, and this is easier to arrange with more periods since then self 2 can in turn be punished for not punishing. In this sense, the assumption of just three periods biases our results against commitment being possible.

If instead the number of shock realizations increases with the number of periods, then the environment becomes considerably harder to analyze. Nonetheless, we believe that our basic insights and techniques remain valid. Moreover, in some circumstances a general T-period model can be analyzed by simply iterating our three-period model. In particular, consider the case where T is an odd number and an independent shock is realized at each of dates 1, 3, 5, ..., $T-2$, with the date-t shock affecting utility only at dates t and $t+1$. This environment can be analyzed recursively using results of our basic three-period model, by replacing $u_3(\tilde{c}_3; \phi)$ with the expected utility promised from date $t+2$ onwards.

[22] So in the two additive shock interpretations discussed, $\underline{\tilde{c}}_t(\phi) = -\phi_t$.

4 Commitment with verifiable savings

To build intuition, we analyze in this section the special case in which saving decisions are verifiable. We say that $\{c(\phi)\}_\phi$ *generates a commitment problem in state* ϕ if self 1, in state ϕ, prefers the consumption prescribed for state $\phi' \neq \phi$; formally, $U^1(c(\phi);\phi) < U^1(c(\phi');\phi)$. Without loss we restrict attention to consumption plans with the property that, in each state, self 0 prefers the consumption prescribed for that state to the consumption prescribed for the other state. Formally, $\{c(\phi)\}_\phi$ must satisfy $U^0(c(\phi);\phi) \geq U^0(c(\phi');\phi)$ for all $\phi, \phi' \in \Phi$. If $\{c(\phi)\}_\phi$ fails this condition, then it is dominated by a non-contingent consumption plan in which the individual gets $c(\phi')$ in both states.

As the following result shows, a commitment problem in state ϕ rules out a commitment problem in state $\phi' \neq \phi$ and can only arise when $c(\phi')$ offers strictly more date-1 consumption than $c(\phi)$.

Lemma 1 *If $\{c(\phi)\}_\phi$ generates a commitment problem in state ϕ, then $c_1(\phi') > c_1(\phi)$ for $\phi' \neq \phi$ and so, in particular, $\{c(\phi)\}_\phi$ does not generate a commitment problem in state ϕ'. Conversely, $\{c(\phi)\}_\phi$ generates a commitment problem in state ϕ if $c_1(\phi') > c_1(\phi)$ and β is sufficiently small.*

We assume, without loss of generality, that $c_1(\theta') \geq c_1(\theta)$; Lemma 1 then implies that if there is a commitment problem, it is in state θ. To simplify notation, we will often refer to $c(\theta)$ and $c(\theta')$ as c and c', respectively.

When does a contract exist that commits the agent to a consumption plan (c, c')? By standard revelation principle arguments (see Myerson 1981), we can restrict attention to a direct revelation mechanism that gives each of selves 1, 2, and 3 a menu of two consumption choices (corresponding to the two states), where each self's menu possibly depends on previous selves' choices. By a similar argument (see, e.g., Cole and Kocherlakota 2001), we can restrict attention to contracts that do not allow the agent to save. Finally, since date 3 is the final consumption date, self 3 will always choose the highest consumption level on his menu, regardless of the true state. As a result, we can restrict attention to contracts in which self 2's consumption choice dictates self 3's consumption.

Both c and c' must be among the consumption choices offered. Therefore, self 1's menu must include c_1 and c_1', and self 2's menu after self 1 chooses c_1 (respectively, c_1') must include (c_2, c_3) (respectively, (c_2', c_3')). As a result, self 1's menu must be the set $\{c_1, c_1'\}$ and self 2's menu must be the set $\{(c_2, c_3), (\hat{c}_2', \hat{c}_3')\}$ (if self 1 chooses c_1) or $\{(\hat{c}_2, \hat{c}_3), (c_2', c_3')\}$ (if self 1 chooses c_1'), where $(\hat{c}_2', \hat{c}_3')$

and $(\hat{c}_2, \hat{c}_3)$ are "punishments" chosen by self 2 if self 1 deviates from the desired consumption plan by choosing c_1 in state θ' or c_1' in state θ, respectively.

What are the incentive compatibility conditions? First, self 1 must be better off choosing c_1 in state θ, i.e., $U^1(c;\theta) \geq U^1(\hat{c};\theta)$, and c_1' in state θ', i.e., $U^1(c';\theta') \geq U^1(\hat{c}';\theta')$, where $\hat{c} = (c_1', \hat{c}_2, \hat{c}_3)$ and $\hat{c}' = (c_1, \hat{c}_2', \hat{c}_3')$ are the punishments he anticipates if he deviates in each respective case.

Second, self 2 must be better off choosing the punishment in each state if and only if self 1 deviates. For the punishment $\hat{c}$, intended for a deviation in state θ, this is the case if and only if $U^2(\hat{c};\theta) \geq U^2(c';\theta)$ and $U^2(c';\theta') \geq U^2(\hat{c};\theta')$. Similarly, for the punishment $\hat{c}'$, intended for a deviation in state θ', this is the case if and only if $U^2(c;\theta) \geq U^2(\hat{c}';\theta)$ and $U^2(\hat{c}';\theta') \geq U^2(c;\theta')$.

Unsurprisingly, when (c, c') does not generate a commitment problem in either state, the entire consumption decision can be delegated to self 1: $\hat{c} = c'$ and $\hat{c}' = c$ satisfy all the above constraints. Further note that even when (c, c') generates a commitment problem in state θ, there is no need for self 2 to punish self 1 for choosing c_1—the smaller of the two consumption choices—in state θ': $\hat{c}' = c$ satisfies all the above constraints in which $\hat{c}'$ appears. As a result, the interesting part of the contracting problem is the choice of $\hat{c}$, the punishment for choosing c_1' in state θ.

As the constraints above indicate, $\hat{c}$ must satisfy three somewhat conflicting criteria. The first two relate to preferences in state θ. First, $\hat{c}$ must be sufficiently unattractive to self 1, relative to c', to deter him from deviating. Second, $\hat{c}$ must be sufficiently attractive to self 2, relative to c', to induce him to punish. These two criteria are satisfied only if $\hat{c}_2 > c_2'$ and $\hat{c}_3 < c_3'$, i.e., if $\hat{c}$ is strictly *front-loaded* relative to c'. This follows from hyperbolic discounting: relative to consumption at date 3, self 2 values consumption at date 2 more than self 1 does. Therefore, if c' were instead front-loaded relative to $\hat{c}$ and self 1 preferred c' to $\hat{c}$, then self 2 would also prefer c' to $\hat{c}$ and would therefore not punish. We use this property, summarized in the following result, throughout the paper.

Lemma 2 *Fix ϕ, $\tilde{c}^a$, and $\tilde{c}^b$ such that $\tilde{c}_1^a = \tilde{c}_1^b$, $U^1(\tilde{c}^a;\phi) \geq U^1(\tilde{c}^b;\phi)$, and $U^2(\tilde{c}^a;\phi) \leq U^2(\tilde{c}^b;\phi)$. Then $\tilde{c}_2^a \leq \tilde{c}_2^b$ and $\tilde{c}_3^a \geq \tilde{c}_3^b$, both with strict inequality if either $U^1(\tilde{c}^a;\phi) > U^1(\tilde{c}^b;\phi)$ or $U^2(\tilde{c}^a;\phi) < U^2(\tilde{c}^b;\phi)$.*

The third criterion that $\hat{c}$ must satisfy relates to preferences in state θ'. Here, $\hat{c}$ must be sufficiently unattractive relative to c' to deter self 2 from punishing. Given that $\hat{c}$ must be strictly front-loaded relative to c' and self 2 must prefer $\hat{c}$ to c' in state θ, we have established that a particular form of *preference reversal* is a necessary condition for commitment.

Condition PR (Preference reversal) *There exists* $\tilde{c}$ *such that* $\tilde{c}_2 > c'_2$, $U^2(\tilde{c};\theta) \geq U^2(c';\theta)$, *and* $U^2(c';\theta') \geq U^2(\tilde{c};\theta')$.

A commitment problem arises because, relative to consumption at future dates, self 1 values current consumption *more* in state θ' than in state θ (and so $c'_1 > c_1$). Preference reversal (PR), however, says that relative to consumption at future dates, self 2 values current consumption *less* in state θ' than in state θ.[23] As the following result illustrates, PR is easily checked by comparing the slopes of self 2's indifference curves through c' in states θ and θ'.

Proposition 1 *When savings are verifiable, commitment to* (c, c') *is possible only if* c' *satisfies PR or, equivalently,* $U_2^2(c';\theta)/U_3^2(c';\theta) \geq U_2^2(c';\theta')/U_3^2(c';\theta')$.[24]

Finally, although PR is in general only a necessary condition for commitment to (c, c'), it is straightforward to show that if the period utility functions u_2 and u_3 are unbounded above and below, then PR is both sufficient and necessary.

5 Commitment with unverifiable savings

In the following sections we analyze the main case, in which the agent's saving decisions are unverifiable. The possibility of unverifiable savings is important because it significantly limits the set of consumption plans to which an individual can commit. Moreover, it is hard to imagine a contracting device that would render savings verifiable, as the individual can always just hold cash from one period to the next. We denote savings carried over from date t to date $t+1$ by $s_t \geq 0$ and, since consumption and withdrawals need not coincide when the agent can save, we denote withdrawals by x to prevent confusion between the two. With slight abuse of notation, we denote the consumption $(-s_1, s_1, 0) + x + (0, -s_2, s_2)$ by $s_1 + x - s_2$. The reader may find it useful to note that savings in this expression are written with respect to self 2, who is the key strategic actor.

[23] Condition PR may remind readers of Maskin's (1999) monotonicity condition. However, while PR may fail in our setting, monotonicity is trivially satisfied as long as some self's preferences differ across the two states. In our setting, the social choice rule of interest is $F(\phi) = c(\phi)$, where the domain of consumption choices is $\mathbb{R}^3$. This social choice rule is monotonic if and only if for all ϕ and $\phi' \neq \phi$, $U^t(c(\phi);\phi) \geq U^t(x;\phi)$ and $U^t(c(\phi);\phi') < U^t(x;\phi')$ for some self $t \in \{1,2,3\}$ (self 0 is non-strategic) and some $x \in \mathbb{R}^3$. As long as some self's preferences differ across the two states, this condition is satisfied.

[24] The discussion above establishes that PR is necessary for commitment. The equivalence of PR and the inequality is immediate from single crossing and our mild assumption that if, at any point, self 2's indifference curves have the same slope in both states, then they coincide everywhere.

As defined, single crossing only applies when the savings inherited from date 1 are the same across states. In the analysis that follows, however, we must repeatedly compare self 2's indifference curves when the savings inherited from date 1 differ across states. To this end, we extend single crossing:

Assumption SCB (Single crossing from below) *Fix* $s_1 \geq 0$, ϕ, $\phi' \neq \phi$, x, x^a, *and* x^b *such that* $x_2 < x_2^a \leq x_2^b$, $U^2(s_1 + x^a; \phi) \geq U^2(s_1 + x; \phi) \geq U^2(s_1 + x^b; \phi)$, *and* $U^2(x; \phi') \geq U^2(x^a; \phi')$. *Then* $U^2(x; \phi') \geq U^2(x^b; \phi')$.

In words, single crossing from below (SCB) says that once self 2's indifference curve in one state with no savings crosses his indifference curve in the other state with positive savings from below, they cannot cross again at higher levels of x_2. Whereas standard single crossing requires that indifference curves with the same savings cross only once, SCB allows indifference curves with different savings levels to cross twice. In the special case of $s_1 = 0$, the two assumptions are equivalent; therefore, SCB implies standard single crossing.

Assumption SCB is mild and is satisfied by a wide class of preferences. In particular, in the case of additive shocks, it is satisfied whenever standard single crossing is. In the case of multiplicative shocks, it is satisfied under the standard assumption that date-2 utility exhibits either constant or decreasing absolute risk aversion (i.e., $-u_2''/u_2'$ either constant or decreasing).

5.1 The formal contracting problem

Just as in the case of verifiable saving decisions, we can restrict attention to direct revelation mechanisms in which self 2 dictates self 3's consumption. Therefore, under a contract in our setting, self 1 first reports a state $\tilde{\phi}^1 \in \Phi$, followed by self 2, who reports a state $\tilde{\phi}^2 \in \Phi$ and a saving decision $\tilde{s}_1 \geq 0$ made by self 1.

Denote by $X(\tilde{\phi}^2, \tilde{s}_1; \tilde{\phi}^1) \in \mathbb{R}^3_+$ the withdrawals specified by a contract X for a given set of reports and assume, without loss of generality, that no savings occur on the equilibrium path of any subgame. The incentive compatibility constraints (which also cover the general n-state case in Section 9) are then as follows: First, self 1 must truthfully report the state and not save; this is an equilibrium action if and only if

$$U^1\left(X(\phi, 0; \phi); \phi\right) \geq U^1(s_1 + X(\phi, s_1; \tilde{\phi}^1); \phi) \text{ for all } \phi, \tilde{\phi}^1 \in \Phi \text{ and } s_1 \geq 0. \qquad (\text{IC}_1)$$

Second, given self 1's actions, self 2 must truthfully report the state and self 1's saving decision, in addition to not saving himself; this is an equilibrium action if and only if

$$U^2(s_1 + X(\phi, s_1; \tilde{\phi}^1); \phi) \geq U^2(s_1 + X(\tilde{\phi}^2, \tilde{s}_1; \tilde{\phi}^1) - s_2; \phi) \text{ for all } \phi, \tilde{\phi}^1, \tilde{\phi}^2 \in \Phi \text{ and } s_1, \tilde{s}_1, s_2 \geq 0. \tag{IC$_2$}$$

In addition, the contract must satisfy $X(\phi, 0; \phi) = c(\phi)$ for all $\phi \in \Phi$, i.e., $\{c(\phi)\}_\phi$ is implemented in equilibrium, and $X_1(\tilde{\phi}^2, \tilde{s}_1; \tilde{\phi}^1) = c_1\left(\tilde{\phi}^1\right)$ for all $\tilde{\phi}^1, \tilde{\phi}^2 \in \Phi$ and $\tilde{s}_1 \geq 0$, i.e., self 2 cannot retroactively change self 1's consumption. Finally, we adopt the mild regularity condition that, for all $\tilde{\phi}^1, \tilde{\phi}^2 \in \Phi$, $X(\tilde{\phi}^2, \bullet; \tilde{\phi}^1)$ is a finite function (of $\tilde{s}_1$), in the sense of having at most finitely many points of discontinuity.[25] When a contract X exists that satisfies all the above constraints, we say that *commitment to* $\{c(\phi)\}$ *is possible* and call such a contract a *commitment contract*.

The agent's ability to privately save restricts him to consumption plans that his future selves do not want to distort by saving. We therefore require $\{c(\phi)\}_\phi$ to satisfy $U^2(c(\phi); \phi) \geq \max_{s_2 \geq 0} U^2(c(\phi) - s_2; \phi)$ and $U^1(c(\phi); \phi) \geq \max_{s_1 \geq 0} U^1(s_1 + c(\phi) - \hat{s}_2(s_1); \phi)$, where $\hat{s}_2(s_1) = \arg\max_{s_2 \geq 0} U^2(s_1 + c(\phi) - s_2; \phi)$, for all $\phi \in \Phi$. Lemma A-3 in the appendix formally establishes these as necessary conditions for commitment to $\{c(\phi)\}_\phi$.

Just as in the case of verifiable savings, we restrict, without loss, attention to consumption plans with the property that, in each state, self 0 prefers the consumption prescribed for that state to the consumption prescribed for the other state. With unverifiable savings, the only difference is that self 0 takes into account future selves' saving decisions. Formally, we require $\{c(\phi)\}_\phi$ to satisfy $U^0(c(\phi); \phi) \geq U^0\left(\hat{s}_1 + c\left(\phi'\right) - \hat{s}_2(\hat{s}_1); \phi\right)$ for all $\phi, \phi' \in \Phi$ such that $\phi \neq \phi'$, where $\hat{s}_1 \equiv \arg\max_{s_1 \geq 0} U^1\left(s_1 + c\left(\phi'\right) - \hat{s}_2(s_1); \phi\right)$ and $\hat{s}_2(s_1) = \arg\max_{s_2 \geq 0} U^2\left(s_1 + c\left(\phi'\right) - s_2; \phi\right)$. If $\{c(\phi)\}_\phi$ fails this condition, then it is dominated by a non-contingent consumption plan that gives the individual $c\left(\phi'\right)$ in both states. Given this restriction, the definition of a commitment problem changes in the obvious way: we say that $\{c(\phi)\}_\phi$ generates a commitment problem in state ϕ if $U^1(c(\phi); \phi) < U^1(\hat{s}_1 + c\left(\phi'\right) - \hat{s}_2(\hat{s}_1); \phi)$, where $\hat{s}_1$ and $\hat{s}_2$ are as defined immediately above. Under this definition of a commitment problem, Lemma 1 holds unchanged for the case of unverifiable savings.

As in the case of verifiable savings we assume, without loss of generality, that $c_1\left(\theta'\right) \geq c_1(\theta)$. We also assume throughout this section that $\{c(\phi)\}_\phi$ generates a commitment problem. It then

[25] This regularity condition is used only in proving Theorem 1, necessity half. It can be relaxed, though only at the cost of introducing economically uninteresting mathematical complexity.

follows from Lemma 1 that there is a commitment problem in state θ but not in state θ', and that $c_1\left(\theta'\right) > c_1\left(\theta\right)$. Just as in the previous section, we will often refer to $c\left(\theta\right)$ and $c(\theta')$ as c and c', respectively.

Since we will ususally speak of the agent deciding how much to withdraw rather than which state to report, it is useful to instead describe the contract as follows: At date 1 the agent (self 1) chooses date-1 withdrawals from the set $\{c_1, c_1'\}$ (and decides how much to save). At date 2 the agent (self 2) chooses date-2 and date-3 withdrawals from one of two withdrawal *schedules* (and decides how much to save): if he chose c_1 at date 1, then he chooses from $\{X(\theta, s_1; \theta)\}_{s_1 \geq 0}$ and $\{X(\theta', s_1; \theta)\}_{s_1 \geq 0}$; if he chose c_1' at date 1, then he chooses from $\left\{X(\theta, s_1; \theta')\right\}_{s_1 \geq 0}$ and $\{X(\theta', s_1; \theta')\}_{s_1 \geq 0}$.

The contracting problem can be simplified significantly. Just as in the case of verifiable savings, we can restrict attention to contracts that give self 2 (c_2, c_3), and let him save whatever he likes, after self 1 chooses c_1. Under this restriction, all incentive compatibility constraints in which the resulting schedules $\{X(\theta, s_1; \theta)\}_{s_1 \geq 0}$ and $\{X(\theta', s_1; \theta)\}_{s_1 \geq 0}$ appear are automatically satisfied. Intuitively, this is simply because self 1 is not tempted by less date-1 consumption in either state; therefore he does not need to be punished for such deviations.

Lemma 3 *Any contract that gives self 2 (c_2, c_3) after self 1 chooses c_1 satisfies (IC_1) and (IC_2) for $\tilde{\phi}^1 = \theta$.*

Similarly, we can restrict attention to contracts in which self 2 gets (c_2', c_3') (and saves whatever he likes) after self 1 either i) correctly chooses c_1' in state θ', and possibly saves, or ii) incorrectly chooses c_1' in state θ and saves more than s_1^*, defined by

$$s_1^* \equiv \sup\left\{s_1 : U^1\left(c; \theta\right) < U^1(s_1 + c' - \hat{s}_2\left(s_1\right); \theta) \text{ where } \hat{s}_2\left(s_1\right) = \arg\max_{s_2 \geq 0} U^2(s_1 + c' - s_2; \theta)\right\}.$$

Formally, we will show that if commitment to (c, c') is possible, then it is possible with a contract that satisfies this restriction. The intuition is much the same as that behind Lemma 3: in case i) above, self 1 is not tempted by less date-1 consumption and in case ii), as the definition of s_1^* makes clear, he is not tempted by the higher date-1 consumption $c_1' - s_1$ in state θ when $s_1 \geq s_1^*$.[26]

Given the above simplifications, it only remains to determine $\left\{X(\theta, s_1; \theta')\right\}_{s_1 \in [0, s_1^*]}$—the punishment chosen by self 2 when self 1 incorrectly chooses c_1' in state θ and saves s_1^* or less. To

[26] We note these simplifications at this point only for expositional ease, and to highlight the simplicity of the contract we eventually derive; all our results prior to Theorem 1 hold independent of them.

simplify notation we write $\hat{X}(s_1) \equiv X(\theta, s_1; \theta')$ and refer to $\hat{X}$ as a contract in what follows. We next turn to characterizing $\hat{X}(s_1)$ for $s_1 \in [0, s_1^*]$.

5.2 General properties of commitment contracts

The agent's ability to save makes the analysis of the contracting problem much more complicated. In particular, self 1 can now choose from a continuum of possible deviations—choose c_1' in state θ and save any positive amount s_1. Each possible deviation must be met with an appropriate penalty chosen by self 2, whose own preferences in turn depend on the savings he inherits from self 1. Consequently, we have a continuum of interlinked contracting problems, indexed by s_1, that cannot be solved independently of each other. Rather, $\hat{X}(s_1)$—the punishment intended for the savings level s_1—must be such that self 2 with inherited savings s_1 prefers it to the punishment intended for any other savings level $\tilde{s}_1$ (formally, this condition is just a particular instance of (IC_2)). Specifically, the contracting problems are tied together by the *self-selection property*

$$U^2(s_1 + \hat{X}(s_1); \theta) \geq U^2(s_1 + \hat{X}(\tilde{s}_1); \theta) \text{ for all } s_1, \tilde{s}_1 \geq 0. \tag{1}$$

To analyze the problem, we first derive a number of properties, discussed below and formally established in the result that follows, that *any* commitment contract must satisfy.

First, $\hat{X}(s_1)$ must be front-loaded relative to c' and must also impose a penalty, in the sense of offering less total consumption than c'. Consequently, the feature of our earlier example that a commitment contract offers *costly excess flexibility* is a general one.[27] Of course, any commitment contract must also offer *costless flexibility*, i.e., the ability to choose either c or c' without penalty.[28]

Front-loading, just as in the case of verifiable saving decisions, follows from Lemma 2 and the fact that there is a commitment problem. The penalty, however, follows from self 2's ability to save. If $\hat{X}(s_1)$—which must be front-loaded relative to c'—offered strictly more consumption than c', then self 2 would never choose c' in state θ', as he should. Rather, he would be better off choosing $\hat{X}(s_1)$ and saving some of the added date-2 consumption.

Second, as s_1 increases, $\hat{X}(s_1)$ must become less front-loaded and impose a smaller penalty. This follows from the effect of self 1's savings on self 2's preferences—as self 2 inherits a higher

[27] By contrast, the punishment $(\hat{c}_2, \hat{c}_3)$ in the verifiable savings case may increase total consumption relative to (c_2', c_3'), with the punishment stemming solely from forcing the agent to consume very little at date 3.

[28] Formally, this is simply the requirement that $X(\phi, 0; \phi) = c(\phi)$ for all $\phi \in \Phi$.

level of savings s_1, he becomes less willing to choose a front-loaded punishment. If instead $\hat{X}(s_1)$ became more front-loaded as s_1 increases, then self 2 with high inherited savings would never choose the point on the $\hat{X}$-schedule intended for him; after all, even self 2 with low inherited savings did not choose that point. Given that $\hat{X}(s_1)$ must become less front-loaded, the penalty must then become smaller. If instead the penalty became larger as s_1 increases, then total consumption $\hat{X}_2(s_1) + \hat{X}_3(s_1)$ and date-2 consumption $\hat{X}_2(s_1)$ would move in the same direction and $\hat{X}(s_1)$ would be dominated by $\hat{X}(\tilde{s}_1)$ for $\tilde{s}_1 < s_1$.

Finally, $\hat{X}_3$ is, up to a boundary condition, completely determined by $\hat{X}_2$, effectively reducing the contracting problem to the choice of a function $\hat{X}_2$. This simplification follows from the requirement, discussed above and formalized by (1), that self 2 always choose the punishment intended for the particular level of savings he inherits. Moreover, when $\hat{X}$ is continuous, the relationship between $\hat{X}_2$ and $\hat{X}_3$ takes the particularly simple form

$$d\hat{X}_3(s_1) = -\frac{U_2^2(s_1 + \hat{X}(s_1); \theta)}{U_3^2(s_1 + \hat{X}(s_1); \theta)} d\hat{X}_2(s_1). \tag{2}$$

Proposition 2 *For any commitment contract $\hat{X}$, (i) $\hat{X}_2(s_1) > c_2'$ and $\hat{X}_2(s_1) + \hat{X}_3(s_1) \leq c_2' + c_3'$ for all $s_1 \in [0, s_1^*)$, (ii) $\hat{X}_2(s_1)$ is weakly decreasing in s_1 and $\hat{X}_2(s_1) + \hat{X}_3(s_1)$ is weakly increasing in s_1,[29] and (iii) if $\hat{X}$ is continuous at s_1, then it satisfies (2).*

5.3 Strong preference reversal

In the case of verifiable saving decisions, preference reversal is necessary for commitment: in order to ensure that self 1 is punished for choosing c_1' in state θ but not punished for doing so in state θ', there must exist a front-loaded punishment that self 2 will choose in state θ but not in state θ'. In the case of unverifiable saving decisions, self 1's ability to save leads to a stronger version of this condition. Recall that as self 2 inherits a higher level of savings s_1, he grows less willing to choose a front-loaded punishment. Therefore, self 2 with inherited savings s_1^* is the *least* willing to choose a front-loaded punishment and self 2 with no inherited savings is the *most* willing to choose a front-loaded punishment. So, to effectively deter self 1 from deviating, there must exist a front-loaded punishment that self 2 with savings s_1^* is willing to choose in state θ but self 2 with no savings will not choose in state θ'. Such a punishment exists only if the following condition—*strong*

[29] As an aside, note that properties (ii) and (iii) in fact hold for all components of the contract $X\left(\tilde{\phi}^2, \cdot; \tilde{\phi}^1\right)$.

preference reversal—is satisfied.

Condition SPR (Strong preference reversal) *There exists $\tilde{x}$ such that $\tilde{x}_2 > c'_2$,*

$$U^2(s_1^* + \tilde{x}; \theta) \geq U^2(s_1^* + c'; \theta), \tag{SPRa}$$

$$U^2(c'; \theta') \geq U^2(\tilde{x} - s_2; \theta') \text{ for all } s_2 \geq 0. \tag{SPRb}$$

Condition SPR extends PR—the preference reversal condition from before—to the case where savings levels differ. As the following result shows, SPR is necessary for commitment.

Proposition 3 *Commitment to (c, c') is possible only if SPR is satisfied.*

Although Proposition 3 is simple to state, its proof is not trivial: while Proposition 2 implies that commitment is possible only if for any $s_1 < s_1^*$ there exists some punishment $\tilde{x}$ that satisfies $\tilde{x}_2 > c'_2$, (SPRb), and (SPRa) with s_1 in place of s_1^*, it is not immediate that there exists $\tilde{x}$ that satisfies $\tilde{x}_2 > c'_2$, (SPRb), and (SPRa) at s_1^* itself.

As its name suggests, however, the basic economic content of SPR is still the same as that of PR—a form of preference reversal. While self 1 values current consumption *more* in state θ' than in state θ, self 2 must value current consumption *less* in state θ' than in state θ. Further, just as with PR, SPR can be checked by comparing the slopes of two indifference curves through c'. This amounts to checking whether SPR holds in the neighborhood of c'.

Lemma 4 *SPR is satisfied if $U_2^2\left(s_1^* + c'; \theta\right) / U_3^2\left(s_1^* + c'; \theta\right) \geq U_2^2(c'; \theta') / U_3^2(c'; \theta')$ and $U_2^2\left(s_1^* + c'; \theta\right) / U_3^2\left(s_1^* + c'; \theta\right) > 1$. Moreover, for additive shocks, the two conditions are equivalent.*

In general, Lemma 4 provides only a sufficient condition for SPR. The reason is that, as the earlier discussion of SCB illustrates, indifference curves may cross more than once when savings levels differ. Therefore, even if the indifference curves cross the wrong way at c', they may cross again at some $\tilde{x}$ such that $\tilde{x}_2 > c'_2$. Allowing for this possibility, SPR can be checked as follows. Define $\hat{x}^0$ as the solution to

$$\min_{\tilde{x}} \{\tilde{x}_2 + \tilde{x}_3\} \text{ s.t. } \tilde{x}_1 = c'_1,\ \tilde{x}_2 \geq c'_2, \text{ and } U^2\left(s_1^* + \tilde{x}; \theta\right) \geq U^2\left(s_1^* + c'; \theta\right).$$

In words, $\hat{x}^0$ is the most severe (in the sense of minimizing total consumption) front-loaded punishment that self 2 with inherited savings s_1^* would ever choose in state θ. As the following result shows, it is then both necessary and sufficient for SPR that self 2 with no inherited savings does not choose $\hat{x}^0$ in state θ', even if he can save, and self 2 with inherited savings s_1^* strictly prefers to borrow at c' in state θ.

Lemma 5 *SPR is satisfied if and only if* $U^2(c';\theta') \geq U^2(\hat{x}^0 - s_2;\theta')$ *for all* $s_2 \geq 0$ *and* $U_2^2\left(s_1^* + c';\theta\right) / U_3^2\left(s_1^* + c';\theta\right) > 1$.

5.4 Main result: necessary and sufficient conditions for commitment

In the previous subsection we showed how self 1's ability to save leads to a stronger necessary condition for commitment (i.e., SPR instead of PR). Self 2's ability to save can limit commitment even further. Specifically, the punishment must not be too front-loaded, or else self 2 would be tempted to make it less front-loaded by saving. Formally, a commitment contract $\hat{X}$ must satisfy the *no-saving* condition

$$U_2^2(s_1 + \hat{X}(s_1);\theta) \geq U_3^2(s_1 + \hat{X}(s_1);\theta) \text{ for all } s_1 \in [0, s_1^*]. \tag{NS}$$

This constraint, which is just a particular instance of (IC_2), significantly limits the severity of punishments that can be imposed on self 1 for deviating.

In order to characterize when commitment is possible, we assume that SPR is satisfied (this is necessary for commitment by Proposition 3) and next construct the *least front-loaded* commitment contract $\hat{X}^*$. The key feature of $\hat{X}^*$ is that it is the commitment contract under which self 2 is least tempted to save. Therefore, if $\hat{X}^*$ does not satisfy NS, then no commitment contract will.

Accordingly, we first define the boundary value $\hat{X}^*(s_1^*)$ as the least front-loaded punishment that is mild enough to induce self 2 with inherited savings s_1^* to choose it in state θ but also severe enough to deter self 2 with no inherited savings from choosing it in state θ'. Formally,

$$\hat{X}_2^*(s_1^*) \equiv \inf\left\{\tilde{x}_2 : \tilde{x}_2 > c_2' \text{ and for some } \tilde{x}_3,\ (\tilde{x}_2, \tilde{x}_3) \text{ satisfies (SPRa) and (SPRb)}\right\},$$

while $\hat{X}_3^*(s_1^*)$ is defined by setting (SPRa) to equality.[30]

As the following result confirms, $\hat{X}^*(s_1^*)$ is indeed the least front-loaded punishment prescribed by *any* commitment contract. (The formal proof is harder than one might suspect, because of

[30] Equivalently, one could define $\hat{X}_3^*(s_1^*)$ by setting (SPRb) to equality. For a formal argument, see the proof of Theorem 1, sufficiency half, Claim 1.

complications similar to those encountered when proving Proposition 3.)

Proposition 4 *For any commitment contract $\hat{X}$, $\hat{X}_2(s_1) \geq \hat{X}_2^*(s_1^*)$ for all $s_1 \in [0, s_1^*)$.*

Thus defined, $\hat{X}^*(s_1^*)$ ensures that self 1 will not deviate to c_1' and save s_1^* in state θ. To see why, first note that, by the definition of s_1^*, self 1 will not deviate in this way if the deviation is followed by c' at dates 2 and 3. Then, since self 2 is indifferent between c' and the front-loaded punishment $\hat{X}^*(s_1^*)$, it follows that self 1 will dislike the deviation even more if it is followed by $\hat{X}^*(s_1^*)$.

We next define the remainder of the contract for $s_1 < s_1^*$, working downwards from its boundary value $\hat{X}^*(s_1^*)$. There are two cases to consider. First, suppose that for some s_1, self 1 strictly prefers not to deviate and save s_1, i.e., $U^1(c;\theta) > U^1(s_1 + \hat{X}^*(s_1);\theta)$. In this case, we can simply keep the contract fixed for $\tilde{s}_1$ immediately below s_1, i.e., $d\hat{X}^*/ds_1 = 0$, without violating self 1's incentive compatibility constraint. In fact, keeping $\hat{X}^*$ constant is the best we can do, since a commitment contract cannot become *less* front-loaded as s_1 decreases (see Proposition 2).

Second, suppose that for some s_1, self 1 is instead indifferent towards deviating and saving s_1, i.e., $U^1(c;\theta) = U^1(s_1 + \hat{X}^*(s_1);\theta)$. In this case, unlike the previous one, it may not be possible to keep the contract fixed without violating self 1's incentive compatibility constraint. If it is possible, then the contract stays fixed as in the previous case. If not, then $\hat{X}^*$ must become more front-loaded as s_1 decreases. Here it is helpful to note that the least front-loaded commitment contract is also the one that punishes self 1 as mildly as possible. This follows from Lemma 2: the more severe the punishment imposed by a commitment contract is, the more front-loaded it must be to ensure that self 2 is willing to impose it. Consequently, as s_1 decreases, the contract should keep the punishment as mild as possible by holding self 1's deviation utility fixed, i.e.,

$$\frac{d}{ds_1}U^1(s_1 + \hat{X}(s_1);\theta) = 0, \tag{3}$$

thereby keeping him indifferent towards deviating for $\tilde{s}_1$ immediately below s_1. Finally, we set $\hat{X}^*$ to also satisfy (2), as every continuous commitment contract must (see Proposition 2).

More formally, $\hat{X}^*$ is defined as follows. Define s_1^{**} as the boundary between the two cases described above, i.e.,

$$s_1^{**} \equiv \inf\left\{s_1 \in [0, s_1^*] : U^1(c;\theta) \geq U^1(\tilde{s}_1 + \hat{X}^*(s_1^*);\theta) \text{ for all } \tilde{s}_1 \in [s_1, s_1^*]\right\}.$$

Then for $s_1 \in [s_1^{**}, s_1^*)$, $\hat{X}^*(s_1) = \hat{X}^*(s_1^*)$. For $s_1 \in [0, s_1^{**})$, the contract is defined by the boundary condition $\hat{X}^*(s_1^{**}) = \hat{X}^*(s_1^*)$ and the pair of differential equations (2) and

$$\frac{d\hat{X}_2}{ds_1} = \min\left\{-\frac{u_1'(c_1' - s_1;\theta) - \beta u_2'(s_1 + \hat{X}_2(s_1);\theta)}{(1-\beta)\,u_2'(s_1 + \hat{X}_2(s_1);\theta)}, 0\right\}, \tag{4}$$

where (4) is simply (3) with a minimization operator to ensure that, in the second case described above, the contract is kept constant whenever possible.

Theorem 1 *Commitment to (c, c') is possible if and only if SPR and NS are satisfied, in which case $\hat{X}^*$ is a commitment contract.*

In the case of additive shocks—or, more generally, when the simple sufficient condition for SPR of Lemma 4 is satisfied—the definition of $\hat{X}^*(s_1^*)$ is particularly simple:[31]

Lemma 6 *If $U_2^2(s_1^* + c';\theta)/U_3^2(s_1^* + c';\theta) \geq U_2^2(c';\theta')/U_3^2(c';\theta')$ and $U_2^2(s_1^* + c';\theta)/U_3^2(s_1^* + c';\theta) > 1$, then $\hat{X}^*(s_1^*) = c'$ and $s_1^{**} = s_1^*$.*

5.5 Sketch proof of Theorem 1

We illustrate two key steps. First—and related to necessity—we show that $\hat{X}^*$ is *pointwise* the least front-loaded commitment contract, in the sense that no commitment contract exists that is less front-loaded at *any* point. In other words, we show that one cannot make the contract *less* front-loaded at one point by making it *more* front-loaded at another point. This is a detail we brushed over in our construction of $\hat{X}^*$ above.

Consider a particular level of date-1 savings s_1, and suppose that a contract $\hat{X}$ has been defined for all savings levels from s_1^* down to $s_1 + \varepsilon$, where ε is small and positive. Recall that (2) places a strong restriction on how the date-2 and date-3 elements of the contract vary as s_1 changes. This condition in turn determines how self 2's utility varies with s_1, namely

$$dU^2(s_1 + \hat{X}(s_1);\theta) = U_2^2 ds_1 + U_2^2 d\hat{X}_2 + U_3^2 d\hat{X}_3 = U_2^2(s_1 + \hat{X}(s_1);\theta)ds_1. \tag{5}$$

The contract specification at $s_1 + \varepsilon$ and (5) together determine how much *utility* the contract must give self 2 at savings level s_1. However, they do not determine whether self 2 should be given this utility through consumption at date 2 or at date 3.

[31] In this special case, Proposition 4 is an immediate corollary of Proposition 2.

The important implication of (5) is that there is no conflict between making the contract less front-loaded *at* s_1 and making it less front-loaded *below* s_1. In fact, making the contract less front-loaded at s_1 actually helps make the contract less front-loaded below s_1, say at $s_1 - \varepsilon$. The reason is that the less front-loaded the contract is at s_1, the lower self 2's utility must be at $s_1 - \varepsilon$ (see (5)), and hence date-2 consumption can be decreased, making the contract less front-loaded at $s_1 - \varepsilon$. In conclusion, $\hat{X}^*$ is pointwise the least front-loaded commitment contract.

Second—and related to sufficiency—we show that, given SPR, $\hat{X}^*$ satisfies all the incentive compatibility constraints except NS. By construction, (4) ensures that (IC_1) is satisfied; therefore, self 1 does not want to deviate. Moreover, it turns out that for continuous commitment contracts, (2) is not only necessary, but also sufficient for self-selection (1);[32] therefore, conditional on punishing, self 2 will always choose the particular punishment intended for the savings level he inherits.

It remains to show that self 2 will punish if and only if self 1 deviates. By construction, the punishment $\hat{X}^*(s_1^*)$ satisfies (SPRa), i.e., self 2 with inherited savings s_1^* prefers it to c' in state θ. Moreover, since punishments are front-loaded, this preference is preserved at lower levels of inherited savings. More precisely, self 2 with inherited savings s_1 prefers the intended punishment $\hat{X}^*(s_1)$ to c' in state θ, i.e.,

$$U^2(s_1 + \hat{X}^*(s_1); \theta) \geq \max_{s_2 \geq 0} U^2\left(s_1 + c' - s_2; \theta\right), \tag{6}$$

as follows. By construction, (6) is satisfied for $s_1 = s_1^*$. It is also satisfied for $s_1 < s_1^*$ since

$$U^2(s_1 + \hat{X}^*(s_1); \theta) \geq U^2(s_1 + \hat{X}^*(s_1^*); \theta) \geq \max_{s_2 \geq 0} U^2\left(s_1 + c' - s_2; \theta\right),$$

where the first inequality follows from self-selection (1) and the second from the fact that $\hat{X}^*$ is front-loaded relative to c'.

Similarly, $\hat{X}^*(s_1^*)$ satisfies (SPRb) by construction, i.e., self 2 with no inherited savings prefers c' to the punishment $\hat{X}^*(s_1^*)$ in state θ'. Since punishments are front-loaded, this preference is preserved at higher levels of inherited savings. More precisely, self 2 with inherited savings s_1 prefers c' to any punishment $\hat{X}^*(\tilde{s}_1)$ in state θ', i.e.,

$$\max_{s_2 \geq 0} U^2(s_1 + c' - s_2; \theta') \geq \max_{s_2 \geq 0} U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta'), \tag{7}$$

[32] For details, see the proof of Theorem 1, sufficiency half, Claim 4, Step A.

as follows. By construction, (7) is satisfied for $s_1 = 0$ and $\tilde{s}_1 = s_1^*$. By self-selection (1), $\hat{X}^*(\tilde{s}_1)$ lies below self 2's indifference curve in state θ, with inherited savings s_1^*, through $\hat{X}^*(s_1^*)$. By SCB, it then follows that $\hat{X}^*(\tilde{s}_1)$ also lies below self 2's indifference curve in state θ', with no inherited savings, through $\hat{X}^*(s_1^*)$. Consequently, (7) is satisfied for $s_1 = 0$ and all $\tilde{s}_1$. Moreover, since $\hat{X}^*$ is front-loaded relative to c', self 2's preference for c' can only strengthen as s_1 increases. As a result, (7) holds for all s_1 and $\tilde{s}_1$.

6 When is commitment possible?

Together, SPR and NS are both necessary and sufficient for commitment. Although their roles in ensuring the possibility of commitment are intuitively straightforward, it is not immediately clear how to relate them to the primitives of our economic setting. In particular, it is not obvious what conditions on the states θ and θ', or on the consumption plans c and c', would correspond to SPR and NS. In this section we address precisely this question in the context of commitment to self 0's most preferred consumption plan $\{c^*(\phi)\}_\phi$, defined by

$$c^*(\phi) \equiv \arg\max_{\tilde{c}} U^0(\tilde{c};\phi) \;\; s.t. \;\; \sum_t \tilde{c}_t \leq W$$

for $\phi \in \{\theta, \theta'\}$. Our first result is that, in this important special case, SPR is easily related to c^* and c'^*:

Proposition 5 *If $c_3'^* \geq c_3^*$ then SPR is satisfied. Moreover, for additive shocks, the two conditions are equivalent.*

In words, the condition $c_3'^* \geq c_3^*$ says that the increase in date-1 consumption afforded by c'^* comes solely at the expense of date-2 consumption (relative to c^*), without negatively impacting date-3 consumption.

Our next result establishes conditions under which the consumption comparison $c_3'^* > c_3^*$ guarantees not only SPR, but also NS—and hence the possibility of fully overcoming the agent's commitment problem.

Proposition 6 *Suppose that $c_3'^* > c_3^*$ and let β^* be the supremum value of β for which (c^*, c'^*) generates a commitment problem. Then commitment to (c^*, c'^*) is possible for all β sufficiently close to β^*.*

Moreover, numerical simulations (available on authors' webpages) indicate that the only cases in which $c_3'^* > c_3^*$ but NS fails are those with very low values of β, e.g., $\beta \approx 0.1$.

A leading class of shocks in which the condition $c_3'^* > c_3^*$ is satisfied is that of *one-period-ahead shocks* in which the agent learns about a change in his utility one period in advance, i.e., the shock only affects date-2 utility. For example, the agent might learn at date 1 whether, at date 2, he will receive vacation, suffer a temporary decline in income,[33] receive a bonus, or, in the procrastination interpretation of our setting, whether his favorite film is showing. In each of these examples, the shock affects only date-2 marginal utility and therefore pushes self 0's most preferred consumption at dates 1 and 3 in the same direction; therefore, whenever $c_1'^* > c_1^*$ then $c_3'^* > c_3^*$ also. By Proposition 6, commitment to (c^*, c'^*) is often possible in these cases.

The opposite of a one-period-ahead shock is a *contemporaneous shock*, in which the agent learns about a change in his utility only contemporaneously, i.e., the shock only affects date-1 utility. In contrast to the case of one-period ahead shocks, commitment is never possible in the case of contemporaneous shocks. This is especially easy to see when shocks are additive: contemporaneous shocks push self 0's most preferred consumption at dates 2 and 3 in the same direction, and so whenever $c_1'^* > c_1^*$ then $c_3'^* < c_3^*$, and Proposition 6 implies that SPR fails.[34]

One-period ahead shocks and contemporaneous shocks are associated with $c_3'^* > c_3^*$ and $c_3'^* < c_3^*$ respectively. The boundary case of $c_3'^* = c_3^*$ is a *timing shock*, since there $(c_1'^*, c_2'^*) = (c_2^*, c_1^*)$. The wedding example of the introduction is one such case; other good examples include the replacement of an old car when it breaks down and the need to make a down payment on a house purchase. By Proposition 5, SPR is satisfied in these cases, and numerical simulations suggest that NS is often satisfied also and so commitment to (c^*, c'^*) is possible.

Our discussion above relates entirely to the important question of whether the agent can fully overcome his commitment problem, i.e., commit to self 0's most-preferred consumption plan (c^*, c'^*). When this is not possible, what can we say about the optimal contract?

[33] For example, at date 1 the agent learns he will lose his job at date 2, but anticipates finding a new job by date 3.

[34] In general, SPR fails for contemporaneous shocks as follows. Note that (A) the key savings level s_1^* is strictly positive whenever (c^*, c'^*) generates a commitment problem and (B) for contemporaneous shocks, self 2's preferences are the same in both states. Consider any $\tilde{x}$ that is strictly front-loaded relative to c'^*, i.e., $\tilde{x}_2 > c_2'^*$, and such that self 2 with no savings prefers c'^* to $\tilde{x}$ in state θ', i.e., $u_2(\tilde{x}_2; \theta') - u_2(c_2'^*; \theta') + \beta(u_3(\tilde{x}_3; \theta') - u_3(c_3'^*; \theta')) \leq 0$. Facts (A) and (B) above, together with strict concavity of u_2, imply that self 2 must then strictly prefer c'^* to $\tilde{x}$ in state θ when he has savings s_1^*, i.e., $u_2(s_1^* + \tilde{x}_2; \theta') - u_2(s_1^* + c_2'^*; \theta') + \beta(u_3(\tilde{x}_3; \theta') - u_3(c_3'^*; \theta')) < 0$. In words, there is no punishment $\tilde{x}$ that self 2 would choose in state θ but not in state θ'. Hence SPR fails.

For concreteness, we focus on the case of additive shocks. Here, the preference reversal condition PR depends only on the shocks and not on the consumption plan (c, c'). At the same time, SPR—a more demanding condition—depends on the consumption plan (c, c') *only* through the key savings level s_1^*. Given these observations, it is immediate that if PR fails, then the agent cannot overcome any commitment problem, even one in which (c, c') has been distorted away from (c^*, c'^*) to make c' less tempting for self 1 in state θ. Rather, (c, c') must be chosen so that there is no commitment problem in the first place. Formally, this *no-temptation* consumption plan solves

$$\max_{\{c(\phi)\}_\phi} \sum_{\phi \in \{\theta, \theta'\}} \Pr(\phi) \, U^0 (c(\phi); \phi)$$

subject to the no-temptation constraint $U^1 (c(\phi); \phi) \geq \max_{s_1 \geq 0} U^1 \left(s_1 + c(\phi') - \hat{s}_2(s_1); \phi\right)$, where $\hat{s}_2(s_1) = \arg\max_{s_2 \geq 0} U^2 \left(s_1 + c(\phi') - \hat{s}_2(s_1); \phi\right)$, and the previously noted resource constraint and no-savings constraints for selves 1 and 2.

If instead PR is satisfied, but either SPR or NS fails at (c^*, c'^*), then the agent may be able to to commit to a consumption plan that self 0 prefers to the no-temptation consumption plan, but not to (c^*, c'^*). In particular, if SPR fails, then at least one of c and c' can be distorted away from (c^*, c'^*) to make c' less tempting for self 1 in state θ. Doing so reduces the key savings level s_1^*; and once s_1^* is reduced enough, SPR is satisfied.

In summary, there are three distinct types of shocks: those for which self 0 can fully overcome his commitment problem; those for which he must settle for a no-temptation consumption plan; and an intermediate class for which self 0 can do better than a no-temptation consumption plan but cannot commit to (c^*, c'^*).

7 Implementation using simple financial instruments

As we stress in the introduction, commitment contracts have characteristics that are widely observed in real-world contractual arrangements. In particular, any commitment contract must offer a combination of *costless* flexibility and *costly* excess flexibility (see discussion prior to Proposition 2). This two-tiered structure is widely observed. For example, many savings accounts allow limited penalty-free withdrawals, followed by withdrawals carrying a penalty rate, and many borrowing arrangements specify interest rates that increase as the agent borrows more. In particular, retirement accounts often have this characteristic. In the U.S., for example, an individual can withdraw a limited amount without penalty (via a 401(k) loan), while larger withdrawals carry a 10% penalty. In the procrastination interpretation of our model, where the consumption good is leisure, the combination of costless and costly flexibility corresponds to the widespread workplace norm that the penalty for missing one deadline is small, but that missing subsequent deadlines carries a penalty, often in the form of more work (for example, the further one is past a referee report deadline, the more one feels the need to produce an especially thorough report).

We next show that in most cases in which commitment to self 0's most preferred consumption plan[35] (c^*, c'^*) is possible, a commitment contract can be constructed by combining simple financial instruments—specifically, savings accounts and lines of credit.[36] The key step in this construction is that the two features of costless flexibility and costly excess flexibility need not be made contingent on one another, but can instead be provided by two *separate* contracts. Specifically, we show that costly excess flexibility can be offered to self 2 regardless of self 1's consumption choice.

Proposition 7 *Suppose that $c_3'^* \geq c_3^*$ (and hence SPR is satisfied).*[37] *Costly excess flexibility can be offered to self 2 unconditionally without affecting commitment. Formally, self 2 can be given the option to increase his date-2 consumption by $\hat{X}_2(s_1) - c_2'^*$ at the cost of decreasing his date-3 consumption by $c_3'^* - \hat{X}_3(s_1)$,* regardless *of self 1's choice between $c_1'^*$ and c_1^*.*

The intuition for Proposition 7 is that, if self 1 correctly chooses the lower consumption c_1^* in state θ, he leaves self 2 with ample consumption, and hence self 2 finds it relatively unattractive to pay a penalty to *further* increase date-2 consumption.

Costly flexibility is readily interpretable as either a savings withdrawal at a penalty rate—just as in the example from the introduction—or as a high-interest loan. By Proposition 7, two separate contracts can be used to provide the agent with i) the costless flexibility he needs to respond to the shock and ii) the incentives to not abuse that flexibility. We next detail how this can be done using simple financial instruments.

At date 0, the agent invests in the following certificates of deposit (CDs): a one-period CD

[35] The focus on self 0's most preferred consumption plan simplifies the proof of Proposition 7, along with the conditions required for it to hold. However, we emphasize that the conclusion holds much more generally: for many consumption plans $\{c(\phi)\}$ for which commitment is possible, this commitment is achievable using only simple financial instruments.

[36] We thank Andrew Postlewaite for this suggestion.

[37] See Proposition 5. Moreover, by the same result $c_3'^* \geq c_3^*$ is necessary for commitment in the case of additive shocks.

with face value $A_1 = c_1^*$, a two-period CD with face value $A_2 = c_1^* + c_2^* - c_1'^*$, and a three-period CD with face value $A_3 = c_3^* + c_1'^* - c_1^*$. At the same time, the agent arranges for a penalty-free line of credit of $L = c_1'^* - c_1^*$. If the line is drawn at date 2, the date-3 repayment is simply L. If instead the line is drawn at date 1, the date-2 and date-3 repayments are $P_2 = c_1^* + c_2^* - c_1'^* - c_2'^*$ and $P_3 = L - P_2 = c_3^* - c_3'^* + c_1'^* - c_1^*$, respectively. Hence if the credit line is drawn at date 1 his consumption is $(A_1 + L, A_2 - P_2, A_3 - P_3) = c'^*$, while if it is drawn at date 2 his consumption is $(A_1, A_2 + L, A_3 - L) = c^*$.

We close this section with a couple of final observations. First, the exact form of a commitment contract depends on the parameters of the problem, such as the degree of impatience (β) and the shocks (θ and θ'). A notable feature of our setting relative to much of the contracting literature is that the agent is happy to truthfully report these details at date 0. Concretely, and in terms of the discussion above, the agent has the right incentives to choose financial contracts with suitable interest rates and penalties for early withdrawals; and in the procrastination setting, he has the right incentives to set his own deadlines. In essence, self 0 is the principal. In this sense, a nice feature of our contracting problem is that it is considerably less informationally demanding than those between a distinct principal and agent(s).

Second, in common with much of the contracting literature, we have conducted our analysis under an assumption of *exclusivity*: after writing the original contract at date 0, the agent cannot enter into additional contracts with other counterparties at dates 1 or 2. In general, assuming exclusivity of savings arrangements is much less plausible than assuming exclusivity of borrowing arrangements. Institutions such as credit registries and collateral registries make it hard to borrow from multiple lenders without their knowledge. Indeed, the presence of such institutions seems important for the very existence of a well-functioning credit market. Regarding exclusivity of savings arrangements, our analysis already allows for the possibility of entering additional one-period savings contracts. Consequently, the only remaining issue is that self 1 might wish to enter a new two-period savings contract. Fortunately, it is straightforward to adapt the original date-0 contract to rule out this possibility, by specifying that the agent should be able to borrow, interest-free, against any savings contract entered by self 1. This effectively transforms a two-period savings contract back into a one-period savings contract.

8 Naïveté

Thus far, we have assumed that the agent is fully self-aware (sophisticated), in the sense that at any date, he correctly anticipates his future selves' preferences. In this section we show that commitment contracts of the kind we analyze above often enable an agent who is partially (but not completely) naïve about his future selves' preferences to commit. We follow the literature and use the specification introduced by O'Donoghue and Rabin (2001): at each date, the agent's true hyperbolic discount rate is β, but he incorrectly believes that his future selves' rate is $\tilde{\beta} > \beta$. We continue to write U^t for self t's true preferences, and use $\tilde{U}^t$ to denote the preferences incorrectly attributed to self t by prior selves.

For transparency, we focus on the case of verifiable savings (parallel considerations apply to the case of unverifiable savings) and, as before, we assume that (c, c') generates a commitment problem in state θ. A necessary condition for commitment to (c, c') is then that there exists a punishment $\hat{c}$ such that $\hat{c}_1 = c'_1$ that satisfies the following three conditions, which overlap with those established in Section 4: First, $\hat{c}$ must give self 1 less utility than c in state θ; formally, $U^1(c;\theta) \geq U^1(\hat{c};\theta)$. Second, self 1 must believe that self 2 will choose $\hat{c}$ in state θ; formally, $\tilde{U}^2(\hat{c};\theta) \geq \tilde{U}^2(c';\theta)$. Note that this condition relates to self 1's incorrect perception of self 2's preferences, not to self 2's true preferences. Finally, for c' to be the equilibrium outcome in state θ', self 2 must prefer c' to $\hat{c}$ in state θ'; formally, $U^2(c';\theta') \geq U^2(\hat{c};\theta')$.

These necessary conditions are, in fact, also sufficient since commitment to (c, c') is possible using a contract in which self 2 is given the choice between $\hat{c}$ and c' whenever self 1 chooses c'_1. To see this, it remains only to show that self 1 chooses c'_1 over c_1 in state θ'. Just as in Section 4, it follows from the conditions $\tilde{U}^2(\hat{c};\theta) \geq \tilde{U}^2(c';\theta)$ and $U^1(c;\theta) \geq U^1(\hat{c};\theta)$ that $\hat{c}$ is strictly front-loaded relative to c'. Given this, $U^2(c';\theta') \geq U^2(\hat{c};\theta')$ implies that $\tilde{U}^2(c';\theta') \geq \tilde{U}^2(\hat{c};\theta')$, i.e., self 1 believes that self 2 will choose c' in state θ', even given self 1's incorrect beliefs about self 2's preferences. Finally, by Lemma 1 and our assumption that (c, c') generates a commitment problem in state θ, $U^1\left(c';\theta'\right) \geq U^1\left(c;\theta'\right)$ and so self 1 indeed chooses c'_1 over c_1 in state θ'.

From these three conditions, one can readily see that partial naïveté generates a tighter set of necessary conditions for commitment. As such, and as one might expect, commitment becomes more difficult as the agent's naïveté increases. We stress, however, that by continuity slight naïveté has only a small impact on the range of circumstances under which commitment is possible.[38]

While commitment is often possible when the agent is partially naïve, naïveté does have a significant impact on how the contract should be interpreted. Under full sophistication, self 0 has every incentive to sign up for a commitment contract. However, under partial naïveté self 0 may incorrectly believe that he does not have a commitment problem. In these circumstances, there is scope for a benevolent government to improve welfare (at least for self 0) by imposing a commitment contract. Rules governing withdrawals from 401(k) accounts—which, as discussed above, qualitatively resemble commitment contracts—are arguably such a case.

However, it also important to note that while a government-mandated commitment contract can improve the welfare of a partially naïve agent, it can actually hurt a very naïve agent, relative to the alternative of simply allowing self 1 to choose freely between c and c'. For simplicity, we illustrate this point for the case of verifiable savings. First, note that in any commitment contract, the punishment $\hat{c}$ must satisfy

$$u_2\left(\hat{c}_2;\theta\right)+u_3\left(\hat{c}_3;\theta\right)<u_2(c_2';\theta)+u_3(c_3';\theta), \tag{8}$$

since otherwise the punishment would not deter self 1 from overconsuming in state θ.[39] Consequently, at date 1 a completely naïve agent (i.e., $\tilde{\beta}=1$) will choose c_1' in state θ, believing that self 2 will choose c_2'. However, after self 1 chooses c_1' self 2 in fact chooses the punishment $(\hat{c}_2,\hat{c}_3)$. Self 0's equilibrium utility in state θ is hence $U^0\left(\left(c_1',\hat{c}_2,\hat{c}_3\right);\theta\right)$. But by (8), this is strictly less than the utility self 0 would get from a contract allowing self 1 to choose freely between c and c', namely $U^0\left(c';\theta\right)$. Consequently, although there is scope for government paternalism to improve welfare if the government has a reasonably precise estimate of the degree of naïveté, such paternalism is dangerous if agents are instead much more naïve than the government believes.[40]

[38] At the extreme of complete naïveté (i.e., $\tilde{\beta}=1$), commitment is clearly impossible: self 1 must strictly prefer c' to $\hat{c}$ but he must also believe that self 2 prefers $\hat{c}$ to c'; this is not possible since self 1 believes that self 2's preferences over consumption at dates 2 and 3 are the same as his own.

[39] Formally, if $u_2\left(\hat{c}_2;\theta\right)+u_3\left(\hat{c}_3;\theta\right)\geq u_2\left(c_2';\theta\right)+u_3\left(c_3';\theta\right)$ then $U^1\left(\left(c_1',\hat{c}_2,\hat{c}_3\right);\theta\right)\geq U^1\left(c';\theta\right)>U^1\left(c;\theta\right)$.

[40] Eliaz and Spiegler (2006) analyze profit maximization by a monopolist who deals with a population of time-inconsistent individuals who differ in their degree of sophistication. The problem noted in the main text suggests that the parallel question of welfare maximization for a population of differentially sophisticated time-inconsistent individuals would also be interesting. We leave this topic for future research.

9 Extension to three or more states

In this section we show how our analysis can be extended to cover the case of more than two shock realizations, i.e., when the state ϕ is drawn from a set Φ with three or more members. In general, increasing the size of the state space adds considerable analytical complexity to the problem. Our main observation in this section, however, is that in the leading special case of additive shocks—$\Phi \subset \mathbb{R}^3$, $u_t(c_t;\phi) \equiv u_t(c_t+\phi_t)$—with ϕ_3 constant across states in Φ, this adds relatively little complexity beyond that already introduced by unverifiable savings.

The formal problem defined in Section 5 covers arbitrary state spaces Φ. Note first that the problem separates into $|\Phi|$ independent subproblems indexed by self 1's consumption choice $c_1(\tilde{\phi}^1)$. For the remainder of the section, fix $\tilde{\phi}^1$ and consider a representative subproblem.

A key observation when shocks are additive is that the date-2 component of the shock, ϕ_2, and self 1's saving decision, s_1, affect self 2's preferences only through their sum $s_1+\phi_2$; formally, $U^2(s_1+x;\phi) = u_2(s_1+\phi_2+x_2)+\beta u_3(\phi_3+x_3)$. Consequently, if $\underline{\phi}$ is the state in which ϕ_2 is minimal, it follows that for all $s_1 \geq 0$ and $\phi \in \Phi$, $X(\phi, s_1;\cdot) \equiv X(\underline{\phi}, s_1+\phi_2-\underline{\phi}_2;\cdot)$.[41] By the same argument, the condition $X(\phi,0;\phi) = c(\phi)$ becomes $X(\underline{\phi},\phi_2-\underline{\phi}_2;\phi) = c(\phi)$.

Let $\check{s}_1 = s_1+\phi_2-\underline{\phi}_2$; self 2's incentive constraints in the $\tilde{\phi}^1$-subproblem then simplify to

$$U^2(\check{s}_1+X(\underline{\phi},\check{s}_1;\tilde{\phi}^1);\underline{\phi}) \geq U^2(\check{s}_1+X(\underline{\phi},\tilde{s}_1;\tilde{\phi}^1)-s_2;\underline{\phi}) \text{ for all } \check{s}_1, \tilde{s}_1, s_2 \geq 0.$$

In words, just as in the two-state problem we only need to ensure that self 2 chooses the menu item intended for a given level of self 1's "savings" $\check{s}_1$—now a composite of the state ϕ and true savings s_1—and does not himself save.

Self 1's incentive constraints for the $\tilde{\phi}^1$-subproblem can be written as

$$\min_{s_1\geq 0 \text{ and } \phi\in\Phi \text{ s.t. } s_1+\phi_2=\check{s}_1+\underline{\phi}_2} \{U^1(c(\phi);\phi) - U^1(s_1+X(\underline{\phi},\check{s}_1;\tilde{\phi}^1);\phi)\} \geq 0 \text{ for all } \check{s}_1 \geq 0.$$

In words, for every composite savings level $\check{s}_1+\underline{\phi}_2$, there is an actual savings level $s_1 \geq 0$ and a state $\phi \in \Phi$ at which self 1's incentive compatibility condition is tightest. Expanding, self 1's

[41] We are implicitly ruling out contracts that require self 2 to break indifference in different ways in different states.

constraint can be written more explicitly as

$$
\begin{aligned}
& u_2(\check{s}_1 + \underline{\phi}_2 + X_2(\underline{\phi}, \check{s}_1; \tilde{\phi}_1)) + u_3(\phi_3 + X_3(\underline{\phi}, \check{s}_1; \tilde{\phi}_1)) \\
\leq & \min_{s_1 \geq 0 \text{ and } \phi \in \Phi \text{ s.t. } s_1 + \phi_2 = \check{s}_1 + \underline{\phi}_2} \{U^1 (c(\phi); \phi) - u_1(-s_1 + \phi_1 + c_1(\tilde{\phi}^1))\},
\end{aligned} \tag{9}
$$

which makes clear the one-dimensional character of the constraint: for each composite savings level $\check{s}_1 + \underline{\phi}_2$, there is a single inequality that $X(\underline{\phi}, \check{s}_1; \tilde{\phi}_1)$ must satisfy.

To summarize, the general contracting problem separates into $|\Phi|$ subproblems. In the important special case of additive shocks, each of these subproblems is just as tractable as our basic two-state problem, since the state enters self 2's preferences in exactly the same way as date-1 savings. Consequently, each of these subproblems can be analyzed in the same way as the basic two-state problem; specifically, a straightforward analogue of Theorem 1 can be used to characterize when each subproblem can solved.

10 Concluding remarks

Our analysis characterizes the circumstances under which an individual can use a contract to completely overcome his commitment problem, even in the face of uncertainty about future consumption needs. In these situations, hyperbolic discounting ceases to affect the individual's behavior. At the same time, there also exist important cases in which such commitment is not possible, even though we have placed absolutely no restrictions on the class of possible commitment devices.

Moreover, although we focus on one particular form of time-inconsistent preferences—namely the present-bias generated by hyperbolic discounting—we believe that our general arguments are widely applicable to other sources of time-inconsistent preferences. For example, an individual may understand today that, in the future, he will misinterpret the relevance of a small number of data points. We leave the extension of our arguments to other settings for future research.

References

[1] Amador, M., I. Werning, and G.M. Angeletos (2003), "Commitment vs. Flexibility," NBER working paper 10151.

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

[3] Ariely, D., and K. Wertenbroch (2002), "Procrastination, Deadlines, and Performance: Self-Control by Precommitment," *Psychological Science* 13, 219-224.

[4] Bénabou, R., and J. Tirole (2002), "Self-Confidence and Personal Motivation," *Quarterly Journal of Economics* 117, 871-915.

[5] Bénabou, R., and J. Tirole (2004), "Willpower and Personal Rules," *Journal of Political Economy* 112, 848-886.

[6] Bernheim, B.D., D. Ray, and S. Yeltekin (1999), "Self-Control, Saving, and the Low Asset Trap," working paper.

[7] Carrillo, J.D., and T. Mariotti (2000), "Strategic Ignorance as a Self-Disciplining Device," *Review of Economic Studies* 67, 529-544.

[8] Cole, H.L., and N.R. Kocherlakota (2001), "Efficient Allocations with Hidden Income and Hidden Storage," *Review of Economic Studies* 68, 523-542.

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

[10] Doepke, M., and R.M. Townsend (2006), "Dynamic mechanism design with hidden income and hidden actions," *Journal of Economic Theory* 126, 235-285.

[11] Eliaz, K., and R. Spiegler (2006), "Contracting with Diversely Naive Agents," *Review of Economic Studies* 73, 689-714.

[12] Frederick, S., G. Loewenstein, and T. O'Donoghue (2002), "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature* 40, 351-401.

[13] He, Zhiguo (2011), "Dynamic Compensation Contracts with Private Savings," working paper.

[14] Jackson, Thomas H. (1986), *The Logic and Limits of Bankruptcy Law*. Cambridge, MA: Harvard University Press.

[15] Kocherlakota, Naryana R. (2004), "Figuring out the impact of hidden savings on optimal unemployment insurance," *Review of Economic Dynamics* 7, 541-554.

[16] Krusell, P., and A.A. Smith (2003), “Consumption-Savings Decisions with Quasi-Geometric Discounting,” *Econometrica* 71, 365-375.

[17] Laibson, D. (1996), “Hyperbolic Discount Functions, Undersaving and Savings Policy,” working paper.

[18] Laibson, D. (1997), “Golden Eggs and Hyperbolic Discounting,” *Quarterly Journal of Economics* 112, 443-477.

[19] Laibson, D., A. Repetto, and J. Tobacman (1998), “Self-Control and Saving for Retirement,” *Brookings Papers on Economic Activity* 57, 91-196.

[20] Luttmer, E.G.J, and T. Mariotti (2003), “Subjective Discounting in an Exchange Economy,” *Journal of Political Economy* 111, 959-989.

[21] Maskin, E. and T. Sjöström (2002), “Implementation Theory,” in K. Arrow, A. Sen, and K. Suzumara, eds., *Handbook of Social Choice and Welfare*, vol. I. New York: Elsevier Science.

[22] Milgrom, P. and C. Shannon (1994), “Monotone Comparative Statics,” *Econometrica* 62, 157-180.

[23] O’Donoghue, T., and M. Rabin (1999a), “Doing it Now or Later,” *American Economic Review* 89, 103-124.

[24] O’Donoghue, T., and M. Rabin (1999b), “Incentives for Procrastinators,” *Quarterly Journal of Economics* 114, 769-816.

[25] O’Donoghue, T., and M. Rabin (2001), “Choice and Procrastination,” *Quarterly Journal of Economics* 116, 121-160.

[26] Palfrey, T. (2002), “Implementation Theory,” in R. Aumann and S. Hart, eds., *Handbook of Game Theory with Economic Applications*, vol. III. New York: Elsevier Science.

[27] Rudin, W. (1976), *Principles of mathematical analysis* (3rd edition). New York: McGraw-Hill.

[28] Serrano, R. (2004), "The Theory of Implementation of Social Choice Rules," *SIAM Review* 46, 377-414.

[29] Shui H. and L. Ausubel, (2004), "Consumer Time Inconsistency: Evidence from an Experiment in the Credit Card Market," working paper.

[30] Skiba, P.M., and J. Tobacman (2008), "Payday Loans, Uncertainty, and Discounting: Explaining Patterns of Borrowing, Repayment, and Default," working paper.

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

Appendix: Proofs

Results omitted from main text

Lemma A-1 (The effect of savings on self 2's preferences) *Fix s_1, ϕ, x^a, and x^b such that $x_2^a \leq x_2^b$, $x_2^a + x_3^a \geq x_2^b + x_3^b$, and $\max_{s_2 \geq 0} U^2 (s_1 + x^a - s_2; \phi) \geq (\leq) \max_{s_2 \geq 0} U^2 \left(s_1 + x^b - s_2; \phi\right)$. Then the same is true for all $\tilde{s}_1 \geq (\leq) s_1$.*

Proof. Define

$$f(s_1) = \max_{s_2 \geq 0} U^2 (s_1 + x^a - s_2; \phi) - \max_{s_2 \geq 0} U^2 \left(s_1 + x^b - s_2; \phi\right).$$

If $x^a = x^b$, then $f(s_1) = 0$ for all s_1; the result is immediate. If $x_2^a = x_2^b$ and $x_2^a + x_3^a > x_2^b + x_3^b$, then $f(s_1) > 0$ for all s_1; the result is immediate. If $x_2^a < x_2^b$ and $x_2^a + x_3^a = x_2^b + x_3^b$, then there exists $\hat{s}_1$ such that for all $s_1 \geq \hat{s}_1$, $f(s_1) = 0$ and for all $s_1 < \hat{s}_1$, $f(s_1) < 0$; the result is immediate.

Finally, consider the case where $x_2^a < x_2^b$ and $x_2^a + x_3^a > x_2^b + x_3^b$. By standard arguments f is differentiable, with derivative

$$f'(s_1) = U_2^2 (s_1 + x^a - s_2^a; \phi) - U_2^2 \left(s_1 + x^b - s_2^b; \phi\right),$$

where $s_2^a = \arg\max_{s_2 \geq 0} U^2 (s_1 + x^a - s_2; \phi)$ and $s_2^b = \arg\max_{s_2 \geq 0} U^2 \left(s_1 + x^b - s_2; \phi\right)$. To establish the result, we show that $f'(s_1) > 0$ whenever $f(s_1) = 0$. Fix s_1 such that $f(s_1) = 0$.

If $x_2^a - s_2^a < x_2^b - s_2^b$ the result is immediate; the remainder of the proof deals with the case $x_2^a - s_2^a \geq x_2^b - s_2^b$. Note that in this case $x_2^a < x_2^b$ implies $s_2^b > 0$. Suppose that, contrary to the claimed result, $f'(s_1) \leq 0$. Then

$$U_3^2\left(s_1 + x^a - s_2^a; \phi\right) \leq U_2^2\left(s_1 + x^a - s_2^a; \phi\right) \leq U_2^2\left(s_1 + x^b - s_2^b; \phi\right) = U_3^2\left(s_1 + x^b - s_2^b; \phi\right),$$

where the final equality follows from the fact that s_2^b is an interior solution; hence $x_3^a + s_2^a \geq x_3^b + s_2^b$. As a result, $s_1 + x^b - s_2^b$ offers at least as much consumption as $s_1 + x^a - s_2^a$ at each date and, since $x_2^a + x_3^a > x_2^b + x_3^b$, strictly more at at least one date. But then $f(s_1) < 0$, contradicting our supposition that $f(s_1) = 0$.

Lemma A-2 (Contract Monotonicity) *Suppose that X is a commitment contract. Then, for any $\phi, \tilde{\phi}$, $X_2\left(\phi, s_1; \tilde{\phi}\right)$ is weakly decreasing in s_1 and $X_2\left(\phi, s_1; \tilde{\phi}\right) + X_3\left(\phi, s_1; \tilde{\phi}\right)$ is weakly increasing in s_1.*

Proof. Fix s_1 and $\tilde{s}_1 > s_1$. From self 2's incentive constraints, $U^2(s_1 + X\left(\phi, s_1; \tilde{\phi}\right); \phi) \geq U^2(s_1 + X\left(\phi, \tilde{s}_1; \tilde{\phi}\right); \phi)$ and $U^2(\tilde{s}_1 + X\left(\phi, \tilde{s}_1; \tilde{\phi}\right); \phi) \geq U^2(\tilde{s}_1 + X\left(\phi, s_1; \tilde{\phi}\right); \phi)$, implying

$$U^2(\tilde{s}_1 + X\left(\phi, \tilde{s}_1; \tilde{\phi}\right); \phi) - U^2(s_1 + X\left(\phi, \tilde{s}_1; \tilde{\phi}\right); \phi) \geq U^2(\tilde{s}_1 + X\left(\phi, s_1; \tilde{\phi}\right); \phi) - U^2(s_1 + X\left(\phi, s_1; \tilde{\phi}\right); \phi).$$

Since date-3 consumption is the same under $\tilde{s}_1 + X\left(\phi, \tilde{s}_1; \tilde{\phi}\right)$ and $s_1 + X\left(\phi, \tilde{s}_1; \tilde{\phi}\right)$ on the one hand, and under $\tilde{s}_1 + X\left(\phi, s_1; \tilde{\phi}\right)$ and $s_1 + X\left(\phi, s_1; \tilde{\phi}\right)$ on the other, it follows from concavity of preferences that $X_2\left(\phi, \tilde{s}_1; \tilde{\phi}\right) \leq X_2\left(\phi, s_1; \tilde{\phi}\right)$. Second, self 2's incentive constraints also imply $U^2(\tilde{s}_1 + X\left(\phi, \tilde{s}_1; \tilde{\phi}\right); \phi) \geq U^2(\tilde{s}_1 + X\left(\phi, s_1; \tilde{\phi}\right) - s_2; \phi)$ for $s_2 = X_2\left(\phi, s_1; \tilde{\phi}\right) - X_2\left(\phi, \tilde{s}_1; \tilde{\phi}\right) \geq 0$. Since date-2 consumption is the same under $\tilde{s}_1 + X\left(\phi, \tilde{s}_1; \tilde{\phi}\right)$ and $\tilde{s}_1 + X_2\left(\phi, s_1; \tilde{\phi}\right) - s_2$, it follows that date-3 consumption is greater under the former, i.e., $X_3\left(\phi, \tilde{s}_1; \tilde{\phi}\right) \geq X_3\left(\phi, s_1; \tilde{\phi}\right) + s_2$. Substituting in for s_2 implies $X_2\left(\phi, \tilde{s}_1; \tilde{\phi}\right) + X_3\left(\phi, \tilde{s}_1; \tilde{\phi}\right) \geq X_2\left(\phi, s_1; \tilde{\phi}\right) + X_3\left(\phi, s_1; \tilde{\phi}\right)$, completing the proof.

Lemma A-3 (No saving from equilibrium consumption) *If commitment to $\{c(\phi)\}_\phi$ is possible then i) $U^1(c(\phi); \phi) \geq \max_{s_1 \geq 0} U^1(s_1 + c(\phi) - \hat{s}_2; \phi)$, where $\hat{s}_2 = \arg\max_{s_2 \geq 0} U^2(s_1 + c(\phi) - s_2; \phi)$, and ii) $U^2(c(\phi); \phi) \geq \max_{s_2 \geq 0} U^2(c(\phi) - s_2; \phi)$.*

Proof. Let X be a commitment contract. ii) is immediate from self 2's IC constraint with $\tilde{\phi}^1 = \tilde{\phi}^2 = \phi$ and $\tilde{s}_1 = s_1 = 0$. For i), note first that, again from self 2's IC constraint, for any $s_1 \geq 0$, $U^2(s_1 + X(\phi, s_1; \phi); \phi) \geq U^2(s_1 + X(\phi, 0; \phi) - \hat{s}_2; \phi)$, where $\hat{s}_2 = \arg\max_{s_2 \geq 0} U^2(s_1 + X(\phi, 0; \phi) - s_2; \phi)$.

Provided either $X_2(\phi, s_1; \phi) \leq X_2(\phi, 0; \phi) - \hat{s}_2$ or $X_3(\phi, s_1; \phi) \geq X_3(\phi, 0; \phi) + \hat{s}_2$, Lemma 2 implies that $U^1(s_1 + X(\phi, s_1; \phi); \phi) \geq U^1(s_1 + X(\phi, 0; \phi) - \hat{s}_2; \phi)$. From self 1's IC constraint with $\tilde{\phi}^1 = \tilde{\phi}^2 = \phi$, $U^1(X(\phi, 0; \phi); \phi) \geq U^1(s_1 + X(\phi, s_1; \phi); \phi)$. Substituting in $X(\phi, 0; \phi) = c(\phi)$ then implies i).

Consequently, it remains only to show that the case $X_2(\phi, s_1; \phi) > X_2(\phi, 0; \phi) - \hat{s}_2$ and $X_3(\phi, s_1; \phi) < X_3(\phi, 0; \phi) + \hat{s}_2$ cannot arise. Suppose to the contrary that both these inequalities hold. From Lemma A-2, $X_2(\phi, s_1; \phi) \leq X_2(\phi, 0; \phi)$, implying $\hat{s}_2 > 0$. But then self 2 would want to save given the consumption plan $s_1 + X(\phi, s_1; \phi)$, which violates self 2's IC constraint. The contradiction completes the proof.

Lemma A-4 (Necessary conditions for commitment) *If commitment to $\{c(\phi)\}_\phi$ is possible then $U_2^2(s_1^* + c'; \theta) \geq U_3^2(s_1^* + c'; \theta)$.*

Proof. Let $\hat{X}$ be a commitment contract. First suppose, to the contrary, that $U_2^2(s_1^* + c'; \theta) < U_3^2(s_1^* + c'; \theta)$. Note that, by continuity of preferences and the definition of s_1^*, there then exists $s_1 < s_1^*$ such that $U_2^2(s_1 + c'; \theta) < U_3^2(s_1 + c'; \theta)$ and $U^1(s_1 + c' - \hat{s}_2; \theta) > U^1(s_1 + c; \theta)$ where, just as in the definition of s_1^*, $\hat{s}_2 = \arg\max_{s_2 \geq 0} U^2(s_1 + c' - s_2; \theta)$. From (IC$_1$), $U^1(s_1 + c; \theta) \geq U^1(s_1 + \hat{X}(s_1); \theta)$ and therefore $U^1(s_1 + c' - \hat{s}_2; \theta) > U^1(s_1 + \hat{X}(s_1); \theta)$; from (IC$_2$) and the fact that $X(0) = c'$, $U^2(s_1 + \hat{X}(s_1); \theta) \geq U^2(s_1 + c' - \hat{s}_2; \theta)$; it then follows from Lemma 2 that $\hat{X}_2(s_1) > c_2' - \hat{s}_2$ and $\hat{X}_3(s_1) < \hat{s}_2 + c_3'$. From the definition of $\hat{s}_2$ and the fact that $U_2^2(s_1 + c'; \theta) < U_3^2(s_1 + c'; \theta)$, we have $U_2^2(s_1 + c' - \hat{s}_2; \theta) = U_3^2(s_1 + c' - \hat{s}_2; \theta)$. But then $U_2^2(s_1 + \hat{X}(s_1); \theta) < U_3^2(s_1 + \hat{X}(s_1); \theta)$, violating (IC$_2$) and contradicting our supposition that $\hat{X}$ is a commitment contract.

Lemma A-5 (Extension of SCB to self 2's indirect utility function with saving) *Fix s_1 and x such that $x_2 > c_2'$, $U^2(s_1 + x; \theta) \geq U^2(s_1 + c'; \theta)$, and $U^2(c'; \theta') \geq U^2(x - s_2; \theta')$ for all $s_2 \geq 0$. Fix $\tilde{x}$ such that $\tilde{x}_2 \geq x_2$, $U^2(s_1 + \tilde{x}; \theta) = U^2(s_1 + c'; \theta)$, and $U_2^2(s_1 + \tilde{x}; \theta) \geq U_3^2(s_1 + \tilde{x}; \theta)$. Then $U^2(c'; \theta') \geq U^2(\tilde{x} - s_2; \theta')$ for all $s_2 \geq 0$.*

Proof. We first show that $U^2(c'; \theta') \geq U^2(\tilde{x} - s_2; \theta')$ for all $s_2 \in [0, \tilde{x}_2 - x_2]$. Note that $U^2(s_1 + \tilde{x}; \theta) = U^2(s_1 + c'; \theta)$ and $U_2^2(s_1 + \tilde{x}; \theta) \geq U_3^2(s_1 + \tilde{x}; \theta)$ imply that $U^2(s_1 + c'; \theta) \geq U^2(s_1 + \tilde{x} - s_2; \theta)$ for all $s_2 \geq 0$, in particular for all $s_2 \in [0, \tilde{x}_2 - x_2]$. Therefore, since $\tilde{x}_2 - s_2 \geq x_2$ for all $s_2 \in [0, \tilde{x}_2 - x_2]$, it follows from SCB that $U^2(c'; \theta') \geq U^2(\tilde{x} - s_2; \theta')$ for all $s_2 \in [0, \tilde{x}_2 - x_2]$.

We next show that $U^2(c'; \theta') \geq U^2(\tilde{x} - s_2; \theta')$ for all $s_2 \geq \tilde{x}_2 - x_2$. Note that $\tilde{x}_2 \geq x_2$, $U^2(s_1 + x; \theta) \geq U^2(s_1 + c'; \theta) = U^2(s_1 + \tilde{x}; \theta)$, and $U_2^2(s_1 + \tilde{x}; \theta) \geq U_3^2(s_1 + \tilde{x}; \theta)$ imply that

$x_2 + x_3 \geq \tilde{x}_2 + \tilde{x}_3$. Therefore it follows from $U^2(c';\theta') \geq U^2(x - s_2;\theta')$ for all $s_2 \geq 0$ that $U^2(c';\theta') \geq U^2(\tilde{x} - s_2;\theta')$ for all $s_2 \geq \tilde{x}_2 - x_2$.

Lemma A-6 (s_1^* under self 0's most preferred consumption plan) *Under self 0's most preferred consumption plan* $\{c^*(\phi)\}_\phi$, $s_1^* \leq c_1'^* - c_1^*$.

Proof. It is sufficient to show that for all $s_1 > c_1'^* - c_1^*$ and $s_2 \geq 0$, $U^1(c^*;\theta) \geq U^1(s_1 + c'^* - s_2)$ or, equivalently, $u(c_1^*;\theta) - u(c_1'^* - s_1;\theta) \geq \beta u(s_1 + c_2'^* - s_2;\theta) - \beta u(c_2^*;\theta) + \beta u(s_2 + c_3'^*;\theta) - \beta u(c_3^*;\theta)$. By the definition of c^*, this inequality holds for $\beta = 1$; since $c_1^* > c_1'^* - s_1$ it follows immediately that it must also hold for all $\beta < 1$.

Proofs of results stated in main text

Proof of Lemma 1. We prove the result for the case of unverifiable savings (see page 14). For the case of verifiable savings, simply restrict s_1 and s_2 to zero below.

If $\{c(\phi)\}_\phi$ generates a commitment problem in state ϕ, then $U^1(c(\phi);\phi) < U^1(\hat{s}_1 + c(\phi') - \hat{s}_2(\hat{s}_1);\phi)$, where $\phi' \neq \phi$, $\hat{s}_1 \equiv \arg\max_{s_1 \geq 0} U^1(s_1 + c(\phi') - \hat{s}_2(s_1);\phi)$, and $\hat{s}_2(s_1) = \arg\max_{s_2 \geq 0} U^2(s_1 + c(\phi') - s_2;\phi)$. By assumption, $U^0(c(\phi);\phi) \geq U^0(\hat{s}_1 + c(\phi') - \hat{s}_2(\hat{s}_1);\phi)$. Then, since $U^0(\hat{s}_1 + c(\phi') - \hat{s}_2(\hat{s}_1);\phi) - U^0(c(\phi);\phi) = U^1(\hat{s}_1 + c(\phi') - \hat{s}_2(\hat{s}_1);\phi) - U^1(c(\phi);\phi) - (1-\beta)(u_1(c_1(\phi') - \hat{s}_1) - u_1(c_1(\phi)))$, it follows immediately that $c_1(\phi') - \hat{s}_1 > c_1(\phi)$ and so, in particular, that $\{c(\phi)\}_\phi$ cannot generate a commitment problem in state ϕ'.

We finally show that if $c_1(\phi') > c_1(\phi)$ and β is small enough, then $\{c(\phi)\}_\phi$ generates a commitment problem in state ϕ or, equivalently, $u(c_1(\phi') - \hat{s}_1;\phi) - u(c_1(\phi);\phi) > \beta u(c_2(\phi);\phi) + \beta u(c_3(\phi);\phi) - \beta u(\hat{s}_1 + c_2(\phi') - \hat{s}_2(\hat{s}_1);\phi) - \beta u(\hat{s}_2 + c_3(\phi');\phi)$. As β approaches zero, the right hand side approaches zero while the left hand side approaches $u_1(c_1(\phi')) - u_1(c_1(\phi)) > 0$, where the latter follows from the fact that $\hat{s}_1 = 0$ for β small enough.

Proof of Lemma 2. The proof is immediate from the fact that $U^1(\tilde{c}^a;\phi) - U^1(\tilde{c}^b;\phi) = U^2(\tilde{c}^a;\phi) - U^2(\tilde{c}^b;\phi) - (1-\beta)(u_2(\tilde{c}_2^a;\phi) - u_2(\tilde{c}_2^b;\phi))$.

Proof of Lemma 3. By the construction of $X(\phi, s_1;\theta)$, (IC$_2$) is satisfied for $\tilde{\phi}^1 = \theta$. (IC$_1$) for $\tilde{\phi}^1 = \theta$ is simply $U^1(c(\phi);\phi) \geq \max_{s_1 \geq 0} U^1(s_1 + c - \hat{s}_2(s_1);\phi)$, where $\hat{s}_2(s_1) = \arg\max_{s_2 \geq 0} U^2(s_1 + c - s_2;\phi)$.

For $\phi = \theta$ this inequality is satisfied by Lemma A-3. For $\phi = \theta'$ it is satisfied because we know (c, c') does not generate a commitment problem in state θ'.

Proof of Proposition 2. **Claim 1** *For all* $s_1 \in [0, s_1^*)$, $\hat{X}_2(s_1) > c_2'$.

Proof of Claim 1. Note that by Lemma A-4 and concavity of preferences, $s_2 = 0$ solves $\max_{s_2 \geq 0} U^2(s_1 + c' - s_2; \theta)$ for all $s_1 \in [0, s_1^*]$. It then follows from continuity of preferences and the definition of s_1^* that $U^1(s_1^* + c'; \theta) = U^1(c; \theta)$. Together with the fact that there is a commitment problem and the concavity of $U^1(s_1 + c'; \theta)$ in s_1, this implies that there exists $\hat{s}_1 \in [0, s_1^*)$ such that for all $s_1 \in [\hat{s}_1, s_1^*)$, $U^1(s_1 + c'; \theta) > U^1(c; \theta)$.

Fix any $s_1 \in [\hat{s}_1, s_1^*)$. From (IC$_1$) for $\left(\phi, \tilde{\phi}^1\right) = (\theta, \theta')$, $U^1(c; \theta) \geq U^1(s_1 + \hat{X}(s_1); \theta)$ and therefore $U^1(s_1 + c'; \theta) > U^1(s_1 + \hat{X}(s_1); \theta)$; from (IC$_2$) for $\left(\phi, \tilde{\phi}^1, \tilde{\phi}^2\right) = (\theta, \theta', \theta')$ and the fact that $X(\theta', 0; \theta') = c'$, $U^2(s_1 + \hat{X}(s_1); \theta) \geq U^2(s_1 + c'; \theta)$; it then follows from Lemma 2 that $\hat{X}_2(s_1) > c_2'$.

From Lemma A-2, it follows that $\hat{X}_2(s_1) > c_2'$ for all $s_1 \in [0, s_1^*)$.

Claim 2 *For all* $s_1 \in [0, s_1^*)$, $\hat{X}_2(s_1) + \hat{X}_3(s_1) \leq c_2' + c_3'$.

Proof of Claim 2. Fix $s_1 \in [0, s_1^*)$ and suppose, contrary to the claimed result, that $\hat{X}_2(s_1) + \hat{X}_3(s_1) > c_2' + c_3'$. By Claim 1, $\hat{X}_2(s_1) > c_2'$. It then follows that $U^2(\hat{X}(s_1) - s_2; \theta') > U^2\left(c'; \theta'\right)$ for $s_2 = \hat{X}_2(s_1) - c_2' > 0$, violating (IC$_2$) for $\left(\phi, \tilde{\phi}^1, \tilde{\phi}^2\right) = (\theta', \theta', \theta)$ and contradicting our supposition that $\hat{X}$ is a commitment contract.

Claim 3 *If* $\hat{X}$ *is continuous at* s_1*, then it satisfies (2).*

Proof of Claim 3. It follows immediately from (1) that for all $s_1, \tilde{s}_1 \geq 0$,

$$\begin{aligned} & U^2(\tilde{s}_1 + \hat{X}(\tilde{s}_1); \theta) - U^2(s_1 + \hat{X}(\tilde{s}_1); \theta) \\ \geq\ & U^2(\tilde{s}_1 + \hat{X}(\tilde{s}_1); \theta) - U^2(s_1 + \hat{X}(s_1); \theta) \geq U^2(\tilde{s}_1 + \hat{X}(s_1); \theta) - U^2(s_1 + \hat{X}(s_1); \theta). \end{aligned} \tag{A-1}$$

Suppose that $\tilde{s}_1 > s_1$ and divide (A-1) everywhere by $\tilde{s}_1 - s_1$, yielding

$$\begin{aligned} \frac{U^2\left(\tilde{s}_1 + \hat{X}(\tilde{s}_1); \theta\right) - U^2\left(s_1 + \hat{X}(\tilde{s}_1); \theta\right)}{\tilde{s}_1 - s_1} &\geq \frac{U^2(\tilde{s}_1 + \hat{X}(\tilde{s}_1); \theta) - U^2(s_1 + \hat{X}(s_1); \theta)}{\tilde{s}_1 - s_1} \\ &\geq \frac{U^2\left(\tilde{s}_1 + \hat{X}(s_1); \theta\right) - U^2\left(s_1 + \hat{X}(s_1); \theta\right)}{\tilde{s}_1 - s_1}. \end{aligned}$$

If $\hat{X}$ is continuous at s_1, then the upper bound and the lower bound both converge to $U_2^2(s_1 +$

$\hat{X}(s_1);\theta)$ as $\tilde{s}_1 \to s_1$; it follows that $dU^2(s_1 + \hat{X}(s_1);\theta) = U_2^2(s_1 + \hat{X}(s_1);\theta)ds_1$. Equation (2) then follows from the identity $dU^2(s_1 + \hat{X}(s_1);\theta) = U_2^2 ds_1 + U_2^2 d\hat{X}_2 + U_3^2 d\hat{X}_3$.

Proof of Proposition 3. Suppose that commitment is possible and define, for all $s_1 \in [0, s_1^*)$,

$$
\begin{aligned}
Y(s_1) = \{\tilde{x} : \tilde{x}_1 = c_1',\ \tilde{x}_2 > c_2',\ & U_2^2(s_1 + \tilde{x};\theta) \geq U_3^2(s_1 + \tilde{x};\theta), \\
& U^2(s_1 + \tilde{x};\theta) \geq U^2(s_1 + c';\theta), \\
& U^2(c';\theta') \geq U^2(\tilde{x} - s_2;\theta') \text{ for all } s_2 \geq 0\}.
\end{aligned}
$$

SPR is satisfied provided that the intersection $\bigcap_{s_1 \in [0,s_1^*)} Y(s_1)$ is nonempty. To establish nonemptiness, let $B_\varepsilon(c')$ be the open set consisting of all points lying within ε of c'. As we show in Claims 1 and 2 below, there exists $\varepsilon > 0$ such that for all $s_1 \in [0, s_1^*)$, $Y(s_1) \backslash B_\varepsilon(c')$ is compact and nonempty. It follows that the intersection $\bigcap_{s_1 \in [0,s_1^*)} Y(s_1) \backslash B_\varepsilon(c')$ is nonempty (see, e.g., Theorem 2.36 in Rudin 1976).

To see why a nonempty intersection $\bigcap_{s_1 \in [0,s_1^*)} Y(s_1)$ implies that SPR is satisfied, let x be a member of this intersection, so for all $s_1 \in [0, s_1^*)$, $x \in Y(s_1)$ and so, in particular, $U^2(s_1 + x;\theta) \geq U^2(s_1 + c';\theta)$. Hence, by continuity of preferences, $U^2(s_1^* + x;\theta) \geq U^2(s_1^* + c';\theta)$. Moreover, since for some s_1, $x \in Y(s_1)$, we also have $x_2 > c_2'$ and $U^2(c';\theta') \geq U^2(x - s_2;\theta')$ for all $s_2 \geq 0$. Hence SPR is satisfied.

Claim 1 *For all $s_1 \in [0, s_1^*)$, $Y(s_1)$ is nonempty and bounded, and the collection $\{Y(s_1)\}_{s_1 \in [0,s_1^*)}$ is nested (i.e., $Y(s_1) \subset Y(\tilde{s}_1)$ for $\tilde{s}_1 < s_1$).*

Proof of Claim 1. The set $Y(s_1)$ is nonempty since commitment is possible (see Proposition 2(i)). To see that the sets are nested, rewrite the two conditions involving s_1 in the definition of $Y(s_1)$ as $u_2'(s_1 + \tilde{x}_2;\theta) \geq \beta u_3'(\tilde{x}_3;\theta)$ and $u_2(s_1 + \tilde{x}_2;\theta) - u_2(s_1 + c_2';\theta) \geq \beta u_3(c_3';\theta) - \beta u_3(\tilde{x}_3;\theta)$. If each of these conditions is satisfied for some s_1 then, by concavity of preferences, the same is true for all lower values of s_1. Finally, boundedness also follows from $u_2'(s_1 + \tilde{x}_2;\theta) \geq \beta u_3'(\tilde{x}_3;\theta)$: for any $\tilde{x} \in Y(s_1)$, $\tilde{x}_2 > c_2'$ and $\tilde{x}_3 < c_3'$, and so $u_2'(s_1 + \tilde{x}_2;\theta) > \beta u_3'(c_3';\theta)$ and $u_3'(\tilde{x}_3;\theta) < \beta u_2'(s_1 + c_2';\theta)$. Given the Inada conditions, these inequalities establish an upper bound for $\tilde{x}_2$ and a lower bound for $\tilde{x}_3$.

Claim 2 *There exists $\varepsilon > 0$ such that for all $s_1 \in [0, s_1^*)$, $Y(s_1) \backslash B_\varepsilon(c')$ is nonempty.*

Proof of Claim 2. We show that $Y(s_1) \backslash B_\varepsilon(c')$ is nonempty for all s_1 sufficiently close to s_1^*; the claim then follows from nestedness.

First, consider the case $U_2^2(s_1^* + c'; \theta) / U_3^2(s_1^* + c'; \theta) < U_2^2(c'; \theta') / U_3^2(c'; \theta')$. Let $\hat{s}_1 \in [0, s_1^*)$ be such that $U_2^2(\hat{s}_1 + c'; \theta) / U_3^2(\hat{s}_1 + c'; \theta) < U_2^2(c'; \theta') / U_3^2(c'; \theta')$. Consequently, there exists $\varepsilon > 0$ such that no element of $Y(\hat{s}_1)$ lies in $B_\varepsilon(c')$. By the nestedness of the sets $Y(s_1)$, the same is true for all $s_1 \in [\hat{s}_1, s_1^*)$. Hence , $Y(s_1) \backslash B_\varepsilon(c') = Y(s_1) \neq \varnothing$ for all $s_1 \in [\hat{s}_1, s_1^*)$.

Second, consider the case $U_2^2(s_1^* + c'; \theta) / U_3^2(s_1^* + c'; \theta) \geq U_2^2(c'; \theta') / U_3^2(c'; \theta')$. Since, by supposition, commitment is possible, it follows from (IC_2) for $\left(\phi, \tilde{\phi}^1, \tilde{\phi}^2, s_1\right) = \left(\theta', \theta', \theta', 0\right)$ and the fact that $X\left(\theta', 0, \theta'\right) = c'$ that $U_2^2(c'; \theta') / U_3^2(c'; \theta') \geq 1$ and so there exists $\hat{s}_1 \in [0, s_1^*)$ such that $U_2^2(\hat{s}_1 + c'; \theta) / U_3^2(\hat{s}_1 + c'; \theta) > 1$. Choose $\lambda > 1$ such that (for any $\varepsilon > 0$) the point $\left(c_1', c_2' + \frac{3}{2}\varepsilon, c_3' - (U_2^2(\hat{s}_1 + c'; \theta) / U_3^2(\hat{s}_1 + c'; \theta)) \frac{3}{2}\varepsilon\right)$ lies within $\lambda\varepsilon$ of c'. Finally, choose $\varepsilon > 0$ such that for all $s_1 \in [\hat{s}_1, s_1^*)$ and all $x \in B_{\lambda\varepsilon}(c')$, $U_2^2(s_1 + x; \theta) / U_3^2(s_1 + x; \theta) > 1$.

Now, consider any $s_1 \in [\hat{s}_1, s_1^*)$. If no member of $Y(s_1)$ lies within ε of c', it is immediate that $Y(s_1) \backslash B_\varepsilon(c')$ is nonempty. Suppose instead that there exists $x \in Y(s_1)$ that lies within ε of c'. Define $\tilde{x}$ by $\tilde{x}_1 = c_1'$, $\tilde{x}_2 = c_2' + \frac{3}{2}\varepsilon$, and $U^2(s_1 + \tilde{x}; \theta) = U^2(s_1 + c'; \theta)$; thus defined, $\tilde{x}$ lies within $\lambda\varepsilon$ of c', and so $U_2^2(s_1 + \tilde{x}; \theta) / U_3^2(s_1 + \tilde{x}; \theta) > 1$. From Lemma A-5, it follows that $U^2\left(c'; \theta'\right) \geq U^2\left(\tilde{x} - s_2; \theta'\right)$ for all $s_2 \geq 0$ and hence $\tilde{x} \in Y(s_1) \backslash B_\varepsilon(c')$, completing the proof of the claim.

Proof of Lemma 4. Suppose that $U_2^2(s_1^* + c'; \theta) / U_3^2(s_1^* + c'; \theta) > 1$ and $U_2^2(s_1^* + c'; \theta) / U_3^2(s_1^* + c'; \theta) \geq U_2^2(c'; \theta') / U_3^2(c'; \theta')$ but, to the contrary, SPR is violated. Recall that, by assumption, $U_2^2(c'; \theta') / U_3^2(c'; \theta') \geq 1$. There are two cases to consider.

First suppose that $U_2^2(c'; \theta') / U_3^2(c'; \theta') > 1$; then there exists $\tilde{x}$ such that $\tilde{x}_2 > c_2'$ and $U^2\left(c'; \theta'\right) = U^2\left(\tilde{x}; \theta'\right) \geq U^2\left(\tilde{x} - s_2; \theta'\right)$ for all $s_2 \geq 0$. Since, by supposition, SPR is violated, $U^2(s_1^* + c'; \theta) > U^2(s_1^* + \tilde{x}; \theta)$ and so there exists $s_1 \in [0, s_1^*)$ such that $U^2(s_1 + c'; \theta) > U^2(s_1 + \tilde{x}; \theta)$ (note that $s_1^* > 0$ whenever there is a commitment problem). By adding date-3 consumption to $\tilde{x}$, one obtains $\tilde{x}'$ such that $\tilde{x}_2' > c_2'$, $U^2(s_1 + c'; \theta) = U^2(s_1 + \tilde{x}'; \theta)$, and $U^2\left(c'; \theta'\right) < U^2\left(\tilde{x}'; \theta'\right)$. Since $U_2^2(s_1^* + c'; \theta) / U_3^2(s_1^* + c'; \theta) \geq U_2^2(c'; \theta') / U_3^2(c'; \theta')$ and $s_1 < s_1^*$, then $U_2^2(s_1 + c'; \theta) / U_3^2(s_1 + c'; \theta) > U_2^2(c'; \theta') / U_3^2(c'; \theta')$. But then $U^2(s_1 + c'; \theta) = U^2(s_1 + \tilde{x}'; \theta)$ and SCB imply that $U^2\left(c'; \theta'\right) \geq U^2\left(\tilde{x}'; \theta'\right)$, a contradiction.

Next suppose that $U_2^2(c'; \theta') / U_3^2(c'; \theta') = 1$. Since, by supposition, $U_2^2(s_1^* + c'; \theta) / U_3^2(s_1^* + c'; \theta) > 1$ there exists $\tilde{x}$ such that $\tilde{x}_2 > c_2'$, $U^2(s_1^* + c'; \theta) = U^2(s_1^* + \tilde{x}; \theta)$, and $\tilde{x}_2 + \tilde{x}_3 < c_2' + c_3'$. Since

$U_2^2\left(c';\theta'\right)/U_3^2\left(c';\theta'\right)=1$ and $\tilde{x}_2+\tilde{x}_3<c_2'+c_3'$ we know $U^2\left(c';\theta'\right)>U^2\left(\tilde{x}-s_2;\theta'\right)$ for all s_2, contradicting our supposition that SPR is violated.

To show the converse in the case of additive shocks, note that, in this case, s_1 and θ_2 enter self 2's utility function $u_2\left(s_1+x_2+\theta_2\right)+\beta u_3\left(x_3+\theta_3\right)$ only through their sum $s_1+\theta_2$. Hence, by standard single crossing, self 2's state-θ indifference curve through any point x with inherited savings s_1 crosses at most once with self 2's state-θ' indifference curve through x with no savings. It follows immediately that $U_2^2\left(s_1^*+c';\theta\right)/U_3^2\left(s_1^*+c';\theta\right)\geq U_2^2(c';\theta')/U_3^2(c';\theta')$ is a necessary condition for SPR.

Finally, we show that $U_2^2\left(s_1^*+c';\theta\right)/U_3^2\left(s_1^*+c';\theta\right)>1$ is a necessary condition for SPR. To establish the contrapositive, suppose that $1\geq U_2^2\left(s_1^*+c';\theta\right)/U_3^2\left(s_1^*+c';\theta\right)$. Hence any $\tilde{x}$ such that $\tilde{x}_2>c_2'$ and $U^2\left(s_1^*+\tilde{x};\theta\right)\geq U^2\left(s_1^*+c';\theta\right)$ must satisfy $\tilde{x}_2+\tilde{x}_3>c_2'+c_3'$. But then $U^2\left(\tilde{x}-s_2;\theta'\right)>U^2\left(c';\theta'\right)$ for $s_2=\tilde{x}_2-c_2'>0$, implying that SPR is not satisfied.

Proof of Lemma 5. Note that $U_2^2\left(s_1^*+c';\theta\right)/U_3^2\left(s_1^*+c';\theta\right)>1$ if and only if $\hat{x}_2^0>c_2'$. Given this, sufficiency is immediate. For necessity, first suppose that $U_2^2\left(s_1^*+c';\theta\right)/U_3^2\left(s_1^*+c';\theta\right)\leq 1$ and so $\hat{x}_2^0\leq c_2'$. By the definition of $\hat{x}^0$, this implies that for all $\tilde{x}$ such that $\tilde{x}_2>c_2'$ and $U^2\left(s_1^*+\tilde{x};\theta\right)\geq U^2\left(s_1^*+c';\theta\right)$, $\tilde{x}_2+\tilde{x}_3>c_2'+c_3'$. As a result, for every such $\tilde{x}$, $U^2\left(c';\theta'\right)<U^2\left(\tilde{x}-s_2;\theta'\right)$ for $s_2=\tilde{x}_2-c_2'$, a violation of SPR.

Second, suppose that $U_2^2\left(s_1^*+c';\theta\right)/U_3^2\left(s_1^*+c';\theta\right)>1$ (and so $\hat{x}_2^0>c_2'$) but that $U^2\left(c';\theta'\right)<U^2\left(\hat{x}^0-s_2^+;\theta'\right)$ for some $s_2^+\geq 0$. Note that by the definition of $\hat{x}^0$, $\tilde{x}_2+\tilde{x}_3\geq\hat{x}_2^0+\hat{x}_3^0$ for all $\tilde{x}$ such that $U^2\left(s_1^*+\tilde{x};\theta\right)\geq U^2\left(s_1^*+c';\theta\right)$ and $\tilde{x}_2\geq c_2'$. We show that SPR is violated by showing that no such $\tilde{x}$ satisfies (SPRb) and $\tilde{x}_2>c_2'$.

On the one hand, if $\tilde{x}_2\geq\hat{x}_2^0-s_2^+$, then (SPRb) is violated since $U^2\left(c';\theta'\right)<U^2\left(\tilde{x}-s_2;\theta'\right)$ when $s_2\geq 0$ is set such that $\tilde{x}_2-s_2=\hat{x}_2^0-s_2^+$. On the other hand, if $\tilde{x}_2<\hat{x}_2^0-s_2^+$, then (SPRb) is violated since, as we next show, $U^2\left(c';\theta'\right)<U^2\left(\tilde{x};\theta'\right)$. Suppose to the contrary that $U^2\left(c';\theta'\right)\geq U^2\left(\tilde{x};\theta'\right)$. By the definition of $\hat{x}^0$, $U^2\left(s_1^*+\hat{x}^0-s_2^+;\theta\right)\leq U^2\left(s_1^*+c';\theta\right)$. So SCB implies that $U^2\left(c';\theta'\right)\geq U^2\left(\hat{x}^0-s_2^+;\theta'\right)$, contradicting our supposition that $U^2\left(c';\theta'\right)<U^2\left(\hat{x}^0-s_2^+;\theta'\right)$.

Proof of Proposition 4. **Claim** *If* $\hat{X}_2^*\left(s_1^*\right)>c_2'$ *then* $U_2^2(s_1^*+\hat{X}^*\left(s_1^*\right);\theta)\geq U_3^2(s_1^*+\hat{X}^*\left(s_1^*\right);\theta)$.

Proof of Claim. Suppose to the contrary that $U_2^2(s_1^*+\hat{X}^*\left(s_1^*\right);\theta)<U_3^2(s_1^*+\hat{X}^*\left(s_1^*\right);\theta)$ and define $\tilde{x}\left(\tilde{x}_2\right)$ by $\tilde{x}_1\left(\tilde{x}_2\right)=c_1'$, $\tilde{x}_2\left(\tilde{x}_2\right)=\tilde{x}_2$, and $U^2(s_1^*+\tilde{x}\left(\tilde{x}_2\right);\theta)=U^2(s_1^*+\hat{X}^*\left(s_1^*\right);\theta)=U^2(s_1^*+c';\theta)$. There then exists $\varepsilon>0$ such that $\hat{X}_2^*\left(s_1^*\right)-\varepsilon>c_2'$ and $\tilde{x}_2(\hat{X}_2^*\left(s_1^*\right)-\varepsilon)+\tilde{x}_3(\hat{X}_2^*\left(s_1^*\right)-\varepsilon)<$

$\hat{X}_2^*(s_1^*) + \hat{X}_3^*(s_1^*)$. As a result, $U^2(c';\theta') > U^2(\tilde{x}(\hat{X}_2^*(s_1^*) - \varepsilon) - s_2;\theta')$ for all $s_2 \geq 0$, contradicting the definition of $\hat{X}^*(s_1^*)$.

Main step of proof. Suppose to the contrary that for some $s_1^+ < s_1^*$, $\hat{X}_2(s_1^+) < \hat{X}_2^*(s_1^*)$ and so $\hat{X}_2^*(s_1^*) > c_2'$ by Proposition 2. For any $\tilde{s}_1$ define $\tilde{x}(\tilde{s}_1)$ by $\tilde{x}_1 = c_1'$, $\tilde{x}_2(\tilde{s}_1) = \hat{X}_2(s_1^+)$, and $U^2(\tilde{s}_1 + \tilde{x}(\tilde{s}_1)) = U^2(\tilde{s}_1 + c';\theta)$. By the definition of $\hat{X}^*(s_1^*)$, $U^2(c';\theta') < U^2(\tilde{x}(s_1^*) - s_2;\theta')$ for some $s_2 \geq 0$. By continuity of preferences, there thus exist $s_1 \in (s_1^+, s_1^*)$ and $s_2 \geq 0$ such that $U^2(c';\theta') < U^2(\tilde{x}(s_1) - s_2;\theta')$.

Since $s_1 > s_1^+$, Proposition 2 implies that $c_2' < \hat{X}_2(s_1) \leq \hat{X}_2(s_1^+) = \tilde{x}_2(s_1)$. By (IC$_2$) for $(\phi, \tilde{\phi}^1, \tilde{\phi}^2) = (\theta, \theta', \theta')$, $U^2(s_1 + \hat{X}(s_1);\theta) \geq U^2(s_1 + c';\theta)$, and by (IC$_2$) for $(\phi, \tilde{\phi}^1, \tilde{\phi}^2) = (\theta', \theta', \theta)$, $U^2(c';\theta') \geq U^2(\hat{X}(s_1) - s_2;\theta')$ for all $s_2 \geq 0$. By the above claim, $U_2^2(s_1^* + \hat{X}^*(s_1^*);\theta) \geq U_3^2(s_1^* + \hat{X}^*(s_1^*);\theta)$. Together with $\tilde{x}_2(s_1^*) = \hat{X}_2(s_1^+) < \hat{X}_2^*(s_1^*)$ and $U^2(s_1^* + \tilde{x}(s_1^*)) = U^2(s_1^* + c';\theta) = U^2(s_1^* + \hat{X}^*(s_1^*);\theta)$, this implies $U_2^2(s_1^* + \tilde{x}(s_1^*);\theta) \geq U_3^2(s_1^* + \tilde{x}(s_1^*);\theta)$. Finally note that since for any $\tilde{s}_1$, $U^2(\tilde{s}_1 + \tilde{x}(\tilde{s}_1)) = U^2(\tilde{s}_1 + c';\theta)$ and $\tilde{x}_2(\tilde{s}_1) = \tilde{x}_2(s_1^*) > c_2'$, we must have $U_2^2(s_1 + \tilde{x}(s_1);\theta) \geq U_3^2(s_1 + \tilde{x}(s_1);\theta)$ for $s_1 < s_1^*$. (This is due to the clockwise rotation of the indifference curve through c' as $\tilde{s}_1$ decreases.) Since $\tilde{x}(s_1)$ satisfies all the conditions of Lemma A-5 with $x = \hat{X}(s_1)$, it follows that $U^2(c';\theta') \geq U^2(\tilde{x}(s_1) - s_2;\theta')$ for all $s_2 \geq 0$, contradicting the prior paragraph.

Proof of Lemma 6. We show that for all $\varepsilon > 0$, there exists $\tilde{x}$ such that $c_2' < \tilde{x}_2 < c_2' + \varepsilon$ and both (SPRa) and (SPRb) are satisfied. Suppose to the contrary that there exists $\varepsilon > 0$ such that this is not the case. An easy adaptation of the proof of Lemma 4 gives a contradiction.

Proof of Theorem 1 (sufficiency) Lemma 3 establishes that (IC$_1$) and (IC$_2$) are satisfied for $\tilde{\phi}^1 = \theta$ by any contract in which self 2 gets (c_2, c_3) after self 1 chooses c_1.

In Claims 1-5 we show that if c' satisfies SPR and $\hat{X}^*$ satisfies NS, then (IC$_1$) and (IC$_2$) are satisfied for $\tilde{\phi}^1 = \theta'$ by any contract in which, after self 1 chooses c_1' and saves s_1, self 2 gets $\hat{X}^*(s_1)$ in state θ and $X^*(s_1)$ in state θ', where $X^*(s_1)$ solves

$$\max_{\tilde{c}} U^2(s_1 + \tilde{c};\phi) \text{ s.t. } \tilde{c}_1 = c_1',\ \tilde{c}_2 \leq c_2',\ \text{and } \tilde{c}_2 + \tilde{c}_3 = c_2' + c_3' \tag{A-2}$$

for $\phi = \theta'$ and $\hat{X}^*(s_1)$ is as defined in the main text for $s_1 \leq s_1^*$, while for $s_1 > s_1^*$ it solves (A-2) for $\phi = \theta$. In words, X^* is the schedule that results from giving self 2 (c_2', c_3') and letting him save whatever he likes. Since we will repeatedly refer to (IC$_1$) and (IC$_2$) in the claims below, recall that

$\hat{X}^*(s_1) \equiv X(\theta, s_1; \theta')$ and $X^*(s_1) \equiv X(\theta', s_1; \theta')$.

Claim 1 $\hat{X}^*(s_1^*)$ *satisfies (SPRb) with equality, i.e.,* $U^2(c'; \theta') = \max_{s_2 \geq 0} U^2(\hat{X}^*(s_1^*) - s_2; \theta')$.

Proof of Claim 1. First suppose that $\hat{X}_2^*(s_1^*) = c_2'$; then $\hat{X}_3^*(s_1^*) = c_3'$. Since, by assumption, $U_2^2(c'; \theta') \geq U_3^2(c'; \theta')$, it follows that (SPRb) holds with equality.

Next suppose that $\hat{X}_2^*(s_1^*) > c_2'$. If, to the contrary, (SPRb) were satisfied strictly, then there would exist $\tilde{x}$ such that $\tilde{x}_2 < \hat{X}_2^*(s_1^*)$ and both (SPRa) and (SPRb) are satisfied, contradicting the definition of $\hat{X}_2^*(s_1^*)$.

Claim 2 $\hat{X}_2^*(s_1) + \hat{X}_3^*(s_1)$ *is weakly increasing in* s_1.

Proof of Claim 2. For $s_1 > s_1^*$, $\hat{X}_2^*(s_1) + \hat{X}_3^*(s_1) = c_2' + c_3'$ by definition. For $s_1 = s_1^*$, $\hat{X}_2^*(s_1) + \hat{X}_3^*(s_1) \leq c_2' + c_3'$ since by Claim 1, $\hat{X}^*(s_1^*)$ satisfies (SPRb) with equality. For $s_1 \in [0, s_1^*]$, $\hat{X}^*$ is continuous and satisfies (2); by NS, it follows that $d\hat{X}_2^*(s_1) + d\hat{X}_3^*(s_1) \geq 0$.

Claim 3 *(IC_1) is satisfied for* $\tilde{\phi}^1 = \theta'$.

Proof of Claim 3. For $\tilde{\phi}^1 = \theta'$ and $\phi = \theta$, (IC_1) reduces to $U^1(c; \theta) \geq U^1\left(s_1 + \hat{X}^*(s_1); \theta\right)$. For $s_1 > s_1^*$, recall that $\hat{X}^*(s_1)$ solves (A-2) for $\phi = \theta$; therefore it follows immediately from the definition of s_1^* that (IC_1) is satisfied.

For $s_1 = s_1^*$, recall that by the definition of $\hat{X}^*(s_1^*)$, $\hat{X}_2^*(s_1^*) \geq c_2'$ and $U^2(s_1^* + c'; \theta) = U^2\left(s_1^* + \hat{X}^*(s_1^*); \theta\right)$; therefore it follows immediately from Lemma 2 that $U^1(s_1^* + c'; \theta) \geq U^1\left(s_1^* + \hat{X}^*(s_1^*); \theta\right)$. Further, since by Lemma 5, $U_2^2(s_1^* + c'; \theta) > U_3^2(s_1^* + c'; \theta)$, it follows from the definition of s_1^* that $U^1(c; \theta) = U^1(s_1^* + c'; \theta)$ and so (IC_1) is satisfied.

For $s_1 \in [s_1^{**}, s_1^*)$, it follows from the definition of s_1^{**} that (IC_1) is satisfied.

Finally, for $s_1 \in [0, s_1^{**}]$, $U^1(s_1 + \hat{X}^*(s_1); \theta)$ is weakly increasing in s_1: either $\frac{d}{ds_1}\hat{X}_2^*(s_1) < 0$, in which case $U^1(s_1 + \hat{X}^*(s_1); \theta)$ is constant by construction, or $\frac{d}{ds_1}\hat{X}_2^*(s_1) = 0$, in which case $\frac{d}{ds_1}U^1(s_1 + \hat{X}^*(s_1); \theta) = -u_1'(c_1' - s_1; \theta) + \beta u_2'(s_1 + \hat{X}_2(s_1); \theta) \geq 0$.

For $\tilde{\phi}^1 = \theta'$ and $\phi = \theta'$, (IC_1) reduces to $U^1(c'; \theta') \geq U^1(s_1 + X^*(s_1); \theta')$. Recall that $X^*(s_1)$ solves (A-2) for $\phi = \theta'$; therefore it follows immediately from our assumption that $U^1(c(\phi); \phi) \geq \max_{s_1 \geq 0} U^1(s_1 + c(\phi) - \hat{s}_2; \phi)$, where $\hat{s}_2 = \arg\max_{s_2 \geq 0} U^2(s_1 + c(\phi) - s_2; \phi)$, that ($IC_1$) is satisfied.

Claim 4 *(IC_2) is satisfied for* $\tilde{\phi}^1 = \theta'$ *and* $\phi = \theta$.

Proof of Claim 4. Case 1. For $\tilde{\phi}^1 = \theta'$, $\phi = \theta$, and $\tilde{\phi}^2 = \theta$, (IC_2) reduces to $U^2(s_1 + \hat{X}^*(s_1); \theta) \geq U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta)$.

Step 1A. We show that for all $s_1, \tilde{s}_1 \in [0, s_1^*]$ and $s_2 \geq 0$, $U^2(s_1 + \hat{X}^*(s_1); \theta) \geq U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta)$. Consider the function $f(\tilde{s}_1) \equiv \max_{s_2 \geq 0} U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta)$ and note that by NS, $f(\tilde{s}_1 = s_1) = U^2(s_1 + \hat{X}^*(s_1); \theta)$; we will therefore show that $f(\tilde{s}_1)$ has a global maximum at $\tilde{s}_1 = s_1$. By standard envelope arguments,

$$\frac{df(\tilde{s}_1)}{d\tilde{s}_1} = U_2^2(s_1 + \hat{X}^*(\tilde{s}_1) - \hat{s}_2; \theta)\frac{d\hat{X}_2^*(\tilde{s}_1)}{d\tilde{s}_1} + U_3^2(s_1 + \hat{X}^*(\tilde{s}_1) - \hat{s}_2; \theta)\frac{d\hat{X}_3^*(\tilde{s}_1)}{d\tilde{s}_1},$$

where $\hat{s}_2 = \arg\max_{s_2 \geq 0} U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta)$. By the definition of $\hat{X}^*$,

$$\frac{df(\tilde{s}_1)}{d\tilde{s}_1} = U_3^2(s_1 + \hat{X}^*(\tilde{s}_1) - \hat{s}_2; \theta)\left(\frac{U_2^2(s_1 + \hat{X}^*(\tilde{s}_1) - \hat{s}_2; \theta)}{U_3^2(s_1 + \hat{X}^*(\tilde{s}_1) - \hat{s}_2; \theta)} - \frac{U_2^2(\tilde{s}_1 + \hat{X}^*(\tilde{s}_1); \theta)}{U_3^2(\tilde{s}_1 + \hat{X}^*(\tilde{s}_1); \theta)}\right)\frac{d\hat{X}_2^*(\tilde{s}_1)}{d\tilde{s}_1}.$$

Observe that $\frac{d}{d\tilde{s}_1} f(\tilde{s}_1)\Big|_{\tilde{s}_1 = s_1} = 0$. Given that $\frac{d}{d\tilde{s}_1}\hat{X}_2^*(\tilde{s}_1) \leq 0$, it suffices to show that

$$\frac{U_2^2(s_1 + \hat{X}^*(\tilde{s}_1) - \hat{s}_2; \theta)}{U_3^2(s_1 + \hat{X}^*(\tilde{s}_1) - \hat{s}_2; \theta)} \geq (\leq) \frac{U_2^2(\tilde{s}_1 + \hat{X}^*(\tilde{s}_1); \theta)}{U_3^2(\tilde{s}_1 + \hat{X}^*(\tilde{s}_1); \theta)} \text{ if } \tilde{s}_1 > (<) s_1,$$

since doing so establishes global concavity of f.

Consider the case of $\tilde{s}_1 > s_1$. If $U_2^2(\tilde{s}_1 + \hat{X}^*(\tilde{s}_1); \theta) = U_3^2(\tilde{s}_1 + \hat{X}^*(\tilde{s}_1); \theta)$, then the required inequality is immediate. If instead $U_2^2(\tilde{s}_1 + \hat{X}^*(\tilde{s}_1); \theta) > U_3^2(\tilde{s}_1 + \hat{X}^*(\tilde{s}_1); \theta)$, then $U_2^2(s_1 + \hat{X}^*(\tilde{s}_1); \theta) > U_3^2(s_1 + \hat{X}^*(\tilde{s}_1); \theta)$, implying $\hat{s}_2 = 0$; the required inequality follows. Finally, consider the case of $\tilde{s}_1 < s_1$. If $U_2^2(s_1 + \hat{X}^*(\tilde{s}_1) - \hat{s}_2; \theta) = U_3^2(s_1 + \hat{X}^*(\tilde{s}_1) - \hat{s}_2; \theta)$, then the required inequality is immediate. If instead $U_2^2(s_1 + \hat{X}^*(\tilde{s}_1) - \hat{s}_2; \theta) > U_3^2(s_1 + \hat{X}^*(\tilde{s}_1) - \hat{s}_2; \theta)$, then $\hat{s}_2 = 0$; the required inequality follows as before.

Step 1B. We show that for all $s_1 > s_1^*$, $\tilde{s}_1 \in [0, s_1^*]$, and $s_2 \geq 0$, $U^2(s_1 + \hat{X}^*(s_1); \theta) \geq U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta)$. By the definition of $\hat{X}^*$, $U^2(s_1 + \hat{X}^*(s_1); \theta) = \max_{s_2 \geq 0} U^2(s_1 + c' - s_2; \theta)$ for all $s_1 > s_1^*$; we therefore show that for all $s_1 > s_1^*$ and $\tilde{s}_1 \in [0, s_1^*]$, $\max_{s_2 \geq 0} U^2(s_1 + c' - s_2; \theta) \geq \max_{s_2 \geq 0} U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta)$. By NS, $U^2(s_1^* + \hat{X}^*(s_1^*); \theta) = \max_{s_2 \geq 0} U^2(s_1^* + \hat{X}^*(s_1^*) - s_2; \theta)$; by the definition of $\hat{X}^*(s_1^*)$, $U^2(s_1^* + c'; \theta) = U^2(s_1^* + \hat{X}^*(s_1^*); \theta)$; it then follows from $\hat{X}_2^*(s_1^*) \geq c_2'$ that $U^2(s_1^* + c'; \theta) = \max_{s_2 \geq 0} U^2(s_1^* + c' - s_2; \theta)$ which, together with Step A, implies that for all $\tilde{s}_1 \in [0, s_1^*]$, $\max_{s_2 \geq 0} U^2(s_1^* + c' - s_2; \theta) \geq \max_{s_2 \geq 0} U^2(s_1^* + \hat{X}^*(\tilde{s}_1) - s_2; \theta)$. Since for all $\tilde{s}_1 \in [0, s_1^*]$, $\hat{X}_2^*(\tilde{s}_1) \geq c_2'$, it then follows from Lemma A-1 that for all $s_1 > s_1^*$ and $\tilde{s}_1 \in [0, s_1^*]$, $\max_{s_2 \geq 0} U^2(s_1 + c' - s_2; \theta) \geq \max_{s_2 \geq 0} U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta)$.

Step 1C. We show that for all $s_1 \in [0, s_1^*]$, $\tilde{s}_1 > s_1^*$, and $s_2 \geq 0$, $U^2(s_1 + \hat{X}^*(s_1); \theta) \geq U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta)$. By the definition of $\hat{X}^*$, $\max_{\tilde{s}_1 > s_1^*, s_2 \geq 0} U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta) = \max_{s_2 \geq 0} U^2(s_1 + c' - s_2; \theta)$ for all $s_1 > s_1^*$; as before, $U^2(s_1 + \hat{X}^*(s_1); \theta) = \max_{s_2 \geq 0} U^2(s_1 + \hat{X}^*(s_1) - s_2; \theta)$; we will therefore show that for all $s_1 \in [0, s_1^*]$, $\max_{s_2 \geq 0} U^2(s_1 + \hat{X}^*(s_1) - s_2; \theta) \geq \max_{s_2 \geq 0} U^2(s_1 + c' - s_2; \theta)$. As before, $\max_{s_2 \geq 0} U^2(s_1^* + \hat{X}^*(s_1^*) - s_2; \theta) = \max_{s_2 \geq 0} U^2(s_1^* + c' - s_2; \theta)$; since $\hat{X}_2^*(s_1^*) \geq c_2'$ it follows from Lemma A-1 that for all $s_1 \in [0, s_1^*]$, $\max_{s_2 \geq 0} U^2(s_1 + \hat{X}^*(s_1^*) - s_2; \theta) \geq \max_{s_2 \geq 0} U^2(s_1 + c' - s_2; \theta)$ which, together with the first step of the proof establishes that for all $s_1 > s_1^*$ and $\tilde{s}_1 \in [0, s_1^*]$, $\max_{s_2 \geq 0} U^2(s_1 + c' - s_2; \theta) \geq \max_{s_2 \geq 0} U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta)$.

Step 1D. It follows directly from the definition of $\hat{X}^*$ that for all $s_1, \tilde{s}_1 > s_1^*$ and $s_2 \geq 0$, $U^2(s_1 + \hat{X}^*(s_1); \theta) \geq U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta)$.

Case 2. For $\tilde{\phi}^1 = \theta'$, $\phi = \theta$, and $\tilde{\phi}^2 = \theta'$, (IC$_2$) reduces to $U^2(s_1 + \hat{X}^*(s_1); \theta) \geq U^2(s_1 + X^*(\tilde{s}_1) - s_2; \theta)$. By the definition of X^*, Step 1C above establishes that for all $s_1 \in [0, s_1^*]$, $\tilde{s}_1 \geq 0$, and $s_2 \geq 0$, $U^2(s_1 + \hat{X}^*(s_1); \theta) \geq U^2(s_1 + X^*(\tilde{s}_1) - s_2; \theta)$ and Step 1D establishes that for all $s_1 > s_1^*$, $\tilde{s}_1 \geq 0$, and $s_2 \geq 0$, $U^2(s_1 + \hat{X}^*(s_1); \theta) \geq U^2(s_1 + X^*(\tilde{s}_1) - s_2; \theta)$.

Claim 5 *(IC$_2$) is satisfied for $\tilde{\phi}^1 = \theta'$ and $\phi = \theta'$.*

Proof of Claim 5. For $\tilde{\phi}^1 = \theta'$, $\phi = \theta'$, and $\tilde{\phi}^2 = \theta$, (IC$_2$) reduces to $U^2(s_1 + X^*(s_1); \theta') \geq U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta')$. We first show that for all $s_1 \geq 0$, $\tilde{s}_1 \in [0, s_1^*]$, and $s_2 \geq 0$, $U^2\left(s_1 + X^*(s_1); \theta'\right) \geq U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta')$. By the definition of X^*, $U^2\left(s_1 + X^*(s_1); \theta'\right) = \max_{s_2 \geq 0} U^2\left(s_1 + c' - s_2; \theta'\right)$; we will therefore show that for all $s_1 \geq 0$ and $\tilde{s}_1 \in [0, s_1^*]$, $\max_{s_2 \geq 0} U^2\left(s_1 + c' - s_2; \theta'\right) \geq \max_{s_2 \geq 0} U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta')$. Consider the case of $s_1 = 0$: since c' is self 0's desired allocation in state θ', $\max_{s_2 \geq 0} U^2\left(c' - s_2; \theta'\right) = U^2\left(c'; \theta'\right)$; by Claim 1, $U^2\left(c'; \theta'\right) = \max_{s_2 \geq 0} U^2(\hat{X}^*(s_1^*) - s_2; \theta')$. By Claim 4, $U^2(s_1^* + \hat{X}^*(s_1^*); \theta) \geq U^2(s_1^* + \hat{X}^*(\tilde{s}_1) - s_2; \theta)$ for all $\tilde{s}_1 \in [0, s_1^*]$ and $s_2 \geq 0$. If $\hat{X}_2^*(\tilde{s}_1) - s_2 > \hat{X}_2^*(s_1^*)$, then it follows from the definition of $\hat{X}^*(s_1^*)$ and SCB that $\max_{s_2 \geq 0} U^2(\hat{X}^*(s_1^*) - s_2; \theta') \geq U^2(\hat{X}^*(\tilde{s}_1) - s_2; \theta)$. If $\hat{X}_2^*(\tilde{s}_1) - s_2 \leq \hat{X}_2^*(s_1^*)$ then, by Claim 2, $\hat{X}_2^*(\tilde{s}_1) + \hat{X}_3^*(\tilde{s}_1) \leq \hat{X}_2^*(s_1^*) + \hat{X}_3^*(s_1^*)$, and so $\max_{s_2 \geq 0} U^2(\hat{X}^*(s_1^*) - s_2; \theta') \geq U^2(\hat{X}^*(\tilde{s}_1) - s_2; \theta)$. As a result, $\max_{s_2 \geq 0} U^2\left(s_1 + c' - s_2; \theta'\right) \geq \max_{s_2 \geq 0} U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta')$ for $s_1 = 0$ and all $\tilde{s}_1 \in [0, s_1^*]$; it then follows from Lemma A-1 that the same is true for all $s_1 \geq 0$ and $\tilde{s}_1 \in [0, s_1^*]$.

Finally, it follows directly from the definitions of $\hat{X}^*$ and X^* that for all $s_1 \geq 0$ and $\tilde{s}_1 > s_1^*$, $U^2\left(s_1 + X^*(s_1); \theta'\right) \geq U^2(s_1 + \hat{X}^*(\tilde{s}_1) - s_2; \theta')$.

For $\tilde{\phi}^1 = \theta'$, $\phi = \theta'$, and $\tilde{\phi}^2 = \theta'$, (IC$_2$) reduces to $U^2(s_1 + X^*(s_1); \theta') \geq U^2(s_1 + X^*(\tilde{s}_1) - s_2; \theta')$; it follows directly from the definition of X^* that it is satisfied.

Proof of Theorem 1 (necessity) Suppose that SPR is satisfied and $\hat{X}^*$ violates NS but, to the contrary, commitment is possible. Let $\hat{X}$ be a commitment contract.

Claim 1 *$\hat{X}^*$ satisfies NS at s_1^*.*

Proof of Claim 1. If $\hat{X}^*(s_1^*) = c'$, this is immediate from Lemma A-4. If instead $\hat{X}^*(s_1^*) \neq c'$, suppose to the contrary that $U_2^2\left(s_1^* + \hat{X}^*(s_1^*);\theta\right) < U_3^2\left(s_1^* + \hat{X}^*(s_1^*);\theta\right)$. But then for $\varepsilon > 0$ sufficiently small, $\hat{X}^*(s_1^*)+(0,-\varepsilon,\varepsilon)$ satisfies both (SPRa) and (SPRb), contradicting the definition of $\hat{X}^*(s_1^*)$.

Claim 2 *There exists $s_1 < s_1^*$ at which $\hat{X}^*$ violates NS and satisfies (IC_1) with equality for $\left(\phi, \tilde{\phi}^1\right) = (\theta, \theta')$.*

Proof of Claim 2. If (IC_1) holds with equality at s_1, the proof is complete. Otherwise, $\hat{X}^*$ must be locally constant at s_1. Suppose for now that the set $\left\{\tilde{s}_1 \in (s_1, s_1^*] : \hat{X}^* \text{ is not locally constant at } \tilde{s}_1\right\}$ is nonempty and let $\underline{s}_1$ be its infimum. Since $\hat{X}^*$ is constant on $[s_1, \underline{s}_1]$ and NS is violated at $s_1 < \underline{s}_1$, NS must also be violated at $\underline{s}_1$. Moreover, since $\hat{X}^*$ is not locally constant at $\underline{s}_1$, (IC_1) holds with equality at $\underline{s}_1$. Finally, note that the above set is indeed nonempty: if it were empty, then the argument just made would imply that NS is violated at s_1^*, contradicting Claim 1.

Claim 3 *There exists $s_1 < s_1^*$ such that $U^2(s_1 + \hat{X}(s_1);\theta) < U^2(s_1 + \hat{X}^*(s_1);\theta)$.*

Proof of Claim 3. Following Claim 2, let $s_1 < s_1^*$ be such that, at s_1, $\hat{X}^*$ violates NS and satisfies (IC_1) with equality for $\left(\phi, \tilde{\phi}^1\right) = (\theta, \theta')$. Since by supposition $\hat{X}$ is a commitment contract, $U^1(s_1 + \hat{X}(s_1);\theta) \leq U^1(s_1 + \hat{X}^*(s_1);\theta)$. We show that $U^2(s_1 + \hat{X}(s_1);\theta) < U^2(s_1 + \hat{X}^*(s_1);\theta)$. Suppose to the contrary that $U^2(s_1 + \hat{X}(s_1);\theta) \geq U^2(s_1 + \hat{X}^*(s_1);\theta)$. Together with Lemma 2, these inequalities imply that $\hat{X}_2(s_1) \geq \hat{X}_2^*(s_1)$ and $\hat{X}_3(s_1) \leq \hat{X}_3^*(s_1)$. Since $\hat{X}^*$ violates NS at s_1, this implies that $\hat{X}$ also violates NS at s_1, contradicting our supposition that $\hat{X}$ is a commitment contract.

Final step of proof. We start by defining $\tilde{X}^*$—a perturbation of $\hat{X}^*$—as follows. Define its boundary $\tilde{X}^*(s_1^*)$ such that $\tilde{X}_2^*(s_1^*) < \hat{X}_2^*(s_1^*)$ and $U^1\left(s_1^* + \tilde{X}^*(s_1^*);\theta\right) = U^1\left(s_1^* + \hat{X}^*(s_1^*);\theta\right)$; it follows from Lemma 2 that $U^2\left(s_1^* + \tilde{X}^*(s_1^*);\theta\right) < U^2\left(s_1^* + \hat{X}^*(s_1^*);\theta\right)$. Given the boundary, define $\tilde{X}^*$ in a parallel way to $\hat{X}^*$. That is, let $\tilde{s}_1^{**} \equiv \sup\left\{s_1 \in [0, s_1^*] : U^1\left(s_1 + \tilde{X}^*(s_1^*);\theta\right) > U^1(c;\theta)\right\}$, with $\tilde{s}_1^{**} = s_1^*$ if the set is empty. Then, for every $s_1 \in [\tilde{s}_1^{**}, s_1^*]$, let $\tilde{X}^*(s_1) = \tilde{X}^*(s_1^*)$, and for every $s_1 \in [0, \tilde{s}_1^{**}]$, define $\tilde{X}^*(s_1)$ by the differential equations (2) and (3).

Following Claim 3, fix $s_1^+ < s_1^*$ such that $U^2(s_1^+ + \hat{X}(s_1^+);\theta) < U^2(s_1^+ + \hat{X}^*(s_1^+);\theta)$. As long as the boundary $\tilde{X}^*(s_1^*)$ is chosen to be close enough to $\hat{X}^*(s_1^*)$, $U^2(s_1^+ + \hat{X}(s_1^+);\theta) <$

$U^2\left(s_1^+ + \tilde{X}^*\left(s_1^+\right);\theta\right)$. Moreover, $U^2(s_1^* + \tilde{X}^*\left(s_1^*\right);\theta) < U^2(s_1^* + \hat{X}\left(s_1^*\right);\theta)$: as noted above, $U^2\left(s_1^* + \tilde{X}^*\left(s_1^*\right);\theta\right) < U^2\left(s_1^* + \hat{X}^*\left(s_1^*\right);\theta\right)$ and, from the definition of $\hat{X}^*\left(s_1^*\right)$ and the supposition that $\hat{X}$ is a commitment contract, $U^2(s_1^* + \hat{X}^*\left(s_1^*\right);\theta) = U^2\left(s_1^* + c';\theta\right) \leq U^2(s_1^* + \hat{X}\left(s_1^*\right);\theta)$.

From the proof of Proposition 2, Claim 4, it is straightforward to see that $U^2(s_1 + \hat{X}\left(s_1\right);\theta)$ is continuous in s_1 and differentiable at any s_1 at which $\hat{X}$ is continuous. Since $\hat{X}$ is continuous at all but at most a finite number of points, it then follows that there exists $s_1^{++} \in \left(s_1^+, s_1^*\right)$ such that $U^2(s_1^{++} + \hat{X}\left(s_1^{++}\right);\theta) > U^2\left(s_1^{++} + \tilde{X}^*\left(s_1^{++}\right);\theta\right)$ and $\frac{d}{ds_1}\, U^2(s_1 + \hat{X}\left(s_1\right);\theta)\Big|_{s_1=s_1^{++}} \geq \frac{d}{ds_1}\, U^2\left(s_1 + \tilde{X}^*\left(s_1\right);\theta\right)\Big|_{s_1=s_1^{++}}$. Equation (2) implies $\frac{d}{ds_1}\, U^2(s_1 + \hat{X}\left(s_1\right);\theta)\Big|_{s_1=s_1^{++}} = U_2^2(s_1^{++} + \hat{X}\left(s_1^{++}\right);\theta)$ and $\frac{d}{ds_1}\, U^2\left(s_1 + \tilde{X}^*\left(s_1\right);\theta\right)\Big|_{s_1=s_1^{++}} = U_2^2\left(s_1^{++} + \tilde{X}^*\left(s_1^{++}\right);\theta\right)$; therefore $\hat{X}_2\left(s_1^{++}\right) \leq \tilde{X}_2^*\left(s_1^{++}\right)$ and $\hat{X}_3\left(s_1^{++}\right) > \tilde{X}_3^*\left(s_1^{++}\right)$. Lemma 2 then implies that $U^1(s_1^{++} + \hat{X}\left(s_1^{++}\right);\theta) > U^1\left(s_1^{++} + \tilde{X}^*\left(s_1^{++}\right);\theta\right)$.

If $U^1\left(s_1^{++} + \tilde{X}^*\left(s_1^{++}\right);\theta\right) = U^1\left(c;\theta\right)$, then $U^1(s_1^{++} + \hat{X}\left(s_1^{++}\right);\theta) > U^1\left(c;\theta\right)$, contradicting our supposition that $\hat{X}$ is a commitment contract. If instead $U^1\left(s_1^{++} + \tilde{X}^*\left(s_1^{++}\right);\theta\right) < U^1\left(c;\theta\right)$ (the opposite inequality is ruled out by the definition of $\tilde{X}^*$), then, by the definition of $\tilde{X}^*$, $\tilde{X}^*\left(s_1^{++}\right) = \tilde{X}^*\left(s_1^*\right)$ and so $\hat{X}_2\left(s_1^{++}\right) \leq \tilde{X}_2^*\left(s_1^{++}\right) = \tilde{X}_2^*\left(s_1^*\right) < \hat{X}_2^*\left(s_1^*\right)$, contradicting Proposition 4 and thus also our supposition that $\hat{X}$ is a commitment contract.

Proof of Proposition 5. We show that $c_3'^* \geq c_3^*$ is equivalent to $U_2^2\left(s_1^* + c'^*;\theta\right)/U_3^2\left(s_1^* + c'^*;\theta\right) \geq U_2^2\left(c'^*;\theta'\right)/U_3^2\left(c'^*;\theta'\right)$, the simple sufficient condition for SPR introduced in Lemma 4. Note that this condition is itself equivalent to $u_2'\left(s_1^* + c_2'^*;\theta\right) \geq u_3'\left(c_3'^*;\theta\right)$ since, by the definitions of c^* and c'^*, $u_2'\left(c_2^*;\theta\right)/u_3'\left(c_3^*;\theta\right) = u_2'\left(c_2'^*;\theta'\right)/u_3'\left(c_3'^*;\theta'\right) = 1$. Hence we show that $c_3'^* \geq c_3^*$ if and only if $u_2'\left(s_1^* + c_2'^*;\theta\right) \geq u_3'\left(c_3'^*;\theta\right)$.

Suppose that $c_3'^* \geq c_3^*$; then $c_1'^* - c_1^* \leq c_2^* - c_2'^*$. From Lemma A-6 we know that $s_1^* \leq c_1'^* - c_1^*$; therefore $s_1^* \leq c_2^* - c_2'^*$. As a result, $u_2'\left(s_1^* + c_2'^*;\theta\right) \geq u_2'\left(c_2^*;\theta\right) = u_3'\left(c_3^*;\theta\right) \geq u_3'\left(c_3'^*;\theta\right)$.

Note that (since $\beta \leq 1$) if $u_2'\left(s_1^* + c_2'^*;\theta\right) \geq u_3'\left(c_3'^*;\theta\right)$, then $s_2 = 0$ solves $\max_{s_2 \geq 0} U^2\left(s_1 + c'^* - s_2;\theta\right)$ for $s_1 \leq s_1^*$ and so s_1^* is simply the largest solution to $U^1\left(s_1 + c'^*;\theta\right) = U^1\left(c^*;\theta\right)$; moreover, because there is a commitment problem, $u_1'\left(c_1'^* - s_1^*;\theta\right) > \beta u_2'\left(s_1^* + c_2'^*;\theta\right)$.

It follows from the above two steps that if $c_3'^* \geq c_3^*$, then $u_2'\left(s_1^* + c_2'^*;\theta\right) \geq u_3'\left(c_3'^*;\theta\right)$, which in turn implies $U^1\left(s_1^* + c'^*;\theta\right) = U^1\left(c^*;\theta\right)$. It follows from concavity of preferences and the definition of c^* that if $c_3'^* = c_3^*$, then $s_1^* = c_1'^* - c_1^* = c_2^* - c_2'^*$ (so that $s_1^* + c'^* = c^*$) and so $u_2'\left(s_1^* + c_2'^*;\theta\right) = u_2'\left(c_2^*;\theta\right) = u_3'\left(c_3^*;\theta\right) = u_3'\left(c_3'^*;\theta\right)$.

To complete the proof, suppose that $c_3'^* < c_3^*$ but, contrary to the claimed result, $u_2'(s_1^* + c_2'^*; \theta) \geq u_3'(c_3'^*; \theta)$. Just as above, $U^1(s_1^* + c'^*; \theta) = U^1(c^*; \theta)$ and $u_1'(c_1'^* - s_1^*; \theta) > \beta u_2'(s_1^* + c_2'^*; \theta)$. To see how $s_1^* + c_2'^*$ evolves as $c_3'^*$ increases, differentiate $U^1(s_1^* + c'^*; \theta) = U^1(c^*; \theta)$ to obtain

$$\frac{d(s_1^* + c_2'^*)}{dc_3'^*} = -\frac{u_1'(c_1'^* - s_1^*; \theta) - \beta u_3'(c_3'^*; \theta)}{u_1'(c_1'^* - s_1^*; \theta) - \beta u_2'(s_1^* + c_2'^*; \theta)}.$$

The denominator is strictly positive and by our supposition that $u_2'(s_1^* + c_2'^*; \theta) \geq u_3'(c_3'^*; \theta)$, the numerator is weakly greater than the denominator; it follows that the derivative is strictly negative. As a result, $u_2'(s_1^* + c_2'^*; \theta)$ strictly increases as $c_3'^*$ increases, leading to a contradiction as $c_3'^*$ approaches c_3^*.

Proof of Proposition 6. By Proposition 5, SPR is satisfied, so by Theorem 1, it suffices to show that $\hat{X}^*$ satisfies NS for all β sufficiently close to β^*. Throughout the proof we regularly add a β argument to s_1^*, $\hat{X}^*$, and U^t to aid exposition.

As a preliminary, note that $\beta^* < 1$, as follows. By assumption, $s_2 + c_3'^* > c_3^*$ for all $s_2 \geq 0$. Therefore, by the definition of c^*, $U^1(s_1 + c'^* - s_2; \theta, \beta = 1) < U^1(c^*; \theta, \beta = 1)$ for all $s_1, s_2 \geq 0$.

For $\beta \leq \beta^*$ such that there is a commitment problem, the proof of Proposition 5 implies (since $c_3'^* > c_3^*$)

$$\frac{u_2'(s_1^*(\beta) + c_2'^*; \theta)}{\beta u_3'(c_3'^*; \theta)} \geq \frac{u_2'\left(c_2'^*; \theta'\right)}{\beta u_3'\left(c_3'^*; \theta'\right)} = \frac{1}{\beta} \geq \frac{1}{\beta^*} > 1. \tag{A-3}$$

The result then follows if we can show that $\max_{s_1 \in [0, s_1^*(\beta)]} \left|\hat{X}_2^*(s_1; \beta) - c_2'^*\right| \to 0$ as $\beta \to \beta^*$. For use below, note that the first inequality in (A-3) and Lemma 6 imply $\hat{X}^*(s_1^*(\beta); \beta) = c'^*$.

Using the simplifications noted for the $c_3'^* \geq c_3^*$ case in the proof of Proposition 5, $\max_{s_1 \geq 0} U^1(s_1 + c'^*; \theta, \beta^*) = U^1(c^*; \theta)$. Define $s_1^*(\beta^*) = \lim_{\beta \to \beta^*} s_1^*(\beta)$, and note that $s_1^*(\beta^*) = \arg\max_{s_1 \geq 0} U^1(s_1 + c'^*; \theta, \beta^*)$. The important implication of these observations is that, if $s_1^*(\beta^*) > 0$,

$$u_1'\left(c_1'^* - s_1^*(\beta); \theta\right) - \beta u_2'\left(s_1^*(\beta) + c_2'^*; \theta\right) \to 0 \text{ as } \beta \to \beta^*. \tag{A-4}$$

Recall that $\hat{X}^*$ satisfies (4) (see page 21). We claim, and show below, that $\hat{X}_2^*(\cdot; \beta)$ is concave for β sufficiently close to β^*. Given this, we know that for any $s_1 \in [0, s_1^*(\beta)]$, $\hat{X}_2^*(s_1; \beta)$ lies between $c_2'^*$ and $c_2'^* - (s_1^*(\beta) - s_1) \frac{d}{ds_1} \hat{X}_2^*(s_1; \beta)\Big|_{s_1 = s_1^*(\beta)}$. If $s_1^*(\beta^*) > 0$, the result then follows since (by $\beta^* < 1$

and (A-4))

$$\left.\frac{d\hat{X}_2^*(s_1;\beta)}{ds_1}\right|_{s_1=s_1^*(\beta)} \to 0 \text{ as } \beta \to \beta^*.$$

If instead $s_1^*(\beta^*)=0$, the result follows simply from the fact that

$$\left.\frac{d\hat{X}_2^*(s_1;\beta)}{ds_1}\right|_{s_1=s_1^*(\beta)} \to -\frac{u_1'(c_1'^*;\theta)-\beta^* u_2'(c_2'^*;\theta)}{(1-\beta^*)\,u_2'(c_2'^*;\theta)} \text{ as } \beta \to \beta^*, \tag{A-5}$$

i.e., the derivative does not explode as β approaches β^*. (Note that the limit is negative, since in this case $s_1=0$ maximizes $U^1(s_1+c'^*;\theta,\beta^*)$.)

It remains only to establish that $\hat{X}_2^*(\cdot;\beta)$ is concave. For this, it is sufficient to establish that $s_1+\hat{X}_2^*(s_1;\beta)$ is weakly increasing in s_1 on $[0,s_1^*(\beta)]$. From (A-4), if $s_1^*(\beta^*)>0$, then for all β sufficiently close to β^*,

$$\frac{d\left(s_1+\hat{X}_2^*(s_1;\beta)\right)}{ds_1} \geq 0 \tag{A-6}$$

is satisfied strictly at $s_1=s_1^*(\beta)$. If instead $s_1^*(\beta^*)=0$, then note that $c_1'^*>c_1^*$ and $c_2'^*<c_2^*$ (since $c_3'^*>c_3^*$) imply $u_2'(c_2'^*;\theta)-u_1'(c_1'^*;\theta)>u_2'(c_2^*;\theta)-u_1'(c_1^*;\theta)=0$; substituting in (A-5) then implies that (A-6) is again satisfied strictly for β sufficiently close to β^*.

Fix any such β close enough to β^* such that (A-6) indeed holds strictly at $s_1=s_1^*(\beta)$. Suppose there exists $s_1\in[0,s_1^*(\beta)]$ such that (A-6) holds at equality. At any such s_1,

$$\frac{d^2\left(s_1+\hat{X}_2^*(s_1;\beta)\right)}{ds_1^2} = \frac{1}{1-\beta}\frac{u_1''(c_1'^*-s_1;\theta)}{u_2'\left(s_1+\hat{X}_2^*(s_1;\beta);\theta\right)} < 0.$$

Consequently, if any such s_1 exists, it follows from continuity that $\frac{d}{ds_1}\left(s_1+\hat{X}_2^*(s_1;\beta)\right)\leq 0$ for all higher values of s_1, contradicting the fact that (A-6) holds strictly at $s_1^*(\beta)$. It follows that (A-6) holds strictly for all $s_1\in[0,s_1^*(\beta)]$, completing the proof.

Proof of Proposition 7. The main step is to show that self 2 will not use the costly flexibility in state θ after self 1 chooses c_1^*, i.e., that for all s_1,

$$\max_{s_2\geq 0} U^2(s_1+c^*-s_2;\theta) \geq \max_{\tilde{s}_1,s_2\geq 0} U^2\left(s_1+c^*+\hat{X}^*(\tilde{s}_1)-c'^*-s_2;\theta\right). \tag{A-7}$$

Note that even if self 2 chooses the costly flexibility in state θ' after self 1 chooses c_1^*, self 1 would still not deviate and choose c_1^* in state θ'. The proof parallels that of Lemma A-6.

To establish (A-7), suppose to the contrary that there exist s_1, $\tilde{s}_1$, $s_2 \geq 0$ such that

$$U^2 \left(s_1 + c^* + \hat{X}^* (\tilde{s}_1) - c'^* - s_2; \theta\right) > \max_{\tilde{s}_2 \geq 0} U^2 (s_1 + c^* - \tilde{s}_2; \theta). \tag{A-8}$$

Note first that since $\hat{X}_2^* (\tilde{s}_1) + \hat{X}_3^* (\tilde{s}_1) \leq c_2'^* + c_3'^*$, it must be that $\hat{X}_2^* (\tilde{s}_1) - c_2'^* - s_2 > 0$. In turn, this implies $c_3'^* - \hat{X}_3^* (\tilde{s}_1) - s_2 > 0$. Moreover, inequality (A-8) implies that $U^2 \left(s_1 + c^* + \hat{X}^* (\tilde{s}_1) - c'^* - s_2; \theta\right) > U^2 (s_1 + c^*; \theta)$ or, equivalently,

$$\begin{aligned} & u_2 \left(s_1 + c_2^* + \left(\hat{X}_2^* (\tilde{s}_1) - c_2'^* - s_2\right); \theta\right) - u_2 (s_1 + c_2^*; \theta) \\ > \;& \beta u_3 (c_3^*; \theta) - \beta u_3 \left(c_3^* - \left(c_3'^* - \hat{X}_3^* (\tilde{s}_1) - s_2\right); \theta\right). \end{aligned} \tag{A-9}$$

Since $c_3'^* \geq c_3^*$ is equivalent to $c_1'^* - c_1^* \leq c_2^* - c_2'^*$ and, by Lemma A-6, $s_1^* \leq c_1'^* - c_1^*$, we know that $s_1^* + c_2'^* \leq c_2^*$. As a result, (A-9) and concavity of preferences imply

$$u_2 \left(s_1^* + c_2'^* + \left(\hat{X}_2^* (\tilde{s}_1) - c_2'^* - s_2\right); \theta\right) - u_2 (s_1^* + c_2'^*; \theta) > \beta u_3 (c_3'^*; \theta) - \beta u_3 \left(c_3'^* - \left(c_3'^* - \hat{X}_3^* (\tilde{s}_1) - s_2\right); \theta\right)$$

or, equivalently, $U^2 \left(s_1^* + \hat{X}^* (\tilde{s}_1) - s_2; \theta\right) > U^2 (s_1^* + c'^*; \theta)$. But since $U^2 (s_1^* + c'^*; \theta) = U^2 \left(s_1^* + \hat{X}^* (s_1^*); \theta\right)$ by construction, this last inequality violates (IC$_2$) for $\left(\phi, \tilde{\phi}^1, \tilde{\phi}^2\right) = \left(\theta, \theta', \theta\right)$, contradicting Theorem 1 and completing the proof.