Hyperbolic Discounting and Consumption

by

David Isaac Laibson

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1994

Author .....
Department of Economics
May 12, 1994

Certified by .....
Olivier Jean Blanchard
Professor of Economics
Thesis Supervisor

Certified by .....
Roland Benabou
Associate Professor of Economics
Thesis Supervisor

Accepted by .....
Olivier Jean Blanchard
Chairman, Departmental Committee on Graduate Students

Hyperbolic Discounting and Consumption

by

David Isaac Laibson

Submitted to the Department of Economics
on May 12, 1994, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Chapter 1: Research on animal and human behavior has led psychologists to conclude that discount functions are approximately hyperbolic (Ainslie, 1992). This thesis characterizes the savings/consumption behavior of sophisticated economic decision-makers with hyperbolic discount functions. Such preferences are characterized by "dynamic inconsistency"; the preferences imply a conflict between the optimal contingent plan from today's perspective and the optimal decision from tomorrow's perspective. To model intertemporal choice when such conflicts arise, I assume that individuals engage in an intertemporal game with themselves (Strotz (1956), Phelps (1968), Peleg and Yaari (1973), Goldman (1980)). Specifically, I reinterpret the infinite-horizon intergenerational consumption game proposed by Phelps and Pollak (1968) as an intra-personal consumption game. I show that hyperbolic discounting generates a coordination problem which leads to the existence of multiple intra-personal equilibria, some of which are Pareto-rankable. (In this intra-personal game two equilibria are Pareto-rankable if all temporal selves of the individual are better off under one of the two equilibria.) I characterize the equilibrium set, and calibrate the model. I interpret this model as a framework for understanding the psychological concept of self-control.

Chapter 2: I analyze the decisions of a consumer with a hyperbolic discount function who has access to a stylized precommitment technology: a partially illiquid financial instrument. The financial instrument is illiquid in the sense that the sale of this asset has to be initiated one period before the actual proceeds are received. The availability of this imperfect precommitment technology has important consequences, because the non-exponential discount structure implies that the decision maker has dynamically inconsistent preferences, and hence has an incentive to try to constrain her own future choices. This framework can be used to resolve many of the empirical puzzles in the consumption/savings literature. First, the model explains why consumption tracks income, while simultaneously explaining how impatient consumers manage to avoid excessive dissaving. Second, the model explains why consumers have a different

propensity to consume out of wealth than they do out of labor income. Third, the model presents a new challenge to Ricardian equivalence. Finally, the model provides two potential explanations for the low level of U.S. savings during the 1980's.

Chapter 3: I analyze the problem of an agent with dynamically inconsistent preferences (hyperbolic discount functions) who has access to a "binding automaton:" a machine which enables the agent to perfectly commit herself to contingent rules linking observable states to observable actions. I assume that effort is not observable, generating an intra-personal principal-agent problem. In equilibrium, the agent exhibits a high marginal propensity to consume (MPC) out of effort-related income (*e.g.* labor income) and a relatively low MPC out of income which is independent of effort (*e.g.* capital gains). I interpret this as a partial explanation of mental accounts (Thaler, 1990) and of self-reward/self-punishment.

Thesis Supervisor: Olivier Jean Blanchard
Title: Professor of Economics

Thesis Supervisor: Roland Benabou
Title: Associate Professor of Economics

Acknowledgments

I have received wonderful support and guidance all along the way. Olivier Blanchard always asked the right questions, and shaped my thinking at almost all key junctures. He knew when to be tough and when to encourage me to explore a poorly structured but promising idea. His insights and encouragement played a central role in the development of the essays which follow. Roland Benabou, Matthew Rabin, and Julio Rotemberg also made numerous critical contributions to my intellectual growth. Each of them constantly pushed me to think more carefully and more clearly. I always felt that I understood things more deeply after speaking to them. I also received invaluable guidance from Ricardo Caballero, Drew Fudenberg, Drazen Prelec, and Jean Tirole. In particular, Drazen introduced me to the field of psychology and economics.

I am also grateful to my friends. They taught me a lot of economics. And they taught me how to relax over a cup of coffee. They kept life fun, and tried (sometimes in vain) to prevent me from losing perspective. I am particularly indebted to Mariann Donovan, Fausto Panunzi, Tom Skinner, Jeromin Zettelmeyer, and my fourth-year officemates: Peter Klibanoff, Owen Lamont, and Fiona Scott-Morton.

Financially, I have been supported by the National Science Foundation and the Alfred P. Sloan Foundation.

Lastly, and undoubtedly most importantly, my family has given me strength, comfort, and inspiration at every stage of my education. In so many ways, my parents and sisters coauthored this thesis with me.

Dedication

To my parents and sisters, lifelong coauthors.

Contents

0.1 Introduction

Research on animal and human behavior has led psychologists to conclude that discount functions are approximately hyperbolic: rewards τ periods in the future are discounted $\frac{1}{\kappa_1+\kappa_2\cdot\tau}$ where the κ's are constants (see Ainslie (1992)).[1] Hyperbolic discount functions imply a monotonically falling discount rate. This discount structure sets up a conflict between today's preferences and the preferences which will be held in the future. For example, from today's perspective, the discount rate between two far off periods, t and $t+1$, is a long-term low discount rate. However, from the time t perspective, the discount rate between t and $t+1$ is a short-term high discount rate. This type of preference "change" is reflected in many common experiences. For example, today I may desire to quit smoking next year, but when next year actually roles around my taste at that time will be to postpone any sacrifices another year.

Hyperbolic discount functions generate a preference structure which is a special case of the general class of "dynamically inconsistent" preferences: *i.e.*, preferences which imply a conflict between the optimal contingent plan from today's perspective and the optimal decision from tomorrow's perspective. Robert Strotz (1956) was the first economist to study dynamically inconsistent preferences. Pollak (1968), Peleg and Yaari (1973), and Goldman (1980) have extended Strotz's work, arguing that when preferences are dynamically inconsistent, dynamic decisions should be modelled as an *intra*-personal game among different temporal selves (*i.e.*, "me today" is modelled as a different player from "me tomorrow").

Despite the availability of this analytic framework, and the substantial body of evidence supporting hyperbolic discounting, few economists have studied the implications of hyperbolic discount functions. Phelps and Pollak (1968) analyze an intergenerational game in which each generation has a discount function which I will argue below is approximately hyperbolic. However, Phelps and Pollak restrict their attention to linear Markov-equilibria, undermining the generality of their analysis. Akerlof (1991) analyzes the behavior of decision-makers who place a special premium on ef-

[1]The citations in this introduction are referenced in the bibliography at the end of chapter one.

fort made in the current period. Such a premium can be interpreted as a reflection of hyperbolic discounting although this interpretation is not made by Akerlof. Akerlof's analysis is inconsistent with the intra-personal game approach as his decision-makers act myopically; they fail to foresee the preference "changes" described above. Finally, Loewenstein and Prelec (1992) also analyze the choices of myopic decision-makers with hyperbolic discount functions.

To date economists have not characterized the unrestricted behavior of sophisticated/rational economic decision-makers with hyperbolic discount functions. However, in pathbreaking work, a pschyiatrist, George Ainslie (1992), has discussed the kinds of qualitative behavior that sophisticated agents with hyperbolic discount functions might exhibit. The primary goal of this thesis is to formalize and extend Ainslie's psychological analysis, by explicitly modelling individual decision-making as an intra-personal game.

The chapters which follow share a common assumption: consumers have hyperbolic discount functions, and consumers understand the intra-personal conflicts that they face. The chapters are differentiated by the assumptions that I make regarding the external constraints in each consumer's environment. In chapter one, a consumer faces an infinite horizon "eat the pie"problem. In this chapter the consumer is assumed to be unable to use commitment to predetermine future choices. In chapter two, a consumer faces a finite horizon consumption smoothing problem in which the consumer can use an illiquid asset to achieve partial commitment. In chapter three, a consumer faces a three-period consumption smoothing problem with an effort decision. Here I assume that the consumer has an extremely sophisticated precommitment mechanism that can tie future observable actions to observable states.

Chapter 1

Self-Control and Saving

1.1 Introduction

This chapter analyzes the "eating a pie" consumption problem when the consumer's discount function is hyperbolic and the consumer can not commit future actions. I adopt a stylized discount function which can capture the qualitative properties of the hyperbolic discount structure, and which nests the standard exponential discount structure as a special case. I analyze an intra-personal delay-of-gratification game which is associated with these preferences. I look for subgame-perfect equilibria of this game and find a multiplicity which is disconcerting because it arises *even* in the special case of exponential discounting. To address this problem, I propose a menu of three refinements, *each* of which eliminates the multiplicity when discounting is exponential. For the exponential case, moreover, each of these refinements admits a single equilibrium which is the standard "Ramsey" outcome.

For the case of hyperbolic discounting, only two of the original three refinements still imply uniqueness. A common equilibrium survives both of these refinements, but this equilibrium is Pareto-dominated by a continuum of other subgame perfect equilibria. (In this intra-personal game two equilibria are pareto-rankable if all temporal selves of the individual are better off under one of the two equilibria.) This suggests that there is no focal equilibrium in the non-exponential discount case. So while there are reasonable ways to rule out multiplicity in the exponential case, such arguments

do not carry over to the case in which discount rates are non-exponential.

This conclusion has interesting consequences for the theory of dynamic choice in general and for consumption theory in particular. In the dynamic choice category there are three conclusions which should be highlighted. First, this approach provides a way of explaining why people with identical preferences might exhibit dramatically different behavior in the same environment. Second, this approach can explain why some people have self-acknowledged "bad habits" which they can't break, while other people with ostensibly identical preferences muster the internal self-discpline/self-control to avoid these "bad habits." The bad habit is a perfect equilibrium which is Pareto-dominated by another perfect equilibrium. Third, the model explains why individual behavior is often perceived to be self-diagnostic of future behavior (*i.e.*, the model explains why individuals often reason, "if I show self-control today, I'll be more likely to show self-control tomorrow").

Finally the model is of direct interest to macroeconomists, since the intertemporal choice problem that I study is a savings/consumption decision. The model explains how an economy's savings rate can be indeterminate in equilibrium. A single calibration of the model produces many potential equilibrium paths. For any reasonable parameter vector the associated range of equilibrium savings rates is quite large, with substantially more dispersion than that exhibited in cross-country savings studies. Hence, the multiple equilibria of this model can explain heterogeneous international savings rates. This suggests that the heterogeneity implied by the model is substantial when normalized by cross-country savings variability.

The chapter is divided into six sections, including this introduction. Section two lays out the formal model, describing the specific intra-personal game that I consider. Section three characterizes the equilibrium set for the case of exponential discounting, and proposes three refinements. Section four applies the refinements of section three to the case of non-exponential discounting. Section five characterizes the equilibrium set when the discount structure is non-exponential. Section six interprets the results and section seven concludes.

1.2 An intra-personal game.

Following Strotz (1955), Pollak (1968), Peleg and Yaari (1972), and Goldman (1980), I model an individual as a composite of autonomous temporal selves. These selves are indexed by their respective periods of control, $(t = 0, 1, 2, \ldots)$, over a consumption decision. During its period of control, self t observes all past consumption levels $(c_0, c_1, c_2, \ldots, c_{t-1})$, and the current wealth level A_t. Self t then chooses a consumption level for period t, which satisfies the restriction,

$$0 \leq c_t \leq A_t. \tag{1.1}$$

Self $t+1$ then "inherits" an asset stock equal to,

$$A_{t+1} = R \cdot (A_t - c_t) \tag{1.2}$$

where R is the constant gross return. The game continues, with self $t+1$ in control. Finally, note that in this game the precommitment solution discussed by Strotz (1956) is implicitly ruled out; *i.e.*, the current self can not make consumption decisions for future selves.[1]

Now it only remains to specify the payoffs of the "players" of this game. Player t receives payoff $U_t(c_0, c_1, \ldots, c_t, \ldots)$ where U_t is a map into the real line or the extended real line $(\overline{\Re} \equiv \Re \cup \{-\infty, \infty\})$.[2] I restrict U_t to focus consideration on a special case which is both economically interesting and analytically tractable.

[1] From the perspective of the time zero self, precommitment would be optimal. However, in the real world precommitment is often not possible, since for most forms of economic activity there do not exist institutional mechanisms for precommitment. Three mutually compatible stories may explain this empirical observation. First, such contracts would be difficult to enforce without creating a costly monitoring structure. Second, in a world of uncertainty, such contracts would be difficult to write down. Specifying all of the potential contingencies is impossible. Finally, precommitment is ethically ambiguous, and, at least in the US, some precommitment contracts are not legally binding. Should a 25 year old self be able to make commitments which the 50 year old self is compelled to follow? Which self has the right to make temporally global decisions. The ethical and legal dimensions of this last question are discussed in Schelling (1983).

[2] It may help to interpret self t's payoff as the expectation at time t of this function.

In particular, I assume

$$U_t(c_0, c_1, \ldots, c_t, \ldots) = u(c_t) + \beta \sum_{i=1}^{\infty} \delta^i u(c_{t+i}) \tag{1.3}$$

where δ and β are discount parameters, and $u(c)$ is a continuous, strictly concave function, $u : [0, \infty) \rightarrow \overline{\Re}$. Unless otherwise stated, I assume that u is a member of the class of CRRA utility functions (with relative risk aversion coefficient $\rho \in (0, \infty)$). When I work in this class, I need $u(c)$ to be well-defined for all $c \in [0, \infty)$. To preserve the continuity property, set $u(0) = \lim_{c \rightarrow 0} u(c)$. Hence, for $\rho \in [1, \infty)$, $u(0) = -\infty$.

There are two reasons, other than analytical tractibility, to focus on the preferences in equation 1.3. The first motivation is that if u is in the CRRA class and $\beta = 1$, the model reduces to the familiar case of exponential discounting with time-additive homothetic preferences. Hence the $\beta \neq 1$ case may be thought of as a perturbation to the "standard" macro preferences. If we care about robustness we probably want to know what happens when such perturbed preferences are considered.

The second motivation is more complex. There is a large body of evidence that discount functions are closely approximated by hyperbolas (*i.e.* the discount function is approximated by the curve $\frac{1}{\kappa_1 + \kappa_2 \tau}$). This observation was first made by Herrnstein (1961), in relation to animal behavior experiments. The work was later extended with human subjects (DeVilliers and Herrnstein (1976)). Small amendments have been subsequently proposed to this hyperbolic structure, but the basic shape has not been challenged. The important characteristic of the hyperbolic discount function is that it discounts relatively more heavily than the exponential for events in the near future, but discounts less heavily for events in the distant future. Psychologists, notably Ainslie (1975, 1986, 1992), Prelec (1989), and Loewenstein and Prelec (1992) believe that such hyperbolic discounting may play an important role in generating problems of self-regulation, and may provide a potential explanation for numerous behavioral anomalies. When $0 < \beta < 1$ the discount structure in equation [1.3] mimics the hyperbolic shape, while maintaining most of the analytical tractibility of the exponential case.

The preferences in [1.3] were first analyzed by Phelps and Pollak (1968). However, their choice of this structure was motivated in a different way. Their game is one of imperfect intergenerational altruism, so the players are (non-overlapping) generations. I assume that the different players are temporally distinct selves of a single person. The mathematical analysis which follows can be applied to either interpretation.

There is another important contrast between this essay and the work of Phelps and Pollak. They confine their analysis to a subset of the joint strategy space by limiting their analysis to symmetric Markov strategies that are linear with respect to the current asset stock. Specifically, they consider equilibria that are supported by the following symmetric strategy for all selves: *Consume at rate* λ *whatever the previous history, (set* $c_t = \lambda A_t$*)*. They look for λ values that support this as a Nash equilibrium.[3] I consider the full strategy space.

Before proceeding with my analysis of this model, it is useful to introduce the following notation. Let H_t be the set of feasible histories of the consumption game at time t. An $h_t \in H_t$ history is a $t+1$-element vector, $(A_0, c_0, c_1, c_2, \ldots, c_{t-1}) \in \Re^t$. Let H_t^F represent the set of feasible histories at time t. Let H^F be the set of all feasible histories. Let $A : H^F \rightarrow [0, \infty)$ be the map from feasible histories to asset stocks, such that $A(A_0, c_0, c_1, c_2, \ldots, c_{t-1}) \equiv R^t A_0 - \sum_{i=0}^{t-1} R^{t-i} c_i$. Hence, $A(h_t)$ is the asset stock available to self t after history h_t. Represent the pure strategy space of self t as,

$$S_t \equiv \{s_t \mid s_t : H_t^F \rightarrow [0, \infty), \quad and \quad 0 \leq s(h_t) \leq A(h_t) \ \forall h_t \in H_t^F\}. \tag{1.4}$$

Define the joint strategy space $S \equiv \prod_{t=0}^{\infty} S_t$. Let S^P represent the set of subgame-perfect equilibria of the consumption game. Finally, let $v(s, t, h_{\hat{t}})$ represent the continuation payoff of self t, after history $h_{\hat{t}}$, when equilibrium strategy s is played from time $\hat{t}$ forward.

[3] Actually, all of their equilibria satisfy the stronger condition of subgame perfection.

1.3 The exponential discount case: $\beta = 1$.

The analysis in this section assumes $\beta = 1$, which implies preferences that are standard to economists: exponential discounting with time-additive utility. These preferences were first analyzed by Ramsey (1928). His method of analysis was to assume that behavior would correspond to the precommitment strategy of self zero.

> **Definition 1:** *A* **Ramsey equilibrium** *is an element of S which is subgame perfect and which maximizes the continuation payoff to self zero in every possible subgame.*

It is easy to show that if $\beta = 1$, u is in the CRRA class, and $\delta R^{1-\rho} < 1$, then there exists a unique Ramsey equilibrium. This equilibrium is summarized by the following consumption rule for all selves: *Consume at rate* $\lambda^R = 1 - (\delta R^{1-\rho})^{\frac{1}{\rho}}$. (The assumption on technology, $\delta R^{1-\rho} < 1$, is standard in the macroeconomics literature, and the condition is assumed to hold for the rest of the essay.) The goal of this section is to determine whether the Ramsey equilibrium is a reasonable prediction for the game-theoretic model of section two. I will ultimately conclude that it is. However, the path to that conclusion is surprisingly challenging.

> **Theorem 1:** *Let u be a CRRA utility function with $\rho \in (0,1)$. Fix $\beta = 1$. Then the Ramsey equilibrium is the unique subgame perfect equilibrium.*

Proof: Based on Theorem 4, and hence postponed.

So far so good. Unfortunately, the conclusion of Theorem 1 does not hold when $\rho \geq 1$. This is particularly worrisome because it is common to calibrate models with $1 \leq \rho \leq 2$.

> **Proposition 1:** *Let u be a CRRA utility function with $\rho \in [1, \infty)$. Fix $\beta = 1$. Let $\{c_t^*\}_{t=0}^{\infty}$ be any feasible consumption path. Then $\{c_t^*\}_{t=0}^{\infty}$ can be supported by a subgame perfect equilibrium.*

Proposition 1 says that anything is possible. At first this appears to be a very strong result, but, it is trivial to prove using a rather perverse weakly dominated strategy.

Proof: Consider the following equilibrium strategy for self t:

> *If nobody else has deviated consume* $c_t = c_t^*$. *If anybody else has deviated consume* $c_t = 0$.

Recall that if $\rho \in [1, \infty)$, then $c_t = 0$ implies $u(c_t) = -\infty$. If some self $s < t$ expects c_t to equal zero, then self s's payoff will be negative infinity. Since deviations produce payoffs of $-\infty$, no self has a strict incentive to deviate. This argument also applies off of the equilibrium path. □

This proof makes use of an unrealistic punishment structure, (though it is subgame perfect). One might want to rule out equilibria which rely on weakly dominated strategies. It is straightforward to show that the only weakly dominated strategy is for a self to consume nothing, or everything. Hence, it seems reasonable to restrict attention to subgame perfect strategies that satisfy the condition $0 < c_t < A_t$ in every subgame in which $A_t > 0$. However, this is not strong enough to eliminate or even to reduce the multiplicy in Proposition 1. This is because it is possible to generate infinite punishments without setting consumption to zero. Consider the case $\rho \in (1, \infty)$. If all selves consume at rate λ, (*i.e.* $c_t = \lambda A_t \ \forall t$), where $\lambda \in [1 - (\delta R^{1-\rho})^{\frac{1}{\rho-1}}, 1)$, then all selves have payoff $-\infty$.[4] Recall that infinite punishments make anything possible as a perfect equilibrium.

The next natural step is to see what happens when we exclude equilibria which are supported by such infinitely bad punishments.

> **Definition 2:** *A perfect equilibrium has* **finite payoffs at t,** *if the equilibrium satisfies the following condition: For every* $h_t \in H_t$ *such that* $A_t > 0$, *all selves* $s \geq t$ *receive finite payoffs if no deviations occur from time t forward.*

If $\rho \geq 1$ and if an associated subgame perfect equilibrium has finite payoffs at t, $\forall t$, then in every subgame with a positive asset stock, the path of future consumption

[4]Such infinitely bad punishments with positive consumption levels can also be generated for the case $\rho = 1$.

levels will satisfy the restriction $0 < c_t < A_t$. Hence the finite payoff condition rules out weakly dominated strategies. In addition, the finite payoff condition rules out all strategies that rely on infinite punishments. However, the finite payoff criterion does not meaningfully change the equilibrium set.

Theorem 2: *Let u be a CRRA utility function with $\rho \in [1, \infty)$. Fix $\beta = 1$. Let $\{c_t^*\}_{t=0}^{\infty}$ be any feasible consumption path, such that $\sum_{i=0}^{\infty} \delta^i u(c_i^*) > -\infty$. Then, $\{c_t^*\}_{t=0}^{\infty}$ can be supported by a perfect equilibrium with finite payoffs at t, $\forall t$.*

Proof: Let $U_t^* \equiv \sum_{i=0}^{\infty} \delta^i u(c_{t+i}^*)$. So U_t^* is the payoff to self t on the equilibrium path. The statement of the Theorem assumes, $U_0^* > -\infty$, which implies that $\forall t \geq 0$, $U_t^* > -\infty$. Let A_t^* be the inherited asset stock of self t on the equilibrium path.

First, I'll consider the case $\rho = 1$, (*i.e.* $u(\cdot) = \ln(\cdot)$). Construct the set $\{\lambda_{r,r'} \mid r = 1, \ldots, \infty \quad r' = r, \ldots, \infty\}$, according to the following rules. For each r, choose $\lambda_{r,r}$ such that $0 < \lambda_{r,r} < 1$, and $U_r^* \geq \delta f(\lambda_{r,r}) + \frac{1}{1-\delta} \ln(A_r^*) + \frac{\delta}{(1-\delta)^2} \ln(R)$, where $f(\lambda) \equiv \frac{1}{1-\delta}[\ln(\lambda) + \frac{\delta}{1-\delta} \ln(1-\lambda)]$. Such a selection is always possible since $f(\lambda) \to -\infty$ as $\lambda \to 0$. Now, given $\lambda_{r,r'}$, with $r' \geq r$, pick $\lambda_{r,r'+1}$ such that $0 < \lambda_{r,r'+1} < 1$ and $\delta f(\lambda_{r,r'+1}) \leq f(\lambda_{r,r'})$. Use this rule to recursively generate the entire set of λ values.

Consider the following strategy:

If no previous self has deviated, consume $c_t = c_t^*$. If self r was the first to deviate, and self r' was the last to deviate, consume at rate $\lambda_{r,r'}$.

The remaining step is to confirm that this strategy supports a perfect equilibrium. Suppose that the current self is self t, and that no previous self has deviated. Then,

$$U_t^* \geq \delta f(\lambda_{t,t}) + \frac{1}{1-\delta} \ln(A_t^*) + \frac{\delta}{(1-\delta)^2} \ln(R)$$

$$= \ln(A_t^*) + \sum_{i=1}^{\infty} \delta^i \ln(A_t^* R^i (1 - \lambda_{t,t})^{i-1} \lambda_{t,t})$$

$$\geq \ln(c) + \sum_{i=1}^{\infty} \delta^i \ln((A_t^* - c) R^i (1 - \lambda_{t,t})^{i-1} \lambda_{t,t}) \quad \forall c \in [0, A_t^*].$$

Hence, no self has an incentive to deviate after histories in which no previous self has deviated.

Now suppose that the history is such that self r was the first to deviate and self r' was the last to deviate. The current period is $t > r'$. WLOG assume $A_t = 1$, and $R = 1$ (for $A_t \neq 1$, $R \neq 1$, carry the appropriate constants through the following

inequalities). If self t follows the equilibrium path, its payoff is

$$\sum_{i=0}^{\infty} \delta^i \ln(A_t(1-\lambda_{r,r'})^i R^i \lambda_{r,r'})$$

$$= f(\lambda_{r,r'}) \geq \delta f(\lambda_{r,r'+1}) \geq \ldots \geq \delta^{t-r'} f(\lambda_{r,t}) \geq \delta f(\lambda_{r,t})$$

$$= \ln(A_t) + \sum_{i=1}^{\infty} \delta^i \ln(A_t(1-\lambda_{r,t})^{i-1} R^i \lambda_{r,t})$$

$$\geq \ln(c) + \sum_{i=1}^{\infty} \delta^i \ln((A_t - c)(1-\lambda_{r,t})^{i-1} R^i \lambda_{r,t}) \quad \forall c \in [0, A_t].$$

Hence, no self has an incentive to deviate after histories in which previously selves have deviated. Finally, note that any constant consumption rate bounded strictly between zero and one implies a finite payoff for all selves. This completes the proof for the case $\rho = 1$.

Now consider the case $\rho > 1$. The general argument used for $\rho = 1$ may also be used for this case. However, two differences arise. First, it is necessary to construct a new set $\{\lambda_{r,r'} \mid r = 1, \ldots, \infty \quad r' = r, \ldots, \infty\}$, which is contingent on the value of ρ. Moreover, when the λ values are chosen, they must all satisfy the condition $\lambda < 1 - (R^{1-\rho}\delta)^{\frac{1}{\rho-1}}$. This guarantees that the continuation payoffs are finite. □

The proof of Theorem 2 is long, but the idea is very simple. Since $\sum_{i=0}^{\infty} \delta^i u(c_i^*) > -\infty$, it can be shown that for all t, $\sum_{i=0}^{\infty} \delta^i u(c_{t+i}^*) > -\infty$. Hence, for each self one can construct a finitely bad punishment that ensures that that self will not want to deviate from the equilibrium path. The challenge is to show that these punishments are credible. This is done with sequences of cascading finite punishments, each one worse than the last.

Theorem 1 establishes that the multiplicity in Proposition 1 does not rely on infinitely bad punishments. One consequence of this Theorem and the previous Proposition is that we may now be very suspect of the concept of subgame perfection. Perfection is obviously not strong enough to ensure that the Ramsey equilibrium is the only equilibrium. The rest of this section considers a menu of three mutually compatible refinements, each of which eliminates the multiplicity and leaves the Ramsey solution as the only subgame-perfect outcome.

1.3.1 Phelps and Pollak's uniqueness result.

As stated earlier, Phelps and Pollak (1968) were the first to consider the game described in section two. They looked for equilibria in which all players follow the same linear, Markov strategy both on and off the equilibrium path: *Consume at rate λ whatever the previous history.* They did note, however, that other equilibria might exist. Given the assumption on technology, they show that there exists a unique Nash equilibrium which is supported by their rule. Using that result it is straightforward to show that subject to their restrictions there exists a unique subgame perfect equilibrium. This is the Ramsey equilibrium. Obviously, we may be interested in seeing what happens when we consider more general strategy spaces.

1.3.2 Bounded payoffs.

The second refinement is related to the earlier idea of eliminating infinite punishments. The new approach is to consider what happens when we eliminate equilibria that rely on cascades of ever-worsening punishments. The approch is motivated by the following result.

Theorem 3: *Let u be any bounded, continuous real-valued function $u : [0,\infty) \to \Re$. Fix $\beta = 1$, and $\delta < 1$. Then every subgame-perfect equilibrium is a Ramsey equilibrium.*

Proof: Let C represent the space of bounded, continuous functions, $f : [0,\infty) \to \Re$. Let

$$V(A) \equiv \left\{ v \mid v = v(s,t,h_t), \text{ for some } s \in S^P, \text{ and } h_t \text{ s.t. } A(h_t) = A \right\}.$$

Let $\underline{v}(A) \equiv \inf V(A)$. So $\underline{v} : [0,\infty) \to \Re$. The Theorem follows from the following three lemmas.

Lemma 1: *$\underline{v}$ is a fixed point of the functional equation,*

$$(Tf)(A) = \max_c \{u(c) + \delta f((A-c)R)\}.$$

Proof of Lemma 1: Fix any $A \in [0,\infty)$. The first step is to show, $\underline{v}(A) \leq \max_c\{u(c) + \delta\underline{v}((A-c)R)\}$. Construct a sequence $\{s^n, t^n, h_t^n\}$ such that $A(h_t^n) = A$ for all n, and $\lim_{n\to\infty} v(s^n, t^n, h_t^n) = \underline{v}(A)$. Then,

$$v(s^n, t^n, h_t^n) = \max_c\{u(c) + \delta v(s^n, t^n + 1, \{h_t^n, c\})\},$$

which implies that,

$$v(s^n, t^n, h_t^n) \geq \max_c\{u(c) + \delta\underline{v}((A-c)R)\}.$$

Taking the limit of the LHS yields,

$$\underline{v}(A) \geq \max_c\{u(c) + \delta\underline{v}((A-c)R)\}. \tag{1.5}$$

The next step is to show, $\underline{v}(A) \leq \max_c\{u(c) + \delta\underline{v}((A-c)R)\}$. For this part of the proof I focus exclusively on subgame-perfect equilibria for which the payoff to self one depends exclusively on A_1. I assume that $A_0 = A$. Hence, it is possible to represent the continuation payoff to player one as $v(s, 1, (A-c_0)R)$. If $s \in S^P$ then

$$\underline{v}(A) \leq \max_c\{u(c) + \delta v(s, 1, (A-c)R)\}.$$

Construct $\{s^n\}_{n=1}^{\infty}$, such that $s^n \in S^P$ for all n, and $v(s^n, 1, \cdot) - \underline{v}(\cdot) \leq \frac{1}{n}$ for all n. (Hence, $v(s^n, 1, (A-c)R)$ uniformly converges to $\underline{v}((A-c)R)$.) So,

$$\underline{v}(A) \leq lim_{n\to\infty} \max_c\{u(c) + \delta v(s^n, 1, (A-c)R)\},$$

which implies that,

$$\underline{v}(A) \leq \max_c\{u(c) + \delta\underline{v}((A-c)R)\}. \tag{1.6}$$

Combining equations 1.5 and 1.6 proves the lemma. □

Lemma 2: $\underline{v} \in C$.

Proof of Lemma 2: First, note that $\underline{v}$ is bounded since u is bounded and $\delta < 1$. It remains to show that $\underline{v}$ is continuous. Fix any $A \in [0,\infty)$, and any sequence $\{A_n\}_{n=1}^{\infty}$ such that $A_n \in [0,\infty)$ for all n, and $\lim_{n\to\infty} A_n = A$. Seek to show $\lim_{n\to\infty} \underline{v}(A_n) = \underline{v}(A)$.

Construct a sequence of subgame-perfect equilibrium strategies, $\{s^n\}_{n=1}^{\infty}$ such that $v(s^n, 0, A_n) - \underline{v}(A_n) \leq \frac{1}{n}$. Construct a second sequence of subgame-perfect equilibrium strategies $\{\hat{s}^n\}_{n=1}^{\infty}$ such that $v(\hat{s}^n, 0, A) - \underline{v}(A) \leq \frac{1}{n}$.

Let $\hat{c}^n(A)$ represent the equilibrium consumption of self 0, according to strategy $\hat{s}^n$. Then,

$$v(s^n, 0, A_n) \geq \underline{v}(A) - [u(\hat{c}^n(A)) - u(\hat{c}^n(A) - (A - A_n))],$$

which implies that,

$$\underline{v}(A_n) + \frac{1}{n} \geq \underline{v}(A) - [u(\hat{c}^n(A)) - u(\hat{c}^n(A) - (A - A_n))].$$

Taking limits of both sides yields,

$$\lim_{n\to\infty} \underline{v}(A_n) \geq \underline{v}(A). \tag{1.7}$$

Let $c^n(A)$ represent the equilibrium consumption of self 0, according to strategy s^n. Then,

$$v(\hat{s}^n, 0, A) \geq \underline{v}(A_n) - [u(c^n(A_n)) - u(c^n(A) + (A - A_n))],$$

which implies that,

$$\underline{v}(A) + \frac{1}{n} \geq \underline{v}(A_n) - [u(c^n(A_n)) - u(c^n(A) + (A - A_n))],$$

Taking limits of both sides yields,

$$\underline{v}(A) \leq \lim_{n\to\infty} \underline{v}(A_n). \tag{1.8}$$

Combining equations 1.7 and 1.8 yields

$$\lim_{n\to\infty} \underline{v}(A_n) = \underline{v}(A). \square$$

Lemma 3: *(Stokey et al. (1989), Theorem 4.6) The functional equation,*

$$(Tf)(A) = \max_c \{u(c) + \delta f((A - c)R)\}$$

has a unique fixed point in C.

Together, Lemmas 1, 2, and 3 imply that the continuation payoffs associated with $\underline{v}$ are equivalent to the Ramsey continuation payoffs. □

Theorem 3 establishes that when $\beta = 1$ and the felicity function is bounded (and continuous), all subgame perfect equilibria are Ramsey equilibria.[5] Note that for each value of ρ the CRRA felicity function is unbounded. This explains why Theorem 3 is consistent with Theorem 2. The goal of this section is to develop an equilibrium refinement for the CRRA consumption game which captures the idea

[5] If u is not strictly concave there may not be a unique Ramsey equilibrium.

of boundedness. Specifically, I restrict attention to equilibria that have the property that for any given equilibrium the set of "normalized" continuation payoffs associated with that equilibrium is bounded. The following definition makes this idea precise.

Definition 3: *Fix a strategy profile, s. Let*

$$\Omega(s) \equiv \{\omega \mid \exists \text{ a period } t, \text{ feasible history } h_t, \text{ s.t. } A(h_t) > 0 \text{ and } A(h_t)^{1-\rho} \cdot \omega = v(s,t,h_t)\}.^{6}$$

We will say that s has **bounded normalized continuation payoffs** *(BNCP) if $\inf \Omega(s)$ is finite. Finally, if u is an instantaneous utility function, let $P^{BNCP}(u)$ represent the associated set of subgame perfect equilibria which satisfy the BNCP property.*

Theorem 4: *Let u be a CRRA utility function. Fix $\beta = 1$. The Ramsey equilibrium is the unique subgame perfect equilibrium with the BNCP property.*

Proof: There are three relevant cases: $\rho \in (0,1)$, $\rho = 1$, and $\rho > 1$. I'll prove the Theorem for the first case. The treatment of the remaining two cases uses the same approach.

Assume $\rho \in (0,1)$, and assume that there exists a perfect equilibrium s with the property that $W(s)$ has a well-defined (finite) inf, $\underline{\omega}(s)$. Let $A(h)$ represent the current asset stock when the previous history is h. Let $\omega(h_t) \equiv v(s,t,h_t)\frac{1-\rho}{(A(h_t))^{1-\rho}}$. The existence of finite $\underline{\omega}(s)$, implies that there exists a sequence of histories $\{h^n_{t(n)}\}_{n=1}^{\infty}$, such that $\underline{\omega}(s) = \lim_{n\to\infty} \omega(h^n_{t(n)})$. Note that the superscript n signifies the location in the sequence. No structure is imposed on the function $t(n)$.

Consider some history h_t. If this history arises, the continuation payoff of self t is given by $\frac{(A(h_t))^{1-\rho}}{1-\rho}\omega(h_t)$. Since s is a perfect equilibrium, this payoff is bounded below by,

$$\frac{c^{1-\rho}}{1-\rho} + \frac{\delta[R(A(h_t)-c)]^{1-\rho}}{1-\rho}\underline{\omega}(s) \quad \forall c \in [0, A(h_t)].$$

Let $\hat{c}$ be the value of c that maximizes that lower bound. Using the first order condition (which is necessary and sufficient for this problem) it is easy to show

$$\hat{c} = \frac{A(h_t)}{1 + (\underline{\omega}(s)\delta R^{1-\rho})^{\frac{1}{\rho}}}$$

[6] If $\rho = 1$, let $\Omega(s)$ be as above but replace $A(h_t)^{1-\rho}\cdot\omega$ with $\frac{1-\delta(1-\beta)}{1-\delta}\ln(A(h_t)) + \omega$.

Plugging this value of c back into the expression for the lower bound and dividing through by $\frac{(A(h_t))^{1-\rho}}{1-\rho}$ yields the following expression for $\omega(h_t)$.

$$\omega(h_t) \geq \left[1 + (\underline{\omega}(s)\delta R^{1-\rho})^{\frac{1}{\rho}}\right]^{\rho}$$

Replacing h_t, with $h^n_{t(n)}$ and taking the limit of the LHS as $n \to \infty$, yields,

$$\underline{\omega}(s) \geq \left[1 + (\underline{\omega}(s)\delta R^{1-\rho})^{\frac{1}{\rho}}\right]^{\rho}.$$

Simplification implies that,

$$\underline{\omega}(s) \geq \left[1 + (\delta R^{1-\rho})^{\frac{1}{\rho}}\right]^{-\rho}.$$

Compare this to the payoff that would be received under the Ramsey equilibrium. It is tedious, but not hard to show that if the current period is t, with asset stock A_t, the Ramsey payoff to self t is equal to

$$\frac{A_t^{1-\rho}}{1-\rho}\left[1 + (\delta R^{1-\rho})^{\frac{1}{\rho}}\right]^{-\rho},$$

so the normalized Ramsey payoff is

$$\left[1 + (\delta R^{1-\rho})^{\frac{1}{\rho}}\right]^{-\rho}.$$

The last inequality on $\underline{\omega}(s)$ implies that the worst normalized continuation payoff of equilibrium s is at least as good as the normalized Ramsey payoff. But the Ramsey payoff is also the upper bound on $W(s)$. It follows that $W(s)$ is a one point set with all continuation payoffs equal to the normalized Ramsey payoff. □

I am now ready to prove Theorem 1.

Proof of Theorem 1: Note that if $\rho \in (0,1)$, then $u(\cdot)$ is bounded from below by 0. Hence, if $\rho \in (0,1)$, then any subgame perfect equilibrium has the BNCP property. By Theorem 4, the Ramsey equilibrium is the unique subgame perfect equilibrium. □

1.3.3 Finite horizons.

Consider the finite-horizon analog to the game described in section 2, with $\beta = 1$. If the horizon is T, there are $T+1$ selves, and preferences are given by

$$U_t(c_0, c_1, \ldots, c_T) = \sum_{i=0}^{T-t} \delta^i u(c_{t+i}). \tag{1.9}$$

In subsection 4.2, I show that for the general case $0 < \beta \leq 1$ the finite horizon game has a unique perfect equilibrium. In addition, I show that as $T \to \infty$ these finite horizon equilibria converge to a perfect equilibrium of the infinite horizon game. For the case $\beta = 1$ this limiting equilibrium corresponds to the Ramsey equilibrium. Hence, finite horizon arguments provide another means of picking out the Ramsey equilibrium.

This completes the analysis when $\beta = 1$. For this case, I am very comfortable settling on the Ramsey equilibrium. However, that is not the central issue. More importantly, it may serve the reader to decide, before proceeding, which arguments in favor of the Ramsey outcome are most persuasive. If these refinements are reasonable you may also want to accept their implications in the case $0 < \beta < 1$.

1.4 Non-exponential discounting: $0 < \beta < 1$.

It is trivial to extend the multiplicity results in Proposition 1 and Theorem 1 to the case $0 < \beta < 1$. Hence, it is interesting to see which of the earlier refinements can eliminate this multiplicity. It turns out that only two of the original three refinements continue to imply uniqueness. These two "successful" refinements are discussed first.

1.4.1 Phelps and Pollak revisited.

Once again, Phelps and Pollak's restriction to the set of symmetric, linear Markov strategy, yields a unique perfect equilibrium. Their equilibrium consumption rate, λ^*

satisfies the equation,

$$\lambda^* = 1 - \left(\delta R^{1-\rho}[\lambda^*(\beta - 1) + 1]\right)^{\frac{1}{\rho}}. \tag{1.10}$$

Henceforth, I will refer to their equilibrium strategy as the Phelps-Pollak strategy. Note that this reduces to the Ramsey equilibrium strategy when $\beta = 1$.

1.4.2 Finite horizons revisited.

In this subsection, I show that there is a unique perfect equilibrium in the finite-horizon game, and I consider its limiting properties as the horizon goes to infinity. This analysis formalizes the arguments in subsection 3.3, by proving more general results. All of the results in the current subsection apply to the case $0 < \beta \leq 1$.

I begin by analyzing finite horizon games. Note that the T-horizon game has $T+1$ players, and preferences given by,

$$U_t(c_0, c_1, \ldots, c_T) = u(c_t) + \beta \sum_{i=1}^{T} \delta^i u(c_{t+i}) \quad \forall t.$$

Proposition 2: *For any T-horizon game, there exists a unique subgame-perfect equilibrium. This equilibrium is Markov perfect.*

Proof: Let s_t^T be a point in the strategy space of self t in a game with horizon T. So $s_t^T : H_t^F \rightarrow [0, \infty)$. Suppose that the T-horizon game has a unique perfect equilibrium. Also suppose that this equilibrium has strategies of the form: $s_t^T(h_t) = \lambda_{T-t} A_t$, for all selves $t \in \{0, 1, \ldots, T\}$. Finally, assume $0 < \lambda_{T-t} < 1$ for all $t \in \{0, 1, 2, \ldots, T-1\}$. Let $V(A, T+1) \equiv \beta\delta \sum_{t=0}^{T} \delta^t u(\lambda_{T-t} A_t)$, where $A_0 = A$, and the rest of the A_t sequence is built up recursively: $A_{t+1} = R(1 - \lambda_{T-t})A_t$. It is easy to show $V_A(A, T+1) > 0$ and $V_{AA}(A, T+1) < 0 \; \forall \; A \in (0, \infty)$, and $\lim_{A \rightarrow 0} V_A(A, T+1) = \infty$. Now consider the behavior of self 0 in a game with horizon $T+1$. Since there is a unique subgame perfect equilibrium in the subgame that arises after self 0's choice, self 0 chooses a consumption level to maximize, $u(c_0) + V(R(A_0 - c_0), T+1)$, subject to the restriction $0 \leq c_0 \leq A_0$. The properties of $u(\cdot)$ and $V(\cdot, T+1)$ imply that this problem has a unique interior solution. It is easy to show that the chosen consumption level is proportional to, but less than, A_0. Hence, there exists a number λ, $0 < \lambda < 1$ such

that $c_0 = \lambda A_0 \ \forall A_0$. Set $\lambda_{T+1} \equiv \lambda$. The proof proceeds by induction. To start the induction, simply observe that $\lambda_0 = 1$. □

Since the unique subgame perfect equilibrium in any finite-horizon game is a sequence of Markov strategies, we can write the equilibrium strategy of self t in a T-horizon game as, $c_t(A_t, T)$.

Proposition 3: *Consider a T-horizon game. The following "Euler equation" holds on the unique equilibrium path.*

$$u'(c_t) = R\delta u'(c_{t+1}) \left[\frac{\partial c_{t+1}(A_{t+1}, T)}{\partial A_{t+1}} (\beta - 1) + 1 \right] \tag{1.11}$$

Proof: Continuing the argument from the proof of Proposition 2, note that in equilibrium the following condition holds for all t :

$$u'(c_t) = RV_A(A_{t+1}, T-t).$$

Note that $V(A_{t+1}, T-t)$ can be reexpressed, $V(A_{t+1}, T-t) =$

$$\beta\delta u(c_{t+1}(A_{t+1}, T)) + \delta V(R(A_{t+1} - c_{t+1}(A_{t+1}, T)), T-(t+1))$$

Taking a partial derivative, yields, $V_A(A_{t+1}, T-t) =$

$$\beta\delta u'(c_{t+1}) \frac{\partial c_{t+1}}{\partial A_{t+1}} + \delta R V_A(R(A_{t+1} - c_{t+1}), T-(t+1)) \left[1 - \frac{\partial c_{t+1}}{\partial A_{t+1}} \right].$$

Finally substitute $u'(c_{t+1})$ for $RV_A(A_{t+2}, T-(t+1))$ to get the required result. □

I can now proceed with the main result of this subsection. Theorem 5 establishes that the finite-horizon equilibrium can be used to pick out the Phelps-Pollak equilibrium in the infinite horizon game.

Theorem 5: *As $T \to \infty$, $c_t(A, T)$ converges pointwise to the function $\lambda^* A$, which is the symmetric Markov strategy in the Phelps and Pollak equilibrium.*

Proof: Recall that the proof of Proposition 2 shows that in a game with horizon T,

the unique perfect equilibrium strategy of self t is to consume at rate λ_{T-t}. Given this observation, it is possible to use Proposition 3 to characterize the consumption of self t in a $T+1$-horizon game.

Recall that Proposition 3 states that the following equation holds on the unique equilibrium path of any finite horizon subgame:

$$u'(c_{t-1}) = R\delta u'(c_t)\left[\frac{\partial c_t(A_t, T)}{\partial A_t}(\beta - 1) + 1\right]$$

Assume that the game has horizon T. Substitute in for $u(\cdot)$, replace c_t with $\lambda_{T-t}A_t$, and replace the partial derivative with λ_{T-t}. Solving for c_{t-1} yields,

$$c_{t-1} = \lambda_{T-(t-1)}A_{t-1},$$

where

$$\lambda_{T-(t-1)} = \frac{\lambda_{T-t}}{[\delta R^{1-\rho}(\lambda_{T-t}(\beta - 1) + 1)]^{\frac{1}{\rho}} + \lambda_{T-t}} \tag{1.12}$$

This implies that in a finite horizon game it is possible to calculate the equilibrium consumption rate of today's self from the equilibrium consumption rate of tomorrow's self. Another way of thinking about this is to say that it is possible to calculate self t's equilibrium strategy in the $T+1$-horizon game if we know self t's equilibrium strategy in the T-horizon game.

So far I've noted the following properties. First $c_t(A_t, T) = \lambda_{T-t}A_t$ both on and off the equilibrium path. Second, the sequence of consumption rates $\{\lambda_r\}_{r=0}^{\infty}$, follows the recursion,

$$\lambda_{r+1} = f(\lambda_r) \equiv \frac{\lambda_r}{[\delta R^{1-\rho}(\lambda_r(\beta - 1) + 1)]^{\frac{1}{\rho}} + \lambda_r}$$

Hence, to prove the Theorem it is sufficient to show that $\lambda_r \to \lambda^*$.

In the argument which follows I'll use the following properties of $f(\cdot)$, which are straightforward to confirm.

- $f(0) = 0$.
- $f(\cdot)$ differentiable on $[0, 1]$.
- $f'(0) > 1$ (using technology assumption $\delta R^{(1-\rho)} < 1$).
- $f'(x) > 0$ on $[0, 1]$.
- $f(1) < 1$.

Let $\overline{\lambda} = \sup\{\lambda \mid \lambda \in [0, 1], \lambda = f(\lambda)\}$. There is at least one fixed point at zero, so $\overline{\lambda}$ exists. In fact, it is possible to show that $\overline{\lambda}$ is strictly greater than zero. This follows from the properties $f(0) = 0$, $f'(0) > 1$, $f(\cdot)$ continuous, $f(1) < 1$, and by application of the Intermediate Value Theorem. Finally, $\lambda_r \to \overline{\lambda}$ since $f'(x) > 0$ on $[0, 1]$, $f(1) < 1$, and $\lambda_0 = 1$.

It only remains to show that $\overline{\lambda} = \lambda^*$. Recall that $\overline{\lambda} = f(\overline{\lambda})$. Both sides of this equation can be divided by $\overline{\lambda}$ since it has been shown that $\overline{\lambda} > 0$. Transforming the

resulting equality it is easy to show that

$$\overline{\lambda} = 1 - (\delta R^{1-\rho})^{\frac{1}{\rho}} \left[\overline{\lambda}(\beta - 1) + 1\right]^{\frac{1}{\rho}}.$$

Note that the unique solution to this equation is λ^*. □

Before proceeding with the next subsection, it is helpful to extend the analysis of this subsection a little further. Recall equation 1.11, and note that Theorem 5 establishes that as the horizon goes to infinity

$$\frac{\partial c_{t+1}(A_{t+1}, T)}{\partial A_{t+1}} \rightarrow \lambda^*$$

Hence, in the limit,[7] equation 1.11 becomes,

$$u'(c_t) = R\delta u'(c_{t+1}) \left[\lambda^*(\beta - 1) + 1\right] \quad \forall t \tag{1.13}$$

Note that this equation and its finite-horizon analog reduce to the standard Euler equation when $\beta = 1$. It is also interesting to observe that equation 1.13 is observationally equivalent to the Euler equation that arises when discounting is exponential with discount rate $\overline{\delta} = \delta \left[\lambda^*(\beta - 1) + 1\right]$. This is an observational equivalence result which is somewhat similar to the widely cited but false observational equivalence claim of Strotz (1955).[8] Note that my derivation of equation 1.13 depends on two assumptions that Strotz did not make. First, my discounting structure is very restricted, reverting to exponential discounting after one period. Second, I assume CRRA preferences. The specific form of my result depends on both of these assumptions.[9]

However, there is a sense in which my observational equivalence claim is just a special case of a more general phenomenon. With CRRA preferences and a fixed

[7]This can be interpreted formally in the following way. Consider the infinite horizon game. Let s^* be the equilibrium of that game which is picked out in Theorem 5. Equation 1.13 holds on the equilibrium path of s^*.

[8]Strotz claimed that dynamically inconsistent agents would behave as if they had an exponential discount rate equal to their instantaneous rate of discount at time zero. See Phelps (1968) for an explanation of what Strotz did wrong.

[9]With completely general discounting a very different "Euler-equation" is generated. This general equation contains an infinite sequence of marginal utilities and propensities to consume.

interest rate, any constant consumption rate implies a linear relationship between $u'(c_t)$ and $u'(c_{t+1})$. If the consumption rate is λ,

$$u'(c_t) = R\hat{\delta}u'(c_{t+1})$$

where $\hat{\delta} = R^{\rho-1}(1-\lambda)^{\rho}$.

1.4.3 An interesting class of equilibria.

I now return to the general infinite-horizon case. Before discussing the remaining two refinements, it is helpful to describe a class of equilibria to which I will later refer. Consider the following symmetric strategy (symmetric in the sense that all selves follow the same consumption rule).

> *Consume a fraction $\lambda_0 \leq \lambda^*$ of the current asset stock unless some prior self has consumed at a rate different than λ_0. In that case, consume at rate λ^*.*

Call this strategy the Self-Diagnostic rule (SD rule). The rule specifies that selves choose a (cooperative) low consumption rate, unless some previous self has violated the rule. In the case of a previous violation, the current self is instructed to consume at the Phelps-Pollak rate. The strategy uses linear consumption rules both on and off the equilibrium path, and admits the Phelps-Pollak strategy as a special case (*i.e.* $\lambda_0 = \lambda^*$). Also note that the equilibrium path is supported by a "focal" strategy, in the sense that the post-deviation phase corresponds to the unique equilibrium picked out in the previous two subsections. Since the SD strategy depends on both λ_0 and λ^*, represent it as SD_{λ_0,λ^*}.

> **Proposition 4:** *Fix $0 < \beta < 1$, and $0 < \delta < 1$. There exists an $\epsilon > 0$ such that for all λ_0 in the interval $(\lambda^* - \epsilon, \lambda^*)$, the SD_{λ_0,λ^*} rule supports a sub-game perfect equilibrium which Pareto-dominates the outcome associated with the Phelps-Pollak equilibrium.*

Note that two equilibria are Pareto-rankable if all temporal selves of the individual are better off under one of the two equilibria. Proposition 4 is proved by extending a result in Phelps and Pollak (1968).

Proof: Phelps and Pollak show that for small ϵ there exists an interval $(\lambda^* - \epsilon, \lambda^*)$, with the property that if λ is in that interval, an infinite path of consumption at rate λ Pareto-dominates the consumption path of the Phelps-Pollak equilibrium.

The remaining step is to show that this Pareto-superior path is supported as a perfect equilibrium by the SD rule. Assume that at time t no previous self has deviated. So the SD rule implies that self t should consume at rate $\lambda \in (\lambda^* - \epsilon, \lambda^*)$. Suppose self t were to deviate. The SD rule dictates that a current deviation is punished by future consumption at rate λ^*. So what is self t's best deviation? Recall that λ^* has the property that if all future selves consume at λ^* then the current self will also want to consume at λ^*. So self t's best deviation is to consume at rate λ^*. However, we know that consumption at rate λ^* forever makes self t no better off than consumption at rate λ forever. So there is no incentive to deviate from the equilibrium path. By definition of λ^* there is also no incentive to deviate after a history which is off of the equilibrium path. □

One useful implication of this proof is that if an SD_{λ_0,λ^*} strategy supports a perfect equilibrium than the corresponding equilibrium path Pareto-dominates the equilibrium path of the Phelps-Pollak equilibrium.

1.4.4 The BNCP refinement.

Recall the notation of section 3.2. It is trivial that if s is an SD equilibrium, $\Omega(s)$ is a two point set. It can also be shown that these two points are finite. Hence, every SD equilibrium is admissible by the BNCP criterion. Since the SD equilibria are indexed by a nondegenerate interval the BNCP criterion is too weak to provide uniqueness.

1.4.5 Which equilibrium is focal when $0 < \beta < 1$?

Having considered the three refinements, it is now possible to ask if there is a focal equilibrium for the case $0 < \beta < 1$. The most likely candidate is the Phelps-Pollak equilibrium. Subsections 4.1 and 4.2 showed that this equilibrium is uniquely chosen by the Phelps-Pollak refinement and the finite horizon refinement. In addition, the

Phelps-Pollak equilibrium has appealingly simple symmetric strategies which satisfy the Markov property that the sufficient statistic for today's consumption decision is today's asset stock. Finally, the Phelps-Pollak equilibrium satisfies the BNCP criterion.

However, there is also a strong argument against the Phelps-Pollak equilibrium. Proposition 4 shows that there exist SD equilibria that Pareto-dominate the Phelps-Pollak equilibrium. Moreover, the SD rules are symmetric and relatively simple (though they do not have the Markov property), and the SD equilibria satisfy the BNCP criterion.

Finally, there are other perfect equilibria outside of the SD class which also Pareto-dominate the Phelps-Pollak equilibrium. Is the Phelps-Pollak outcome, an SD outcome, or some other outcome most likely? When $\beta = 1$ we had a clear answer. For the case $0 < \beta < 1$, there is not an obvious choice.[10]

1.5 Characterization of the equilibrium set.

This section characterizes the set of subgame perfect equilibria which survive the BNCP criterion, when u is a CRRA utility function. Hence the section characterizes the set $P^{BNCP}(u)$, which I will henceforth shorten to P^{BNCP}. The characterization depends upon two assumptions (in addition to the earlier technology assumption) which respectively restrict the range of δ and β, the discount parameters in my model. The restriction on δ takes the form,

$$\text{(A1)} \qquad (\delta R^{1-\rho})^{\frac{1}{\rho}} > \tfrac{1}{2}$$

This inequality is satisfied for sufficiently large δ. The restriction on β is less clear-cut. In particular, I assume,

$$\text{(A2)} \qquad \beta \text{ sufficiently close to 1.}$$

[10] I am currently working on renegotiation refinements. Preliminary work in this area indicates that some renegotiation criteria, (*e.g.*, Farrell and Maskin's (1989) weak renegotiation-proofness criterion), continue to admit multiple equilibria.

Recall that $\beta = 1$ is the standard exponential discount case. When the model is calibrated with reasonable parameter values, assumption (A2) is not restrictive – *i.e.* the results which depend on (A2) hold for β values in an interval $(\underline{\beta}, 1]$, where $\underline{\beta}$ is "close" to zero. At the end of this section I present some examples which support this claim.

Consider the set, P^{BNCP}, which contains all of the subgame perfect strategy profiles of the BNCP class. Let $\underline{\omega}$ represent the worst normalized continuation payoff in the set of all normalized continuation payoffs of strategy profiles in P^{BNCP}.

$$\underline{\omega} \equiv \min\{\, \omega \,|\, \exists\, s \in P^{BNCP}, \text{ s.t. } \omega \in \Omega(s)\}.$$

For now, I'll assume that $\underline{\omega}$ is well-defined. A formal existence result appears in the main theorem.

Before proceeding to the theorem, it may be helpful to emphasize the reason that we care about $\underline{\omega}$. Let,

$$\omega(\{\lambda_n\}_{n=m}^{\infty}) \equiv u(\lambda_m) + \beta \sum_{i=1}^{\infty} \delta^i u(R^i \lambda_{m+i} \prod_{j=0}^{i-1}(1 - \lambda_{m+j})),$$

so $\omega(\{\lambda_n\}_{n=m}^{\infty})$ is the normalized payoff to the current self if the path of current and future consumption rates is given by $\{\lambda_n\}_{n=m}^{\infty}$.

Proposition 5: *A path of consumption rates, $\{\lambda_t\}_{t=0}^{\infty}$ can be supported by a perfect equilibrium with bounded normalized continuation payoffs iff $\underline{\omega} \leq \omega(\{\lambda_n\}_{n=m}^{\infty})\ \forall\ m \geq 0$.*

Proof: Necessity follows immediately from the definition of $\underline{\omega}$. To show sufficiency, it is helpful to introduce the following notation. Let $\lambda(s,t,h)$ represent the equilibrium consumption rate of self t after observing history h when the equilibrium strategy profile is s. Pick $s^* \in P^{BNCP}$, t^*, and h^*, such that [11]

$$\underline{\omega} = \frac{v(s^*, t^*, h^*)}{A(h^*)^{1-\rho}}.$$

[11] When $\rho = 1$, find $s^* \in P^{BNCP}$, t^*, and h^*, such that $\underline{\omega} = v(s^*, t^*, h^*) - \frac{1-\delta(1-\beta)}{1-\delta} \ln(A(h^*))$.

Note that such a triplet must exist by the definition of $\underline{\omega}$. Finally, for any two time periods r and t, with $r < t$, let $h(r,t) \equiv \{h^*, \hat{\lambda}_r, \hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_{t-1}\}$, where the respective $\hat{\lambda}$'s are the realized sequence of consumption rates from time r to time $t-1$, and h^*, is the history of consumption rates defined above. Now take a path of consumption rates, $\{\lambda_t\}_{t=0}^{\infty}$, such that $\underline{\omega} \leq \omega(\{\lambda_n\}_{n=m}^{\infty})\ \forall\ m \geq 0$. Consider the following strategy profile:

> *If no previous self has deviated, self t consumes at rate λ_t. If $r < t$ was the first self to deviate, self t consumes at rate $\lambda(s^*, t^* + t - r, h(r,t))$.*

This profile is a subgame perfect equilibrium which supports the proposed path of consumption rates. □

The main theorem depends on the property that the Ramsey consumption path, (*i.e.* $\lambda_t = \lambda^R \equiv 1 - (\delta R^{1-\rho})^{1/\rho}\ \ \forall t$), can be supported by a subgame perfect strategy profile with bounded normalized continuation payoffs. Proposition 6 provides a sufficient condition for this property.

> **Proposition 6:** *Given (A1) and (A2), the Ramsey consumption path can be supported by a subgame perfect strategy profile with bounded normalized continuation payoffs.*

Proof: Let $\omega(\lambda) \equiv \omega(\lambda, \lambda, \lambda, \ldots)$. Let $f(\beta) \equiv \omega(\lambda^R) - \omega(\lambda^*(\beta))$, where $\lambda^*(\beta)$ is the Phelps and Pollak consumption rate. This notation is used to emphasize that $\omega(\lambda^*(\beta))$ depends on β in two ways. First, β is a discount rate, so changes in β affect the value of future consumption. Second, β is in the implicit equation which determines λ^*. Note that β only influences $\omega(\lambda^R)$ through the former mechanism as λ^R does not depend on β.

The body of this proof characterizes the value of $f(\cdot)$ in a neighborhood of $\beta = 1$. First, $f(1) = 0$ since $\lambda^*(1) = \lambda^R$. To evaluate $f(\cdot)$ at β values just below unity, I consider $f'(1)$ and $f''(1)$.

$$f'(\beta) = \frac{\partial \omega(\lambda^R)}{\partial \beta} - \frac{\partial \omega(\lambda^*)}{\partial \beta} - \frac{\partial \omega(\lambda^*)}{\partial \lambda^*}\frac{d\lambda^*}{d\beta}$$

Note that $f'(1) = 0$, as $\frac{\partial \omega(\lambda^R)}{\partial \beta}\Big|_{\beta=1} = \frac{\partial \omega(\lambda^*)}{\partial \beta}\Big|_{\beta=1}$, and, $\frac{\partial \omega(\lambda^*)}{\partial \lambda^*}\Big|_{\beta=1} = \frac{\partial \omega(\lambda^R)}{\partial \lambda^R}\Big|_{\beta=1} = 0$. Tedious algebraic manipulations reveal that

$$f''(1) = \frac{1}{\rho}(\lambda^R)^{-\rho}(1-\lambda^R)(1-2\lambda^R),$$

which is positive by (A1) (recall that $\lambda^R = 1 - (\delta R^{1-\rho})^{1/\rho}$). Given that $f(1) = 0$, $f'(1) = 0$, and $f''(1) > 0$, there exists an interval $(\underline{\beta}, 1)$ such that $f(\beta) > 0\ \forall$

$\beta \in (\underline{\beta}, 1)$. Hence, for sufficiently large $\beta < 1$, and δ satisfying (A2), $f(\beta)$ is positive and the $SD_{\lambda^R,\lambda^\bullet}$ rule is a subgame perfect equilibrium. □

Theorem 5: *Assume that the Ramsey consumption path can be supported by a subgame perfect strategy profile with bounded normalized continuation payoffs. Then $\underline{\omega}$ is well defined, and*

$$\underline{\omega} = \omega(\lambda^D(\overline{\lambda}^P), \overline{\lambda}^P, \lambda^R, \ldots) = \omega(\overline{\lambda}^P, \lambda^R, \ldots),$$

where,

$$\lambda^D(\lambda^P) \equiv \underset{\lambda \in [0,1]}{\text{argmax}}\ \omega(\lambda, \lambda^P, \lambda^R, \ldots),$$

$$\overline{\lambda}^P \equiv \max\{\lambda^P \in (0,1) \,\big|\, \omega(\lambda^D(\lambda^P), \lambda^P, \lambda^R, \ldots) = \omega(\lambda^P, \lambda^R, \ldots)\}.$$

Proof: First, note that $\lambda^D(\lambda^P)$ is a function on $(0,1)$. Let

$$g(\lambda^P) \equiv \omega(\lambda^P, \lambda^R, \ldots) - \omega(\lambda^D(\lambda^P), \lambda^P, \lambda^R, \ldots).$$

Let $\lambda^M \equiv \text{argmax}_{\lambda \in [0,1]} \omega(\lambda, \lambda^R, \ldots)$. The following results are straightforward to confirm:

- λ^M is a point in $(0,1)$,
- $\frac{\partial^2 g(\lambda^P)}{(\partial \lambda^P)^2} < 0$ for all $\lambda^P > \lambda^M$,
- $g(\lambda^M) > 0$,
- $\lim_{\lambda^P \to 1} g(\lambda^P) < 0$.

Hence, $\overline{\lambda}^P$ is well-defined. The rest of the proof is based on the following result.

Lemma 4: *Given $\underline{\omega}$, let $\overline{\lambda} \equiv sup\{\lambda \mid \omega(\lambda, \lambda^R, \ldots) = \underline{\omega}\}$. Then,*

$$\underline{\omega} = \omega(\overline{\lambda}, \lambda^R, \ldots) = \omega(\lambda^D(\overline{\lambda}), \overline{\lambda}, \lambda^R, \ldots).$$

Proof of Lemma 4: Fix any subgame perfect equilibrium, s, with normalized continuation payoff to self 0 of ω. Then,

$$\omega \geq u(\lambda) + \delta[(1-\lambda)R]^{1-\rho}[\omega(s,1,\lambda) - (1-\beta)u(\lambda_1(\lambda))], \text{ for all } \lambda \in [0,1],$$

where $\omega(s,1,\lambda)$ represents the normalized payoff to self 1 under equilibrium s if self 0 has played λ; and $\lambda_1(\lambda)$, represents the equilibrium map from self 0's action λ, to self 1's action, λ_1. By assumption, $\underline{\omega} \leq \omega(s,1,\lambda)$, for all s, λ, so,

$$\omega \geq u(\lambda) + \delta[(1-\lambda)R]^{1-\rho}[\underline{\omega} - (1-\beta)u(\lambda_1(\lambda))], \text{ for all } \lambda \in [0,1].$$

It is straightforward to confirm that

$$\overline{\lambda} = \sup\{\lambda \mid \exists \{\lambda_n\}_{n=1}^{\infty} \text{ such that}, \omega(\lambda, \{\lambda_n\}_{n=1}^{\infty}) \geq \underline{\omega}\}.$$

Hence, $\lambda_1(\lambda) \leq \overline{\lambda}$, implying,

$$\omega \geq u(\lambda) + \delta[(1-\lambda)R]^{1-\rho}[\underline{\omega} - (1-\beta)u(\overline{\lambda})], \text{ for all } \lambda \in [0,1],$$

which in turn implies,

$$\omega \geq \omega(\lambda^D(\overline{\lambda}), \overline{\lambda}, \lambda^R, \ldots).$$

By definition of $\underline{\omega}$, there exists a sequence of ω's which converge to $\underline{\omega}$, and for each ω in this sequence the previous inequality holds. Hence,

$$\underline{\omega} \geq \omega(\lambda^D(\overline{\lambda}), \overline{\lambda}, \lambda^R, \ldots). \tag{1.14}$$

Construct a monotonically increasing sequence, $\{\lambda_n\}_{n=1}^{\infty}$ which converges to $\overline{\lambda}$. By definition of $\overline{\lambda}$ the following *equilibrium path* strategy is supportable with a subgame-perfect equilibrium:

Self 0: consume at rate $\lambda^D(\lambda^n)$.
Self 1: consume at rate λ^n.
Selves $t \geq 2$*: consume at rate* λ^R.

Hence $\underline{\omega} \leq \omega(\lambda^D(\lambda^n), \lambda^n, \lambda^R, \ldots)$. Letting λ^n go to unity (and noting that λ^D is continuous in its argument), yields

$$\underline{\omega} \leq \omega(\lambda^D(\overline{\lambda}), \overline{\lambda}, \lambda^R, \ldots). \tag{1.15}$$

Combining equations 1.14 and 1.15, yields,

$$\underline{\omega} = \omega(\lambda^D(\overline{\lambda}), \overline{\lambda}, \lambda^R, \ldots),$$

proving the lemma. □

To complete the proof of Theorem 5 it is sufficient to show that given $\underline{\omega}$, $\overline{\lambda} = \overline{\lambda}^P$. Suppose $\overline{\lambda} > \overline{\lambda}^P$. Then, by definition of $\overline{\lambda}^P$,

$$\omega(\overline{\lambda}, \lambda^R, \ldots) \neq \omega(\lambda^D(\overline{\lambda}), \overline{\lambda}, \lambda^R, \ldots),$$

contradicting Lemma 4. Alternatively, assume, $\overline{\lambda} < \overline{\lambda}^P$. By definition of $\overline{\lambda}^P$, the following *equilibrium path* strategy is supportable by a subgame-perfect equilibrium:

Self 0: consume at rate $\lambda^D(\overline{\lambda}^P)$.
Self 1: consume at rate $\overline{\lambda}^P$.
Selves $t \geq 2$*: consume at rate consume at rate* λ^R.

Hence, by definition of $\overline{\lambda}^P$ and $\overline{\lambda}$ the normalized payoff to self 0 is

$$\omega(\lambda^D(\overline{\lambda}^P), \overline{\lambda}^P, \lambda^R, \ldots) = \omega(\overline{\lambda}^P, \lambda^R, \ldots) < \omega(\overline{\lambda}, \lambda^R, \ldots) = \underline{\omega},$$

contradicting the definition of $\underline{\omega}$. Hence, $\overline{\lambda} = \overline{\lambda}^P$. □

Theorem 5 characterizes the worst continuation payoff which can be supported by a perfect equilibrium of the BNCP class. Proposition 7, which follows, characterizes the best continuation payoff which can be supported by a perfect equilibrium of the BNCP class. Let,

$$\overline{\omega} \equiv \max\{\, \omega \,|\, \exists\, s \in P^{BNCP}, \text{ s.t. } \omega \in \Omega(s)\}.$$

Proposition 7: *Assume that the Ramsey consumption path can be supported by a subgame perfect strategy profile with bounded normalized continuation payoffs. Then $\overline{\omega}$ is equal to the normalized payoff which would be achieved by self zero if self zero could precommit all future consumption rates.*

Proof: It is sufficient to note that from the perspective of self zero the optimal path of consumption rates from period one forward is given by $\lambda_t = \lambda^R$. □

Theorem 5 and Proposition 7 rely on an assumed property: there exists an element of P^{BNCP} which supports the Ramsey consumption path. It is useful to know if this property holds when the model is calibrated with standard parameter values. Proposition 6 establishes that assumptions (A1) and (A2) are sufficient for this property to hold. Hence, an interesting exercise is to determine the restrictiveness of (A1) and (A2). Do these assumptions rule out any of the parameter values we would like to use to calibrate this model?

Condition (A1) can be analyzed directly. Table 1 examines three leading cases which span the range of ρ values from which most consumption models are calibrated. Inspection of the table immediately reveals that (A1) is not restrictive since preferences are usually calibrated with $1 \leq \rho \leq 2$, and $.90 \leq \delta \leq 1$.

Condition (A2) cannot be analyzed as easily. Recall, that I seek to describe the calibration range over which there exists a subgame perfect equilibrium which implements the Ramsey path. A sufficient condition for the existence of such an equilibrium, is the existence of an $SD_{\lambda^R,\lambda^\bullet}$ equilibrium. This approach is taken in the Proposition 6 proof. Table 2 extends this analysis by mapping ρ, δ pairs into intervals of β values which have the property that if β is in that interval a $SD_{\lambda^R,\lambda^\bullet}$ equilibrium exists. Specifically, Table 2 reports the intervals of β values for which

$$\omega(\lambda^R, \ldots) \geq \omega(\lambda^*, \ldots).$$

Table 2 demonstrates that (A2) is not restrictive. For each of the examples that I consider there is a large range of β values for which the Ramsey path can be

Table 1: Restrictiveness of assumption (A1)

	(A1) given ρ from column 1	(A1) given ρ and given $R = 1.03$
$\rho = \frac{1}{2}$	$\delta^2 R > \frac{1}{2}$	$\delta > .70$
$\rho = 1$	$\delta > \frac{1}{2}$	$\delta > .50$
$\rho = 2$	$\sqrt{\frac{\delta}{R}} > \frac{1}{2}$	$\delta > .26$

Table 2: Restrictiveness of assumption (A2)

β intervals for which $\exists\ SD_{\lambda_R,\lambda_*}$ equilibrium given $R = 1.03$, and various ρ, δ pairs		
	$\delta = .9$	$\delta = .95$
$\rho = \frac{1}{2}$	$.334 < \beta \leq 1$	$.245 < \beta \leq 1$
$\rho = 1$	$.003 < \beta \leq 1$	$.012 < \beta \leq 1$
$\rho = 2$	$0 < \beta \leq 1$	$0 < \beta \leq 1$

implemented. Moreover, for $\rho = 2$ the entire unit interval (open at zero) is in the acceptable region.

1.6 Interpreting the results.

In this section I argue that the preceding multiplicity is actually a good thing, since it provides a potential explanation for some puzzling economic phenomena. I consider the consequences of this analysis for the theory of dynamic choice, and then turn to some specific applications which may be of interest to macroeconomists.

1.6.1 Dynamic choice.

There are three general conclusions which I want to highlight. First, since there is not a focal equilibrium in the case $0 < \beta < 1$, it seems reasonable to predict that two people with identical preferences would exhibit different behavior in the same environment. Hence, the model generates heterogeneous behavior without making recourse to heterogeneous preferences.

Second, this theory explains why some people have self-acknowledged "bad habits" which they can't break, while other people with identical preferences are able to exert what might be called "self-control" to avoid these "bad habits." The bad habit is a perfect equilibrium which is Pareto-dominated by another perfect equilibrium. Consider a person whose selves are playing the Phelps-Pollak equilibrium, and contrast this individual with a different person whose selves are consuming at a lower rate because they are playing a Pareto-superior equilibrium of the *SD* class. The Phelps-Pollak individual may wish that she had the "self-control" of the *SD* individual. However, the Phelps-Pollak individual is in equilibrium, and, short of some kind of renegotiation, will not achieve the better outcomes of the *SD* person.

Third, the theory makes sense of the observation that behavior is often perceived to be self-diagnostic. On the equilibrium path of the *SD* equilibrium, one self-indulgent act—a high consumption rate today—begets a sequence of self-indulgent acts—high consumption rates in the future.

1.6.2 Macroeconomic applications.

The model is of direct interest to macroeconomists, since the intertemporal choice problem that I study is a savings/consumption decision. Most importantly, the model explains how an economy's/individual's savings rate can be indeterminate in equilibrium, without using a traditional externality argument. The scope of this indeterminacy is demonstrated in the following example.

Let $\underline{\lambda}$ be the smallest λ_0 value for which the $SD_{\lambda_0,\lambda^\bullet}$ strategy supports a perfect equilibrium. It can be shown that for all $\lambda \in [\underline{\lambda}, \lambda^*]$, the $SD_{\lambda,\lambda^\bullet}$ rule supports a perfect equilibrium. Table 3 provides examples of the $[\underline{\lambda}, \lambda^*]$ interval for a small set of β, δ pairs when $\rho = 1$.[12]

Table 3: $[\underline{\lambda}, \lambda^*]$

	$\beta = .25$	$\beta = .50$	$\beta = .75$	$\beta = 1.00$
$\delta = .975$	[.003,.093]	[.011,.049]	[.019,.033]	[.025,.025]
$\delta = .950$	[.009,.174]	[.024,.095]	[.038,.066]	[.050,.050]
$\delta = .925$	[.017,.245]	[.040,.140]	[.059,.098]	[.075,.075]
$\delta = .900$	[.030,.308]	[.057,.182]	[.080,.129]	[.100,.100]

Consider the following particular example. If $\beta = .50$, $\delta = .975$, and an SD

[12]It can be shown that $\rho = 1$ implies that the interval is independent of R. $\lambda^* = \frac{1-\delta}{1-\delta(1-\beta)}$, and $\underline{\lambda}$ is the smaller of the two solutions to the following non-linear equation:

$$(\lambda^*)^{-1} \ln\left(\frac{\lambda}{\lambda^*}\right) + \frac{\beta\delta}{(1-\delta)^2} \ln\left(\frac{1-\lambda}{1-\lambda^*}\right) = 0.$$

equilibrium is adopted, then the individual's long-run consumption rate can be as low as .011 and as high as .049. It is helpful to contrast two hypothetical people (who may be interpreted as two small open economies). Suppose the first person is in an SD equilibrium with $\lambda = .011$. Suppose the second person is in an SD equilibrium with $\lambda = .049$. Take the interest rate to be 3%, (*i.e.* R=1.03). Then the consumption of individual one will grow exponentially over time, while the consumption of individual two will fall exponentially; individual one exhibits a savings rate of 62.2% of income, while individual two exhibits a savings rate of -67.8% of income. However, in terms of their deep preferences these two people are completely identical. They function in the same institutional environment. The only difference is that they implement different perfect equilibria. Moreover, it is hard to argue in favor (from a positive perspective) of one equilibrium over the other. The high consumption rate equilibrium is the Phelps-Pollak equilibrium, and has all of the nice properties discussed above. However, it is Pareto-dominated by the low consumption rate equilibrium. Which equilibrium is more likely to occur?

This analysis suggests a potential rationale for government intervention to try to raise the national savings rate. I have analyzed a model in which low savings rate are Pareto-inferior, but nevertheless arise as an equilibrium outcome. Hence it may make sense to use social institutions to pick out welfare-enhancing equilibria. Perhaps schools should teach students to save.

Finally, the analysis of this chapter can be interpreted as a general methodological critique of the "Euler equation" approach to consumption. When β is less than one there are multiple equilibria, and hence, the usual consumption Euler equation no longer holds. (If a unique Euler equation did hold than there could not be multiple equilibria.) Moreover, on a generic subgame-perfect equilibrium path there will be no systematic relationship between the interest rate and the path of marginal utilities. Marginal utilities take on very little importance because for generic equilibria the decision to deviate is not based on considerations of local perturbations to the consumption path.[13]

[13]An exception to this observation is the Phelps-Pollak equilibrium.

1.7 Conclusion.

This chapter characterizes the equilibria that arise when dynamic savings decisions are modelled as an intra-personal game. I present three refinements which admit a unique equilibrium (the Ramsey equilibrium) when discounting is exponential. However, only two of these refinements continue to admit a unique equilibrium when discounting is approximately hyperbolic. Moreover, the unique equilibrium in those two "successful" cases turns out to be Pareto-inferior in the class of subgame perfect equilibria. This leads me to conclude that there is no single focal equilibrium in the case of hyperbolic discounting. I consider several consequences of this result, with emphasis on the indeterminacy of the national savings rate.

There are three extensions on which I am currently working. The first is to incorporate renegotiation refinements. Preliminary work in this area indicates that some renegotiation criteria, (*e.g.* Farrell and Maskin's (1989) weak renegotiation-proofness criterion) continue to admit multiple equilibria. Second, I have used hyperbolic discounting to develop a model of precommitment (see Chapter 2). Third, I have used hyperbolic discounting to develop a model of self-reward and mental accounts (see Chapter 3).

1.8 References

Ainslie, George W. (1975) "Specious Reward: A Behavioral Theory of Impulsiveness and Impulsive Control." *Psychological Bulletin,* 82, 463-96.

Ainslie, George W. (1986) "Beyond Microeconomics. Conflict Among Interests in a Multiple Self as a Determinant of Value." in Jan Elster, ed., *The Multiple Self,* Cambridge: Cambridge University Press, 133-175.

Ainslie, George W. (1992) *Picoeconomics,* Cambridge: Cambridge University Press.

Akerlof, George A. (1991) "Procrastination and Obedience." *American Economic Review,* AEA Papers and Proceedings, 1-19.

DeVilliers, P., and Richard J. Herrnstein (1976) "Toward a Law of Response Strength." *Psychological Bulletin,* 83, 1131-53.

Farrell, Joseph and Eric Maskin (1989) "Renegotiation in Repeated Games." *Games and Economic Behavior.* 1, 327-360.

Goldman, Steven M. (1980) "Consistent Plans." *Review of Economic Studies,* 47, 533-37.

Herrnstein, Richard J. (1961) "Relative and Absolute Strengths of Response as a Function of Frequency of Reinforcement." *Journal of the Experimental Analysis of Animal Behavior,* 4, 267-72.

Loewenstein, George, and Drazen Prelec. (1992) "Anomalies in Intertemporal Choice: Evidence and an Interpretation." *Quarterly Journal of Economics,* 57:2, 573-598.

Peleg, Bezalel, and Menahem E. Yaari. (1973) "On the Existence of a Consistent Course of Action when Tastes are Changing," *Review of Economic Studies,* 40, 391-401.

Phelps, E. S., and R. A. Pollak. (1968) "On Second-best National Saving and Game-equilibrium Growth." *Review of Economic Studies,* 35, 185-199.

Pollak, R. A. (1968) "Consistent Planning." *Review of Economic Studies,* 35, 201-208.

Prelec, Drazen. (1989) "Decreasing Impatience: Definition and Consequences." Harvard Business School Working Paper.

Ramsey, Frank P. (1928) "A Mathematical Theory of Saving." *Economic Journal,* 38, 543-559.

Schelling, Thomas C. (1983) "Ethics, Law, and the Exercise of Self-Command." in Sterling M. McMurrin, ed., *The Tanner Lectures on Human Values IV.* Salt Lake City: University of Utah Press, 43-79.

Stokey, Nancy L., Robert E. Lucas Jr., and Edward C. Prescott (1989) *Recursive Methods in Economic Dynamics.* Cambridge MA: Harvard University Press.

Strotz, Robert H. (1955) "Myopia and Inconsistency in Dynamic Utility Maximization." *Review of Economic Studies,* 23, 165-180.

Chapter 2

Golden Eggs and Hyperbolic Discounting

2.1 Introduction

> Use whatever means possible to remove a set amount of money from your bank account each month before you have a chance to spend it.[1]
> —savings advice in *New York Times'* "Your Money" column

Many people place a premium on the attribute of self-control. Individuals that have this capacity are able to stay on diets, carry through exercise regimens, show up to work on time, and live within their means. Self-control is so desirable that most of us complain that we don't have enough of it. Fortunately there are ways to compensate for this shortfall. One of the most widely used techniques is precommitment. For example, signing up to give a seminar is an easy way to commit oneself to write a paper. The best evidence that such commitments matter is that they create constraints (*e.g.*, deadlines) which generally end up being binding.

Robert Strotz (1956) was the first economist to formalize a theory of precommitment and to show that precommitment mechanisms could be potentially important determinants of economic outcomes. He showed that when individuals' discount func-

[1] "How to Get Ready for Retirement: Save, Save Save." *New York Times.* March 13, 1993, p. 33.

tions are non-exponential they will strictly prefer to constrain their own future choices. Strotz noted that such precommitment decisions are commonly observed.

> "...we are often willing even to pay a price to precommit future actions (and to avoid temptation). Evidence of this in economic and other social behaviour is not difficult to find. It varies from the gratuitous promise, from the familiar phrase 'Give me a good kick if I don't do such and such' to savings plans such as insurance policies and Christmas Clubs which may often be hard to justify in view of the low rates of return. (I select the option of having my annual salary dispersed to me on a twelve- rather than on a nine-month basis, although I could use the interest!) Personal financial management firms, such as are sometimes employed by high-income professional people (*e.g.* actors), while having many other and perhaps more important functions, represent the logical conclusion of the desire to precommit one's future economic activity. Joining the army is perhaps the supreme device open to most people, unless it be marriage for the sake of 'settling down.' The worker whose income is garnished *chronically* or who is continually harassed by creditors, and who, when one oppressive debt is paid, immediately incurs another is commonly precommiting. There is nothing irrational about such behavior (quite the contrary) and attempts to default on debts are simply the later consequences which are to be expected. Inability to default is the force of the precommitment."

Strotz's list is clearly not exhaustive. In general all illiquid assets provide a form of precommitment, though there are sometimes additional reasons that consumers might hold such assets (*e.g.*, high expected returns and diversification). A pension or retirement plan is the clearest example of such an asset. Many of these plans benefit from favorable tax treatment, and most of them effectively bar consumers from using their savings before retirement. For IRA's, Keogh plans, and 401(K) plans, consumers can access their assets, but they must pay an early withdrawal penalty. Moreover, borrowing against some of these assets is legally treated as an early withdrawal, and hence also subject to penalty. A less transparent instrument for precommitment is an investment in an illiquid asset which generates a steady stream of benefits, but which is hard to sell due to substantial transactions costs, informational problems, and/or incomplete markets. Examples include purchasing a home, buying consumer durables, and building up equity in a personal business. Finally, there exists a class of assets which provide a store of illiquid value, like savings bonds, and certificates of

deposit. All of the illiquid assets discussed above have the same property as the goose that laid golden eggs. The asset is long-lived and promises to generate substantial benefits in the long run. However, these benefits are difficult, if not impossible, to realize immediately. Trying to do so will result in a substantial capital loss.

Instruments with these *golden eggs* properties make up the overwhelming majority of assets held by the U.S. household sector. For example, the Federal Reserve System publication *Balance Sheets For the U.S. Economy 1960-91* reports that the household sector held assets of $24.5 trillion at year-end 1991. Over two-thirds of these assets were illiquid, including $4.3 trillion of pension fund and life insurance reserves, $3.8 trillion of residential structures, $2.9 trillion of land, $2.4 trillion of equity in non-corporate business, $2.1 trillion of consumer durables, and *at least* $.5 trillion of other miscellaneous categories. Finally, note that social security wealth and human capital, two relatively large components of illiquid wealth, are not included in the Federal Reserve Balance Sheets.

Despite the abundance of precommitment mechanisms, and Strotz's well-known theoretical work, intra-personal precommitment phenomena have generally been ignored by economists. This deficit is probably explained by the fact that precommitment will only be chosen by decision-makers whose preferences are dynamically inconsistent, and most economists have avoided studying such problematic preferences. However, there is a substantial body of evidence that preferences are dynamically inconsistent. Specifically, research on animal and human behavior has led psychologists to conclude that discount functions are approximately hyperbolic (see Ainslie, 1992).

Hyperbolic discount functions are characterized by a relatively high discount rate over short horizons and a relatively low discount rate over long horizons. This discount structure sets up a conflict between today's preferences, and the preferences which will be held in the future. For example, from today's perspective, the discount rate between two far off periods, t and $t+1$, is the long-term low discount rate. However, from the time t perspective, the discount rate between t and $t+1$ is the short-term high discount rate. This type of preference change is reflected in many common experiences. For example, this year I may desire to start an aggressive savings plan

next year, but when next year actually roles around my taste at that time will be to postpone any sacrifices another year. In the analysis which follows the decision maker foresees these conflicts and uses a stylized precommitment technology to partially limit the options available in the future.

This framework can be used to resolve many of the outstanding empirical puzzles in the consumption/savings literature. First, I am able to explain why consumption tracks income, while simultaneously explaining how impatient consumers manage to avoid excessive dissaving. Second, the model explains why consumers have a different propensity to consume out of wealth then they do out of labor income. Third, the model explains why Ricardian equivalence should not hold even in an economy characterized by an infinitely-lived representative agent. Finally, the model suggests two reasons for the low level of U.S. and international savings during the 1980's.

The body of this essay formalizes these claims. Section 2 lays out the model. Equilibrium outcomes are characterized in Section 3. Section 4 considers the implications of the model for the macroeconomic issues highlighted above. Section 5 concludes with a discussion of ongoing work.

2.2 The consumption decision

The large number of precommitment devices, discussed above, is good news for consumers. They have access to a wide array of assets that effectively enable them to achieve many forms of precommitment. However, from the perspective of an economist, the abundance poses a challenge. Summarizing this complex institutional setting is difficult, particularly if analytical tractability is a goal. It is hard to model the institutional richness in a realistic way without generating an extremely burdensome number of state variables.

I consider a highly stylized precommitment technology which is amenable to an analytic treatment. Specifically, I assume that consumers may invest in two instruments: a liquid asset x and an illiquid asset z. Instrument z is illiquid in the sense

that a sale of this asset has to be initiated one period before the actual proceeds are received. So a current decision to liquidate part or all of an individual's z holding will generate cash flow next period.[2] By contrast, consumers have instantaneous access to their x holdings.

In later sections I embed consumers in a general equilibrium model. Now, however, I consider the consumer in isolation, and assume that the consumer faces a deterministic sequence of interest rates and wages. For simplicity I assume that asset z and asset x have the same rate of return.

The consumer makes consumption/savings decisions in discrete time $t \in \{1, 2, ..., T\}$. Every time period, t, is divided into four subperiods. In the first subperiod, production takes place. The consumer's liquid assets x_{t-1} and non-liquid assets z_{t-1} — both chosen at time period $t-1$ — yield a gross return of $R_t = 1 + r_t$, and the consumer inelastically supplies one unit of labor. In the second subperiod the consumer receives deterministic labor income y_t and gets access to her liquid savings, $R_t \cdot x_{t-1}$. In the third subperiod the consumer chooses current consumption

$$c_t \leq y_t + R_t x_{t-1}.$$

In the fourth subperiod the consumer chooses her new asset allocations, x_t and z_t, subject to the constraints

$$y_t + R_t(z_{t-1} + x_{t-1}) - c_t = z_t + x_t,$$

$$x_t, z_t \geq 0.$$

The consumer begins life with exogenous endowments $x_0, z_0 \geq 0$.

Restricting z_t to be greater than or equal to zero reflects a liquidity constraint. The results of this essay do not depend on this assumption. However, restricting x_t to be greater than or equal to zero is a critical assumption. If the consumer could

[2] One could alternately assume that instantaneous access to asset z is possible with a sufficiently high transaction cost.

set x_t to any value (positive or negative) then she could always perfectly commit her consumption level. For example, if she foresaw a high level of income next period, she could set x_t negative to force tomorrow's self to save some of that income. This could be implemented by writing a contract with an outside agent requiring the consumer to transfer funds to the outside agent, which the outside agent would then deposit in some (illiquid) account of the consumer. I rule out such contracts—and thereby effectively constrain x_t to be positive—for four reasons.

First, and least importantly, writing such contracts is costly. Second, such contracts are susceptible to renegotiation by tomorrow's self, and in any finite-horizon environment, the contract would unwind. (In the last period renegotiation would occur, generating renegotiation in the second to last period, etc.) Third, writing such a contract reveals a lack of self-control, which is an undesirable signal in a culture that highly values self-control. Fourth, and most importantly, such contracts are explicitly unenforceable in the U.S.[3] To make such a contract work, tomorrow's self must be penalized for not paying the specified funds to the outside agent (note that the transfer is not in the interest of tomorrow's self). However, U.S. courts will not enforce any contract with a penalty of this kind.

> Applying the principle of 'just compensation for the loss or injury actually sustained' to liquidated damage provisions, courts have [...] refused enforcement where the clause agreed upon is held to be *in terrorem*—a sum fixed as a deterrent to breach or as security for full performance by the promisor, not as a realistic assessment of the provable damage. Thus, attempts to secure performance through *in terrorem* clauses are currently declared unenforceable even where the evidence shows a voluntary, fairly bargained exchange (Goetz and Scott, 1977, p. 555)

U.S. contract law is based around the "fundamental principle that the law's goal on breach of contract is not to deter breach by compelling the promisor to perform, but rather to redress breach by compensating the promisee" (Farnsworth, 1990, p. 935). Hence, courts allow contracts to specify "liquidated damages" which reflect losses

[3]I am indebted to Robert Hall for pointing out this fact to me.

likely to be experienced by the promisee, but courts do not allow "penalties" which are independent of such losses. In our case, the promisee—the outside agent—experiences no loss if the consumer fails to make the payment. Hence, the contract may not specify a penalty or liquidated damage, so the contract is incapable of compelling tomorrow's self to make the payment. For a more extensive of these issues see Farnsworth (1990, pp. 935-46), Goetz and Scott (1977), and Rea (1984).

I am now ready to move onto a discussion of preferences. At time t, the consumer has a time-additive utility function U_t with an instantaneous utility function with constant relative risk aversion, ρ. Consumers are assumed to have a discount function of the type proposed by Phelps and Pollak (1967) in a model of intergenerational altruism, and which I subsequently used (see chapter one) to model intra-personal dynamic conflict:

$$U_t = E_t \left\{ u(c_t) + \beta \sum_{\tau=1}^{T-t} \delta^\tau u(c_{t+\tau}) \right\} \tag{2.1}$$

I adopt Equation (1) to capture the properties of a generalized hyperbolic discount function: $\frac{1}{\kappa_1 + \kappa_2 \cdot \tau}$, where the κ's are constants, and τ is the time interval between the current period and the consumption event. An important characteristic of the hyperbolic discount function is that the period-to-period discount rate falls monotonically, a property which is implied (weakly) by Equation (1) (assuming $\beta \leq 1$). Falling discount rates imply that short-horizon events are discounted at a higher discount rate than long-horizon events. Ainslie (1992) reviews a large body of direct and indirect evidence that shows that human and animal behavior is best modelled with discount functions with this property.

The preferences given by Equation (1) are dynamically inconsistent, in the sense that contingent plans made at date t are not optimal from the perspective of the decision-maker at date $t+1$. To see this, note that the marginal rate of substitution between periods $t+1$ and $t+2$ from the perspective of the decision maker at time t is given by, $\frac{u'(c_{t+1})}{\delta u'(c_{t+2})}$, which is not equal to the marginal rate of substitution between those same periods from the perspective of the decision maker at time $t+1$: $\frac{u'(c_{t+1})}{\beta \delta u'(c_{t+2})}$.

To analyze equilibrium behavior in this environment it is standard practice to for-

mally model a consumer as a sequence of temporal selves making choices in a dynamic game (*cf.* Pollak (1968), Peleg and Yaari (1973), and Goldman (1980)). Hence, a $T-$period consumption problem translates into a $T-$period game, with T players, or "selves," indexed by their respective periods of control over the consumption decision. (Note that self t is in control during all of the subperiods at time t.) I look for subgame perfect equilibrium (SPE) strategies of this game.

In chapter one, I considered a related consumption/savings game with no precommitment technology, and infinitely-lived agents. I showed that although the economy has no production or consumption externalities, there exists a continuum of pareto rankable subgame perfect equilibria. I characterized the equilibrium set and considered the macroeconomic implications of the multiplicity for savings and growth. In the current chapter, I eliminate the multiplicity by focusing on finite horizon games.

It is helpful to introduce some standard notation which will be used in the analysis which follows. Let h_t represent a (feasible) history at time t, so h_t represents all the moves which have been made from time 0 to time $t-1$: $\{x_0, z_0, (c_\tau, x_\tau, z_\tau)_{\tau=1}^{t-1}\}$. Let S_t represent the set of feasible strategies for self t. Let $S = \prod_{t=1}^{T} S_t$ represent the joint strategy space of all selves. If $s \in S$, let $s|h_t$ represent the path of consumption and asset allocation levels from t to T which would arise if history h_t were realized, and selves t to T played the strategies given by s. Finally, let $U_t(s|h_t)$ represent the continuation payoff to self t if self t expects the consumption and asset allocation levels from t to T to be given by $s|h_t$.

2.3 Equilibrium strategies

This section characterizes the equilibrium strategies of the game described above. Recall that the agent faces a deterministic (time-varying) sequence of interest rates and a deterministic (time-varying) labor income sequence. Unfortunately, for general interest rate and labor income sequences, it is not possible to use marginal conditions to characterize the equilibrium strategies. This non-marginality property is related to the fact that selves who make choices more than two periods from the end of the

game face a non-convex reduced form choice set, where the reduced form choice set is defined as the consumption vectors which are attainable assuming that all future selves play equilibrium strategies. The non-convexity in the reduced form choice set of self t generates discontinuous equilibrium strategies for self t, which in turn generate discontinuities in the equilibrium payoff map of self $t-1$. Formally,

$$U_{t-1}(s|h_{t-1}\,|\,s_{t+\tau} \text{ are SPE strategies } \forall \tau > 0) \tag{2.2}$$

is discontinuous with respect to s_{t-1}. Discontinuity of (2.2) implies that marginal conditions may not be satisfied at an interior optimum. For an exposition of these problems see Laibson (1993). Related issues are also discussed in Peleg and Yaari (1973) and Goldman (1980).

I have, however, found a restriction on the labor income process which eliminates these problems:

$$\text{A1}: \; u'(y_t) \geq \beta\delta^{\tau}\left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(y_{t+\tau}) \quad \forall \; t,\tau \geq 1.$$

Calibration of the model reveals that this assumption is not restrictive. Ainslie (1992) reviews evidence that the one-year discount rate is at least $\frac{1}{3}$. This suggests that β should be calibrated in the interval $(0, \frac{2}{3})$ (assuming that δ is close to unity). To see what this implies consider the following example. Assume $R_t = R$ for all t, $\delta R = 1$, and $u(\cdot) = \ln(\cdot)$. Then A1 is satisfied if, for all t, $y_t \in [\underline{y}, \frac{1}{\beta}\underline{y}]$. If $\beta = \frac{2}{3}$ this interval becomes $[\underline{y}, \frac{3}{2}\underline{y}]$, and as β falls the interval grows even larger.

Before characterizing the equilibria of this restricted game, it is helpful to introduce the following definitions. First, we will say that a joint strategy, s, is *resource exhausting,* if $s|h_{T-1}$ is characterized by $z_T = x_T = 0$, for all feasible h_{T-1}. Second, we will say that a sequence of feasible consumption/savings actions, $\{c_{\hat{t}}, x_{\hat{t}}, z_{\hat{t}} \ldots, c_T, x_T, z_T\}$ *satisfies P1-P4* if $\forall t \geq \hat{t}$,

P1 $u'(c_t) \geq \max_{\tau \in \{1,...,T-t\}} \beta\delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(c_{t+\tau})$

P2 $u'(c_t) > \max_{\tau \in \{1,...,T-t\}} \beta\delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(c_{t+\tau}) \Rightarrow c_t = y_t + R_t x_{t-1}$

P3 $u'(c_{t+1}) < \max_{\tau \in \{1,...,T-t-1\}} \delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(c_{t+1+\tau}) \Rightarrow x_t = 0$

P4 $u'(c_{t+1}) > \max_{\tau \in \{1,...,T-t-1\}} \delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(c_{t+1+\tau}) \Rightarrow z_t = 0$

Finally, we will say that a joint strategy $s \in S$ *satisfies P1-P4* if for any feasible history $h_{\hat{t}}$, $s|h_{\hat{t}}$ satisfies P1-P4.

It is now possible to state the main Theorem of the chapter. This theorem establishes that the consumption game has a unique equilibrium, and characterizes this equilibrium.

Theorem 1: *Fix any T–period consumption game with exogenous variables satisfying A1. There exists a unique resource-exhausting joint strategy, $s^* \in S$, that satisfies P1-P4, and this strategy is the unique subgame perfect equilibrium strategy of this game.*

Theorem 1 is proved with four intermediate lemmas. These lemmas apply to the game described in Theorem 1.

Lemma 1: *Let s be a resource exhausting element of the joint strategy space S. Assume that s satisfies P1-P4. Then for all histories h_t, strategy s implies $c_t \geq y_t$.*

Proof: Use induction to prove result. Fix a period t and feasible history, h_t. Let $s|h_t = \{c^A_{t+\tau}, x^A_{t+\tau}, z^A_{t+\tau}\}_{\tau=0}^{T-t}$. Assume $c_{t+\tau} \geq y_{t+\tau}$ $\forall \tau \geq 1$. By P1, $u'(c_t) \geq \max_{\tau \in \{1,...,T-t-1\}} \beta\delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(c_{t+\tau})$. If this inequality is strict, then P2 implies $c_t \geq y_t$. So WLOG assume, $u'(c_t) = \max_{\tau \geq 1} \beta\delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(c_{t+\tau})$.

$$
\begin{aligned}
u'(c_t) &= \max_{\tau \in \{1,...,T-t-1\}} \beta\delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(c_{t+\tau}) && \text{by assumption} \\
&\leq \max_{\tau \in \{1,...,T-t-1\}} \beta\delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(y_{t+\tau}) && \text{by assumption} \\
&\leq u'(y_t) && \text{by A1}
\end{aligned}
$$

So $u'(c_t) \leq u'(y_t)$, and hence $c_t \geq y_t$. After confirming that $c_T \geq y_T$, (by resource exhaustion), proof is completed by applying standard induction argument. □

Lemma 2: *Let s^A and s^B be resource exhausting elements of the joint strategy space S. Assume that s^A and s^B satisfy P1-P4. Let $\{c_t^A, x_t^A, z_t^A\}_{t=1}^T$ and $\{c_t^B, x_t^B, z_t^B\}_{t=1}^T$ be the respective paths of actions generated by s^A and s^B. Fix a particular value of t, and assume $c_{t+\tau}^A \geq c_{t+\tau}^B \ \forall \tau \geq 1$, with $c_{t+\tau}^A > c_{t+\tau}^B$ for at least one $\tau \geq 1$. Then $c_t^A \geq c_t^B$.*

Proof: By P1, $u'(c_t^A) \geq \max_{\tau \in \{1,\ldots,T-t-1\}} \beta\delta^\tau \left(\prod_{i=1}^\tau R_{t+i}\right) u'(c_{t+\tau}^A)$. If this is satisfied with equality, then,

$$\begin{aligned}
u'(c_t^B) &\geq \max_{\tau \in \{1,\ldots,T-t-1\}} \beta\delta^\tau \left(\prod_{i=1}^\tau R_{t+i}\right) u'(c_{t+\tau}^B) && \text{by P1} \\
&\geq \max_{\tau \in \{1,\ldots,T-t-1\}} \beta\delta^\tau \left(\prod_{i=1}^\tau R_{t+i}\right) u'(c_{t+\tau}^A) && \text{by assumption} \\
&= u'(c_t^A) && \text{by assumption.}
\end{aligned}$$

Hence, $u'(c_t^B) \geq u'(c_t^A)$, implying $c_t^A \geq c_t^B$. So WLOG assume,

$$u'(c_t^A) > \max_{\tau \in \{1,\ldots,T-t-1\}} \beta\delta^\tau \left(\prod_{i=1}^\tau R_{t+i}\right) u'(c_{t+\tau}^A).$$

By P2, $c_t^A = y_t + R_t x_{t-1}$. If $t=1$, then $c_t^A \geq c_t^B$ since $c_1^B \leq y_1 + R_1 x_0 = c_1^A$. So WLOG assume $t \geq 2$. If $x_{t-1}^B \leq x_{t-1}^A$ then $c_t^B \leq y_t + R_t x_{t-1}^B \leq y_t + R_t x_{t-1}^A = c_t^A$. So WLOG assume $x_{t-1}^B > x_{t-1}^A \geq 0$.

$$\begin{aligned}
0 &< \sum_{\tau=1}^{T-t} \left(\prod_{i=1}^\tau R_{t+i}^{-1}\right)(c_{t+\tau}^A - c_{t+\tau}^B) && \text{by assumption} \\
&\leq \sum_{\tau=1}^{T-t} \left(\prod_{i=1}^\tau R_{t+i}^{-1}\right)(c_{t+\tau}^A - y_{t+\tau}) && \text{by Lemma 1} \\
&= R_t(x_{t-1}^A + z_{t-1}^A) - c_t^A + y_t^A && \text{by res. exhaust.} \\
&= R_t z_{t-1}^A && \text{as } c_t^A = y_t + R_t x_{t-1}.
\end{aligned}$$

So $z_{t-1}^A > 0$, and,

$$\begin{aligned}
u'(c_t^B) &\geq \max_{\tau \in \{1,\ldots,T-t-1\}} \delta^\tau \left(\prod_{i=1}^\tau R_{t+i}\right) u'(c_{t+\tau}^B) && \text{by P3 \& } x_{t-1}^B > 0 \\
&\geq \max_{\tau \in \{1,\ldots,T-t-1\}} \delta^\tau \left(\prod_{i=1}^\tau R_{t+i}\right) u'(c_{t+\tau}^A) && \text{by assumption} \\
&\geq u'(c_t^A) && \text{by P4 \& } z_{t-1}^A > 0.
\end{aligned}$$

Hence, $u'(c_t^B) \geq u'(c_t^A)$, implying $c_t^A \geq c_t^B$. □

Lemma 3: *Let s be a resource exhausing element of the joint strategy space S. Assume that s satsifies P1-P4. Let $\{c_t, x_t, z_t\}_{t=1}^T$ represent the path of actions generated by s. Then $c_t = y_t + R_t x_{t-1}$ $\forall t \geq 2$.*

Proof: Suppose $c_t < y_t + R_t x_{t-1}$ for some $t \geq 2$ and look for a contradiction. By P1 and P2,

$$u'(c_t) = \max_{\tau \in \{1,\ldots,T-t-1\}} \beta\delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i} \right) u'(c_{t+\tau}),$$

so

$$u'(c_t) < \max_{\tau \in \{1,\ldots,T-t-1\}} \delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i} \right) u'(c_{t+\tau}).$$

Hence, by P3, $x_{t-1} = 0$. So $c_t < y_t$, which contradicts Lemma 1. □

Lemma 4: *Let $\{c_t, x_t, z_t\}_{t=1}^T$ be a solution path to the following problem.*

$$\begin{array}{rl} \max_{\{c_t,x_t,z_t\}_{t=1}^T} & u(c_1) + \beta \sum_{\tau=1}^{T-1} \delta^\tau u(c_{1+\tau}) \\ s.t. & x_t, z_t \geq 0 \quad \forall t \geq 1 \\ & c_t \leq y_t + R_t x_{t-1} \quad \forall t \geq 1 \\ & x_t + z_t = R_t(x_{t-1} + z_{t-1}) + y_t - c_t \quad \forall t \geq 1 \\ & x_0, z_0 \textit{ fixed} \\ & x_T = z_T = 0 \\ & \{c_t, x_t, z_t\}_{t=2}^T \textit{ satisfies P1-P4} \end{array}$$

Then $\{c_t, x_t, z_t\}_{t=1}^T$ satisfies P1-P4.

Proof: The first step in the proof is to show that the solution set of the program above is a subset of the solution set of the program below.

$$\begin{array}{rl} \max_{\{c_t,x_t,z_t\}_{t=1}^T} & u(c_1) + \beta \sum_{\tau=1}^{T-1} \delta^\tau u(c_{1+\tau}) \\ \text{s.t.} & x_t, z_t \geq 0 \quad \forall t \geq 1 \\ & c_t \leq y_t + R_t x_{t-1} \quad \forall t \geq 1 \\ & x_t + z_t = R_t(x_{t-1} + z_{t-1}) + y_t - c_t \quad \forall t \geq 1 \\ & x_0, z_0 \text{ fixed} \\ & x_T = z_T = 0 \\ & c_2 \geq y_2 \\ & c_t = y_t + R_t x_{t-1} \quad \forall t \geq 3 \end{array}$$

Henceforth I will refer to these respectively as program I and program II. Note that program II is a convex program with linear constraints, so the Kuhn-Tucker first order conditions are necessary and sufficient for a global optimum. I will return to this fact later in the proof.

The following notation will be used to prove the lemma. Let Ω represent the set of all real vectors, $\omega = \{c_t, x_t, z_t\}_{t=1}^{T}$. Let $C_I \subset \Omega$ ($C_{II} \subset \Omega$) represent the subset of vectors in Ω which satisfy the constraints of program I (II). Let $C_I^* \subset \Omega$ ($C_{II}^* \subset \Omega$) represent the subset of vectors in Ω which are solutions to program I (II).

The first step in the proof is to show $C_I \subset C_{II}$. Fix any $\omega \in C_I$, and let $\omega = \{c_t, x_t, z_t\}_{t=1}^{T}$. Note that the first five constraints of program I are identical to the first five constraints of program II. Also note that if $\{c_t, x_t, z_t\}_{t=2}^{T}$ satisfies P1-P4, then by Lemma 1, $c_2 \geq y_2$, and by Lemma 3, $c_t = y_t + R_t x_{t-1}$ $\forall t \geq 3$. Hence, $\omega \in C_{II}$, implying that $C_I \subset C_{II}$.

The next step is to show $C_I^* \subset C_{II}^*$. Fix any $\omega \in C_I^*$. Fix any $\omega' \in C_{II}^*$, and let $\omega' = \{c_t, x_t, z_t\}_{t=1}^{T}$. Define $\hat{x}_1$ such that $c_2 = y_2 + R_2 \hat{x}_1$. Let ω'' be equivalent to ω' except that x_1 is replaced by $\hat{x}_1$, and z_1 is replaced by $\hat{z}_1 = z_1 - (\hat{x}_1 - x_1)$. Let $U(\omega)$ represent the value of the objective function evaluated at ω. Consider the following two properties of $\omega'' : \omega'' \in C_{II}$, $U(\omega') = U(\omega'')$. Recall that $\omega' \in C_{II}^*$. Then ω'' must also be an element of C_{II}^*. Hence ω'' must satisfy the Kuhn-Tucker conditions of program II, (since the conditions are necessary and sufficient). Using the Kuhn-Tucker conditions and the definition of ω'' it is straightforward to show $\omega'' \in C_I$. Note that $\omega'' \in C_{II}^*$ and $\omega \in C_I^* \subset C_I \subset C_{II}$ imply that $U(\omega'') \geq U(\omega)$. Note that $\omega \in C_I^*$ and $\omega'' \in C_I$ imply that $U(\omega'') \leq U(\omega)$. Hence, $U(\omega) = U(\omega'')$, which implies that $U(\omega) = U(\omega')$. So $\omega' \in C_{II}^*$ and $\omega \in C_I^* \subset C_I \subset C_{II}$ imply that $\omega \in C_{II}^*$. Hence, $C_I^* \subset C_{II}^*$.

I am now ready to complete the proof of the lemma. Let ω be a solution to program I, and let $\omega = \{c_t, x_t, z_t\}_{t=1}^{T}$. So $\{c_t, x_t, z_t\}_{t=2}^{T}$ satisfies P1-P4. Since $C_I^* \subset C_{II}^*$, ω must also satisfy the necessary and sufficient Kuhn-Tucker conditions of program II. Combining these constraints it is straightforward to show that $\{c_t, x_t, z_t\}_{t=1}^{T}$ satisfies P1-P4. $\square$

Proof of main theorem: Suppose there exists two resource exhausting joint strategies, $s^A, s^B \in S$, that satisfy P1-P4. Fix any period t, and any feasible history h_t. Let $s^A|h_t \equiv \{c^A_{t+\tau}, x^A_{t+\tau}, z^A_{t+\tau}\}_{\tau=0}^{T-t}$, $s^B|h_t \equiv \{c^B_{t+\tau}, x^B_{t+\tau}, z^B_{t+\tau}\}_{\tau=0}^{T-t}$, By resource exhaustion and Lemma 2, $c^A_{t+\tau} = c^B_{t+\tau}$ $\forall\ \tau \geq 0$. Hence, by Lemma 3, $x^A_{t+\tau} = x^B_{t+\tau}$ $\forall\ \tau \geq 0$. This in turn implies $z^A_{t+\tau} = z^B_{t+\tau}$ $\forall\ \tau \geq 0$, as a result of the savings constraints. Because the proof started with arbitrary h_t, we can conclude that $s^A = s^B$ proving that there exists a unique resource exhausting joint strategy, $s^* \in S$, that satisfies P1-P4. The second part of the Theorem follows from this uniqueness result, Lemma 4, and a standard induction argument. □

Some of the examples in the remainder of the chapter consider the infinite-horizon game which is analogous to the finite-horizon game discussed above. When doing so I will focus consideration on the equilibrium that is the limit (as the horizon goes to infinity) of the unique finite-horizon equilibrium.

2.4 Analysis

In the following subsections I use the golden eggs model to explain several prominent empirical puzzles in the consumption literature.

2.4.1 Comovement of consumption and income

There is a growing body of evidence that household consumption flows track corresponding household income flows "too" closely, generating violations of the life-cycle/permanent-income consumption model. In particular, household consumption is too sensitive to expected transitory movements in household income. The literature which documents this anomaly is quite large; some notable contributions have been made by Hall and Mishkin (1982), Zeldes (1989), Carroll and Summers (1991), Flavin (1991), Carroll (1992), and Shea (1992).[4]

[4] Although Runkle (1989) is unable to reject the permanent income hypothesis, there are reasons to believe his test lacks power (see Shea (1992)).

My model provides an explanation for these anomalous empirical regularities, since consumption decisions in the model are closely related to the current level of labor income. Recall that Lemma 3 established that on the equilibrium path consumption is exactly equal to the current level of cash flow: $c_t = y_t + R_t x_{t-1}$. This however does not (by itself) imply that consumption will track labor income. That is because x_{t-1} is endogenous. It is important to understand what determines the choice of x_{t-1}. I begin with the observation that on the equilibrium path the marginal propensity to consume out of additional cash flow is unity.

Proposition 1: *Fix a consumption game in which inequality A1 is strictly satisfied. Let $c_t = c_t(R_t x_{t-1}, R_t z_{t-1})$ represent the equilibrium (Markov) consumption strategy of self t. Then,*

$$\frac{\partial c_t}{\partial (R_t x_{t-1})} = 1 \quad \forall t \geq 2,$$

when the partial derivative is evaluated on the equilibrium path.

Proof: By P1, $u'(c_t) \geq \beta\delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(c_{t+\tau})$ $\forall t \geq 2, \tau \geq 0$. Suppose this inequality is satisfied exactly for some t, τ pair. Then $x_{t-1} = 0$ by P3. Hence,

$$\begin{aligned} u'(y_t) &= u'(c_t) && \text{by Lemma 3} \\ &= \beta\delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(c_{t+\tau}) && \text{by assumption} \\ &\leq \beta\delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(y_{t+\tau}) && \text{by Lemma 3} \end{aligned}$$

But $u'(y_t) \leq \beta\delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(y_{t+\tau})$ violates A1, (as A1 is assumed to hold strictly). So WLOG, assume $u'(c_t) > \beta\delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(c_{t+\tau})$ $\forall t \geq 2$, $\tau \geq 0$. Hence, for sufficiently small $|\epsilon| > 0$, $u'(c_t + \epsilon) > \beta\delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(c_{t+\tau})$ $\forall t \geq 2$, $\tau \geq 0$. So by P3 and the uniqueness result of Theorem 1, in the subgame starting after any sufficiently small perturbation to the liquid asset stock, the equality $c_t = y_t + R_t x_{t-1}$, continues to hold, and hence, $\frac{\partial c_t}{\partial (R_t x_{t-1})} = 1$. □

Proposition 1 states that on the equilibrium path each self would choose to consume more during its period of control but is constrained by the level of available cash flow. Self $t-1$ has chosen x_{t-1} to optimally constrain — from the perspective

of self $t-1$ — the consumption of self t. In this way "early" selves manipulate the cash flow process by keeping most assets in the illiquid instrument. Hence, at any given moment the consumer is effectively liquidity constrained, though the constraint is self-imposed.

However, there are limits to the ways in which "early" selves can constrain the choices of "later" selves. Self $t-1$ can only deny self t access to assets which have been accumulated in the past. Self $t-1$ can not deny t access to y_t, labor income at time t. So when y_t is particularly high (*i.e.* cash flow at time t is uncontrollably high), consumption at time t will also be high. This implies that on the equilibrium path, predictable movements in income will tend to be reflected in movements in consumption.

An example may help to make this more concrete. Assume that labor income follows the process: $y_t = \overline{y}(1+g)^{t-1}$, when t is odd, $y_t = \underline{y}(1+g)^{t-2}$, when t is even. Assume that the interest rate is constant and $(1+g)^\rho = \delta R$. (This last relationship is motivated by the steady state results below.) Assume that $\overline{y}$ and $\underline{y}$ are related by the equation:

$$u'(\overline{y}) = \beta\delta R u'(\underline{y}).$$

Finally, assume that $x_0 = 0$, and $0 \leq z_0 < \hat{z}$, where $\hat{z}$ is given by

$$u'(\overline{y}) = \delta R u'(\underline{y} + (R^2 - (1+g)^2)\hat{z}).$$

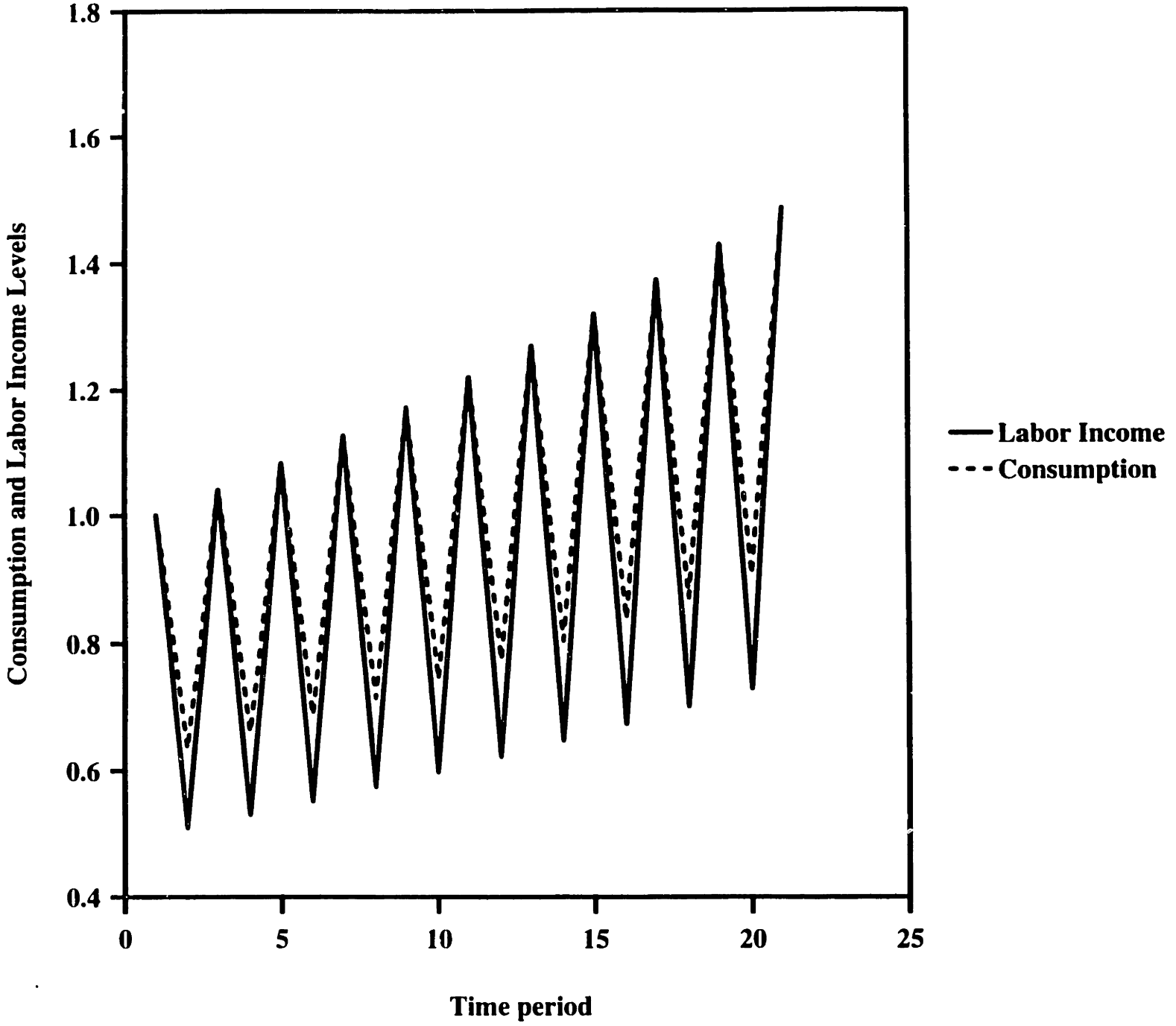
Then the equilibrium consumption path is

$$c_t = y_t + I(t \text{ even})(R^2(1+g)^{t-2} - (1+g)^t)z_0,$$

where $I(t$ even$)$ is the indicator function for even periods.

Figure 1 graphs the labor income path and equilibrium consumption path, using parameter values, $\rho = 1$, $\beta = .5$, $R = 1.04$, $g = .02$, $\overline{y} = \frac{z_0}{3} = 1$. Two properties stand out. First, the illiquid asset is exclusively used to augment consumption in the even periods, *i.e.*, in the periods with relatively low labor income. However, this increase

Figure 1: Labor Income and Consumption



is not sufficient to smooth consumption. A regression of Δc_t on Δy_t yields a coefficient of .75. Since the income process is completely deterministic this implies predictable movements in income are associated with predictable changes in consumption. Hence, consumption tracks income.

2.4.2 Aggregate saving

Carroll (1992) has attempted to explain the comovement of consumption and income by postulating a relatively steep (exponential) intertemporal discount function. (He assumes an annual discount rate of .10.) He exploits the fact that high rates of time discount will generate relatively high levels of consumption-income comovement assuming that agents can not insure against uncertain income flows. However, this approach has an important drawback: high rates of time discount also imply low levels of capital accumulation. Aiyagari (1992) has shown that with a standard general equilibrium model with precautionary savings it is very difficult to reconcile a .10 discount rate with the relatively high historical capital-output ratio ($\frac{K}{Y} = 3$).

By contrast, the golden eggs model generates consumption-income comovement without sacrificing capital formation. This is because in equilibrium decisions to dissave out of the illiquid asset stock do not depend on β. Self t will not be able to consume dissaved assets immediately, so self t does not consider tradeoffs between consumption today and consumption tomorrow when dissaving from the illiquid instrument. Instead self t considers tradeoffs between consumption at $t+1$ and consumption at periods after $t+1$. The value of β is superfluous for such a decision, and hence the steady state capital stock is independent of β.

The following general equilibrium analysis formalizes this claim. Assume that there exists a continuum of individual agents indexed by the unit interval. Individual decision and state variables are represented with an i index (*e.g.* $c_t(i)$). Assume that there exists an aggregate Cobb-Douglas production function given by $Y_t = A_t K_t^{\alpha} L_t^{1-\alpha}$, where $K_t = \int_0^1 [x_{t-1}(i) + z_{t-1}(i)]di$, and $L_t = \int_0^1 L_t(i)di$. Recall that labor is assumed to be supplied inelastically, so $L_t(i) = 1$. In competitive equilibrium labor income of agent i at time period t is given by $y_t(i) = (1-\alpha)Y_t$. Equilibrium also implies

that $r_t = \alpha \frac{Y_t}{K_t} - d$ where d is the rate of depreciation. Finally, A_t is assumed to grow exogenously at rate g_A, so in steady state, capital and output must grow at rate $\frac{g_A}{1-\alpha} \equiv g$.

Proposition 2: *There exists a unique steady state which satisfies A1. In that steady state*

$$(1+g)^\rho = R\delta. \tag{2.3}$$

Proof: If a steady state satisfies A1, then $(1+g)^\rho \geq R\delta$. Moreover, it is easy to show that there exists a steady state at which $(1+g)^\rho = R\delta$. Suppose there exists a steady state at which $(1+g)^\rho > R\delta$. Then by P1-P4, $\lim_{t\to\infty} z_t = \lim_{t\to\infty} x_t = 0$, which implies that no such steady state could exist. □

The important property of the steady state identified in Proposition 2, is that the parameter β does not appear in the equation relating the discount rate and the growth rate. So we are free to calibrate β to generate excess sensitivity, while using δ to match the historical capital-output ratio of three. Any $\beta \in (0,1)$ will suffice to generate consumption-income tracking for sufficiently low z_0. Meanwhile, δ can be chosen to satisfy the equation,

$$\rho g \approx r - (1-\delta) = \alpha \frac{Y}{K} - d - (1-\delta), \tag{2.4}$$

which is a log-linearized version of Equation 2.3. Setting $\delta = 1$, rationalizes $\frac{K}{Y} = 3$, assuming that the other parameters in the equation take standard values: $\rho = 2$, g=.02, α=.36, d =.08.

Hence, the model is able to deliver both excessive consumption-income comovement, and a sufficient level of capital accumulation in a general equilibrium framework.

2.4.3 Asset-specific MPC's

Thaler (1990) argues that consumers have different marginal propensities to consume for different categories of assets. For example, he presents evidence that an unexpected increase in the value of an equity portfolio will have a very small effect on consumption, while an unexpected job-related bonus will be immediately consumed. Thaler divides consumer wealth into three categories: current income, net assets, and future income. He cites a wide body of evidence which suggests that "the MPC from [current income] is close to unity, the MPC from [future income] is close to zero, and the MPC from [net assets] is somewhere in between." Thaler explains this behavior by postulating that consumers use a system of non-fungible mental accounts to guide rule-of-thumb decision-making. By contrast, the golden eggs model predicts that even fully-rational consumers will exhibit asset specific MPC's.[5]

The golden eggs model draws an important distinction between current and future cash flow. As argued above (Proposition 1), the MPC out of current cash flow is unity. In this section I contrast this MPC with its analog for illiquid assets. At first glance it is not clear how to best make this comparison. I'll consider two approaches.

Proposition 3: *Fix a consumption game in which inequality A1 is strictly satisfied. Let* $c_t = c_t(R_t x_{t-1}, R_t z_{t-1})$ *represent the equilibrium (Markov) consumption strategy of self* t*. Then,*

$$\frac{\partial c_t}{\partial (R_t z_{t-1})} = 0 \quad \forall t \geq 2.$$

when the partial derivative is evaluated on the equilibrium path.

Proof: WLOG, assume $u'(c_t) > \beta\delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(c_{t+\tau})$ $\forall t \geq 2$, $\tau \geq 0$, (see Proof of Proposition 1). Hence, for sufficiently small $|\epsilon| > 0$, $u'(c_t) > \beta\delta^\tau \left(\prod_{i=1}^{\tau} R_{t+i}\right) u'(c_{t+\tau} + \left(\prod_{i=1}^{\tau} R_i\right)\epsilon)$ $\forall t \geq 2$, $\tau \geq 0$. So current consumption does not change when z_{t-1} is perturbed. □

[5]Chapter three proposes another hyperbolic discounting model which generates some mental accounting behavior. In chapter three, rational consumers set up a system of self-rewards and self-punishments to motivate later selves to exert high effort. Chapter 3 discusses effort-related mental accounts while the current chapter discusses liquidity related mental accounts.

This result is not surprising, since on the equilibrium path the individual always faces a self-imposed liquidity constraint. Small perturbations to the illiquid asset stock are not sufficient to stop the current self's liquidity constraint from being binding. A more interesting question to ask is how a perturbation to z_{t-1} affects the current choice of x_t. Note that liquid assets set aside today will be consumed tomorrow (Lemma 3). Unfortunately, the value of $\frac{\partial x_t}{\partial(R_t z_{t-1})}$ can take on any value between zero and one. The partial derivative is equal to zero if the equilibrium value of x_t is equal to zero. The partial derivative is equal to unity if t is the penultimate period of the game. It would be helpful to develop an MPC measure which provides a representative value of $\frac{\partial x_t}{\partial(R_t z_{t-1})}$.

Proposition 4: *Fix any $\infty-$horizon consumption game with $R_t = R\ \forall t$. Fix a particular value of $\tau \geq 1$. Assume $\{y_t\}_{t=1}^{\infty}$ satisfies A1, and $y_{t+\tau} = (1+g)^{\tau} y_t$, $\forall t \geq 0$. Assume $(1+g)^{\rho} = \delta R$. Let $x_t = x_t(R_t x_{t-1}, R_t z_{t-1})$ represent the equilibrium (Markov) consumption strategy of self t. Let,*

$$1 - \mathrm{MPC}_t^z \equiv \left[\prod_{i=0}^{\tau-1}\left(1 - \frac{\partial x_{t+i}}{\partial(R z_{t+i-1})}\right)\right]^{\frac{1}{\tau}},$$

evaluated on the equilibrium path. Then,

$$\mathrm{MPC}_t^z = \mathrm{MPC}^z \equiv 1 - (\delta R^{1-\rho})^{\frac{1}{\rho}} \quad \forall t \geq 2.$$

The Proposition is derived from the following lemmas.

Lemma 5: *Fix the economy described in Proposition 4. On the equilibrium path of this game $u'(c_t) = (\delta R)^{\tau} u'(c_{t+\tau}) \forall t \geq 2$, (with τ fixed from Proposition 4).*

Proof of Lemma 5: Suppose $u'(c_t) < (\delta R)^{\tau} u'(c_{t+\tau})$ for some $t \geq 2$. Then P3 implies, $x_{t-1} = 0$, implying,

$$\begin{aligned} u'(y_t) &= u'(c_t) && \text{by Lemma 3} \\ &< (\delta R)^{\tau} u'(c_{t+\tau}) && \text{by assumption} \\ &\leq (\delta R)^{\tau} u'(y_{t+\tau}) && \text{by Lemma 3} \end{aligned}$$

Hence, $y_t^{-\rho} < (\delta R)^\tau y_{t+\tau}^{-\rho}$, implying, $y_t > (1+g)^\tau y_{t+\tau}$, which contradicts the assumptions of Proposition 4.

Alternately, suppose $u'(c_t) > (\delta R)^\tau u'(c_{t+\tau})$ for some $t \geq 2$. Then,

$$\begin{aligned} u'(y_t) &\geq u'(c_t) && \text{by Lemma 1} \\ &> (\delta R)^\tau u'(c_{t+\tau}) && \text{by assumption} \\ &= (\delta R)^\tau u'(y_{t+\tau} + Rx_{t+\tau-1}) && \text{by Lemma 3} \end{aligned}$$

Note that $u'(y_t) = (\delta R)^\tau u'(y_{t+\tau})$, follows from the assumptions. So the previous inequalities imply, $x_{t+\tau-1} > 0$, which together with P3 implies,

$$u'(c_{t+\tau}) \geq \max_{n \geq 1} (\delta R)^n u'(c_{t+\tau+n}).$$

In addition, $x_{t+\tau-1} > 0$, implies that $z_{t-1} > 0$, which together with P4 implies $u'(c_t) \leq \max_{n\geq 1}(\delta R)^n u'(c_{t+n})$. So there exists a finite $\hat{t} \in \{t+1, t+2, \ldots, t+\tau-1\}$, such that $u'(c_{\hat{t}}) > \max_{n\geq 1}(\delta R)^n u'(c_{\hat{t}+n})$. Hence, by P4, $z_{\hat{t}-1} = 0$, contradicting the result that $x_{t+\tau-1} > 0$. □

Lemma 6: *Fix the economy described in Proposition 4. On the equilibrium path of this game* $x_{t+\tau} = (1+g)^\tau x_t$, $z_{t+\tau} = (1+g)^\tau z_t$, $\forall t \geq 1$, *(with* τ *fixed from Proposition 4).*

Proof of Lemma 6: By Lemma 5, $u'(c_t) = (\delta R)^\tau u'(c_{t+\tau})$ $\forall t \geq 2$. Combining this with Lemma 3 implies, $u'(y_t + Rx_{t-1}) = (\delta R)^\tau u'(y_{t+\tau} + Rx_{t+\tau-1})$. The assumptions, $(1+g)^\rho = \delta R$ and $y_{t+\tau} = (1+g)^\tau y_t$ can be used to simplify the previous equation, yielding, $x_{t+\tau-1} = (1+g)^\tau x_{t-1}$ $\forall t \geq 2$. Note that resource exhaustion and Lemma 3 together imply $z_t = \sum_{i=1}^{\infty} R^{-i} x_{t+i}$ $\forall t \geq 1$. So $z_{t+\tau} = \sum_{i=1}^{\infty} R^{-i} x_{t+\tau+i} = \sum_{i=1}^{\infty} R^{-i}(1+g)^\tau x_{t+i} = (1+g)^\tau z_t$ $\forall t \geq 1$. □

Proof of Proposition 4: To prove this proposition I consider two games: an original game, and a perturbed game. The perturbed game is identical to the original game except that in the perturbed game illiquid assets are slightly higher at time zero. I adopt the following notation: Δa represents the difference between variable a in the

perturbed game and variable a in the original game. Then Lemma 3 implies,

$$\Delta z_{t+\tau-1} = R^{\tau}(\Delta z_{t-1})\left(1 - \frac{\Delta x_t}{\Delta(Rz_{t-1})}\right)\cdots\left(1 - \frac{\Delta x_{t+\tau-1}}{\Delta(Rz_{t+\tau-2})}\right),$$

for all $t \geq 2$. Hence,

$$\begin{aligned} 1 - MPC_t^z &\equiv lim_{\Delta z_{t-1}\to 0}\left[\left(1 - \tfrac{\Delta x_t}{\Delta(Rz_{t-1})}\right)\cdots\left(1 - \tfrac{\Delta x_{t+\tau-1}}{\Delta(Rz_{t+\tau-2})}\right)\right]^{\frac{1}{\tau}} \\ &= lim_{\Delta z_{t-1}\to 0}\left(\tfrac{\Delta z_{t+\tau-1}}{\Delta z_{t-1}}\right)^{\frac{1}{\tau}}\tfrac{1}{R} \\ &= \tfrac{(1+g)}{R} \\ &= (\delta R^{1-\rho})^{\frac{1}{\rho}} \end{aligned}$$

where the second to last equality follows from Lemma 6. □

Note that Proposition 4 assumes that the growth rate of labor income is related to the return on capital by the steady state equation in Proposition 2: $(1+g)^{\rho} = \delta R$. Note also that the resulting measure of the marginal propensity to consume, $1 - (\delta R^{1-\rho})^{\frac{1}{\rho}}$ is equivalent to the marginal propensity to consume in the standard Ramsey model with no liquidity constraints and exponential discount function δ^t. For all reasonable parameter values MPC^z is close to zero, contrasting sharply with the unity marginal propensity to consume out of liquid assets.

2.4.4 Ricardian equivalence

In the economy analyzed in this chapter the sequence of exogenous cash flows matters, in a way which is independent of the present value of those cash flows. This is immediately apparent from Figure 1. Because taxation schemes affect these exogenous cash flows, Ricardian equivalence will be violated. Moreover, the model generates such violations even when the consumer has a large asset stock at all times. Hence, Ricardian equivalence is violated for all agents, not, as implied by previous models, just those with no savings.

2.4.5 Declining savings rates in the 1980's

The illiquid asset model may help to explain why the U.S. savings rate was low during the 1980's. I pursue two approaches in this subsection. The first explanation is driven by the fact that during the 1980's a relatively large proportion of national income was realized as cash flow to consumers. The second explanation is driven by developments in the consumer credit market.

Hatsopoulos, Krugman, and Poterba (1989) document the fact that cash flow to consumers (as a percentage of NNP) was high during the 1980's relative to the 1970's. They report that from 1970-79 cash flow averaged 77.9 percent of NNP, while the corresponding number for the 1980-87 period was 80.8 percent. They trace this increase to several sources, notably higher interest income (4.5 percentage points), higher transfers (2.2 percentage points), and higher after-tax cash from takeovers (0.6 percentage points). (There were offsetting falls in cash flow in the following categories: labor income (-.3 percentage points), non-interest capital income in disposable income (-2.0 percentage points), and taxes (-2.1 percentage points).)

Using aggregate data, Hatsopoulos *et al.* estimate a high marginal propensity to consume out of current cash flow. Coupling this result with the higher cash flow levels, they are able to explain most of the savings decline in the 1980's. However, they do not explain *why* consumers should have such a high propensity to consume out of cash flow. The golden eggs model complements their analysis by providing a model which explains the high MPC.

The golden eggs model suggests another explanation for the low level of savings during the the past decade. The 1980's was a period of rapid expansion in the U.S. consumer credit market, and increasing access to credit has reduced the effectiveness of traditional commitment devices. For example, it is now possible to get instant credit lines in many department stores. The golden eggs model predicts that the elimination of commitment devices would lower the level of capital accumulation. I will show that if the credit market were to become sufficiently sophisticated that consumers could instantaneously borrow against their illiquid assets, then the steady state capital-output ratio would fall. I calibrate this fall at the end of the section.

Proposition 5: *Consider the general equilibrium economy analyzed above, but now assume that consumers can instantaneously borrow against their illiquid asset. This economy is equivalent to one in which there is no illiquid asset. In such an economy there exists a unique steady state, such that*

$$(1+g)^{\rho} = \beta\delta R + (1-\beta)\delta(1+g) \tag{2.5}$$

Proof: Chapter one analyzes the economy without the precommitment technology. I show that the infinite horizon equilibrium which corresponds to the limit of the finite horizon equilibria is characterized by constant proportional consumption of the wealth stock, where wealth is defined as the sum of financial assets and the discounted value of future labor income. Let λ represent the coefficient of proportionality; I show that λ is given by,

$$\lambda = 1 - \left[\delta R^{1-\rho}(\lambda(\beta-1)+1)\right]^{\frac{1}{\rho}}.$$

With proportional consumption, the steady state condition is,

$$R(1-\lambda) = (1+g).$$

Solving these equations to eliminate λ yields Equation 2.5. □

Corollary: *In the steady state characterized in Proposition 5 the capital-output ratio is less than the steady state capital-output ratio in the economy with the precommitment technology.*

Proof: Let R^* represent the steady state gross interest rate in the economy with precommitment. Recall Proposition 2: $(1+g)^{\rho} = \delta R^*$. Using Proposition 5 it follows that $R - R^* = (r-g)(1-\beta) > 0$ as $r > g$ is required for the existence of a steady state. □

Table 1 presents evidence about the magnitude of the reduction in steady state capital. I assume that ρ and δ are related by, $(1.02)^{\rho} = \delta(1.04)$. This is the steady state equation for the economy with the precommitment technology, assuming that the real

interest rate in the precommitment economy is .04, and the exogenous growth rate in the precommitment economy is .02. I restrict attention to parameter pairs that satisfy this equation because I am only interested in parameter pairs which are consistent with the steady state relationship which held before the 1980's. Substituting this restriction into Equation 2.5, simplifying, and letting $g = .02$ yields

$$\delta(1.04) = \beta\delta R + (1-\beta)\delta(1.02) \tag{2.6}$$

Note that this reduced-form steady state relationship is independent of ρ and δ, and depends exclusively on β. Table 1 presents steady state r and $\frac{K}{Y}$ values for a range of β values. (The capital-output ratio is calculated with the standard relationship: $r = \alpha\frac{Y}{K} - d$.)

Table 1: Steady state interest rates and capital-output ratios in economies without precommitment technology.

	Δr	$\Delta\frac{K}{Y}$
$\beta = .25$	.060	−1.00
$\beta = .50$	.020	−0.43
$\beta = .75$	.007	−0.16
$\beta = 1.00$	.000	−0.00

For a β value of .5, elimination of the precommitment technology raises the real interest rate two percentage points. This corresponds to a capital stock reduction of 43 percent of output.

2.5 Evaluation and extensions

I have analyzed the consumption problem of a dynamically inconsistent decision-maker who has access to a crude precommitment mechanism. The model helps to explain many of the empirical puzzles in the consumption literature, notably consumption-income tracking and asset-specific MPC's. However, the model has several drawbacks which suggest four important areas to pursue extensions.

The first problem is that the golden eggs model does not explain how consumers accumulate assets in the first place. Note that consumption is always greater than labor income on the equilibrium path. However this is less of a problem than it might first appear. Although there is evidence that individuals often consume less then they earn in labor income, most of this saving is non-discretionary (*e.g.*, pension contributions, life-insurance payments, mortgage payments, and other payments to creditors). Bringing such "non-discretionary savings" into the model can be done very simply. For example, the consumer could elect to take on a 30-period mortgage obligation at time zero, represented by a mortgage payment of m for the next 30 periods. Then the consumer's cash flow at time $t \leq 30$ would be $y_t + R_t x_{t-1} - m$, which would be less than y_t if m were greater than $R_t x_{t-1}$. An almost identical way to model non-discretionary savings would be to let the consumer set x_{t-1} itself less than zero, (*e.g.*, $\underline{x} \leq x_{t-1}$, where $\underline{x} < 0$).

The second problem associated with the model is the anomalous prediction that consumers will always face a binding self-imposed liquidity constraint. For example, the model predicts that consumers should have no liquid funds in their bank accounts and that they should exhaust all of their credit card liquidity. These predictions are not validated by most consumers' experiences. However, this observation can be made consistent with the golden eggs model by introducing a precautionary savings/liquidity motive to the model. For example, consider a continuous-time analog of the golden eggs model, and assume that instantaneous liquidity needs arrive with some hazard rate. Then in equilibrium the consumer will only rarely completely exhaust her liquidity. Formalizing this intuition is part of my current research agenda.

The third problem with the golden eggs model is that some consumers may not need to use external precommitment devices (like illiquid assets) to achieve self-control. Consumers may have internal self-control mechanisms, like "willpower" and "personal rules." In chapter one I analyze an infinite-horizon consumption/savings game with no external precommitment technology and find a multiplicity of Pareto-rankable equilibria. I interpret this multiplicity as a potential model for self-control.

The fourth problem with the golden eggs model is that some consumers may have access to an array of "social" commitment devices that are far richer than the simple illiquid asset proposed in this essay. In chapter three I analyze the problem of a consumer who can use social systems like marriage, work, and friendship to achieve personal commitment. The models proposed in chapters one and three are alternatives to the approach taken in the current chapter. Future work should attempt to compare and contrast these models with empirical analysis.

2.6 References

Ainslie, George W. (1992) *Picoeconomics,* Cambridge: Cambridge University Press.

Aiyagari, S. Rao. (1992) "Uninsured Idiosyncratic Risk and Aggregate Saving," Working Paper 502, Federal Reserve Bank of Minneapolis.

Board of Governors of the Federal Reserve System. (1992) *Balance Sheets For the U.S. Economy 1960-91,* Washington D.C.

Carroll, Christopher D. (1992) "The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence," *Brookings Papers on Economics Activity,* Vol 2, 61-156.

Carroll, Christopher D. (1992) "How does future income affect current consumption?" Board of Governors of the Federal Reserve System (Jan.).

Carroll, Christopher D., and Lawrence H. Summers. (1991) "Consumption Growth Parallels Income Growth: Some New Evidence," pp. 305-343 in B. Douglas Bernheim and John Shoven, eds, *National Saving and Economic Performance,* Chicago: Chicago University Press.

Farnsworth, E. Allan. (1990) *Contracts,* Boston: Little, Brown and Company.

Flavin, Marjorie. (1991) "The joint consumption/asset demand decision: a case study in robust estimation." NBER working paper no. 3802.

Goetz, Charles J. and Robert E. Scott. (1977) "Liquidated Damages, Penalties and the Just Compensation Principle: Some Notes on an Enforcement Model and a Theory of Efficient Breach." *Columbia Law Review,* 77, 554-594.

Goldman, Steven M. (1980) "Consistent Plans." *Review of Economic Studies,* 47, 533-37.

Hall, Robert E. and Frederic S. Mishkin. (1982) "The sensitivity of consumption to transitory income: estimates from panel data on households," *Econometrica,* 50, 461-481.

Hatsopoulos, George N., Paul R. Krugman, James M. Poterba (1989) "Overconsumption: The Challenge to U.S. Economic Policy," American Business Conference working paper.

Laibson, David I. (1993a) "Notes on a Precommitment Problem." MIT mimeo.

Peleg, Bezalel, and Menahem E. Yaari. (1973) "On the Existence of a Consistent Course of Action when Tastes are Changing," *Review of Economic Studies,* 40, 391-401.

Phelps, E. S., and R. A. Pollak. (1968) "On Second-best National Saving and Game-equilibrium Growth." *Review of Economic Studies,* 35, 185-199.

Pollak, R. A. (1968) "Consistent Planning." *Review of Economic Studies,* 35, 201-208.

Rea, Samuel A. Jr. (1984) "Efficiency Implications of Penalties and Liquidated Damages," *Journal of Legal Studies,* 13, 147-167.

Runkle, David E. (1991) "Liquidity constraints and the permanent-income hypothesis." *Journal of Monetary Economics,* 27, 73-98.

Shea, John. (1992) "Union Contracts and the Life Cycle-Permanent Income Hypothesis." Mimeo.

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

Thaler, Richard H. (1990) "Saving, Fungibility, and Mental Accounts." *Journal of Economic Perspectives,* 4:1, 193-205.

Zeldes, Stephen P. (1989) "Consumption and liquidity constraints: an empirical investigation." *Journal of Political Economy,* 97, 305-46.

Chapter 3

Mental Accounts, Self-Control, and an Intrapersonal Principal-Agent Problem

3.1 Introduction

When an individual's preferences are dynamically inconsistent she strictly prefers to constrain her own future choices (Strotz, 1956). Consider the following fanciful advice for a decision-maker who is looking for a way to achieve such commitment: Build a machine to follow you wherever you go, and program the machine to generate extremely aversive stimuli whenever you "misbehave." Let's call the machine a "binding automaton."

At first inspection the advice is absurd. But we do have access to social institutions which approximate the operations of the hypothesized machine. Parents, friends, supervisors, religious leaders, and spouses help us make commitments in ways which mimic the working of a binding automaton. I can promise my spouse to be home by eight. I can promise my workplace supervisor to finish a project in two weeks. Such communications implicitly generate a system of incentives which helps commit me to honor my promise.

In general, any organization/group in which I am a member has a seemingly

endless and often unspoken list of expectations/rules/norms to which I commit when I join the group. For example, my friends see me wearing a very luxurious new coat, and ask me about it. If I acknowledge that I spent $400 dollars on the coat they call me a spendthrift and effectively censure me. If I tell them that I found it on sale, they complement me on my good taste. Hence, membership in my social group is associated with strong incentives to not purchase luxurious clothing at retail prices.

This analysis suggests a theory of organizations and groups based on the commitment value of these institutions. Examples would include firms and schools (committing me to be productive), religions and marriage (committing me to be "virtuous"), twelve-step groups (committing me to be abstemious), etc. With these examples in mind, it may be reasonable to assume that individuals have access to social institutions which act effectively as binding automata. This is a strong assumption, but one which is worth considering as an important benchmark. The analysis which follows takes this assumption as a starting point and sees where it leads. Specifically, I analyze the behavior of an agent who has access to a binding automaton which enables her to perfectly commit to any contingent rule linking observable states to observable actions.

In a world where *all* states and actions are observable, the binding automaton assumption dramatically simplifies analysis. Actions are simply the optimal contingent rules from the perspective of the self that builds the binding automaton. However, in an economy where some actions and/or states are not observable the analysis is more complex, and that is the world which I will analyze. Specifically, I assume that the binding automaton can not directly observe effort.

The problem that I consider is an intra-personal principal-agent problem. The principal is the early self who builds the binding automaton. The agent is the later self who exerts effort in response to the incentive system created by the binding automaton. I assume that both the principal and the agent have a hyperbolic discount function.[1] This discount structure implies that the principal prefers relatively less consumption and relatively more effort during the agent's period of control than does

[1]See the introduction for an overview of hyperbolic discount functions.

the agent.

This intra-personal principal-agent problem can be compared and contrasted with the standard principal-agent problem from the industrial organization literature (*e.g.*, Shavell (1979), Holmström (1979), Grossman and Hart (1983)). In both problems effort is not observable, and the principal sets up an incentive system to motivate effort. However, in the intra-personal principal-agent model the utility of the principal and the agent are linked in a way which bears no resemblance to the utility relationship in the standard principal-agent model.

In the intra-personal principal-agent problem, the principal is forced to strike a balance between consumption smoothing, which the principal wants, and instantaneous gratification, which is used to motivate the agent to exert high effort. This fundamental tension explains many heretofore puzzling anomalies, including self-reward/self-punishment and some mental accounts. The principal, or early self, would like the later self to exert high effort *and* smooth consumption. But in order to extract high effort, the principal has to give the later self a strong incentive. The incentive takes the form of the following norm: if you (the later self) obtain a good-effort related outcome, then you can splurge. Hence, the equilibrium path is characterized by a high marginal propensity to consume out of labor income. This is self-reward. However, the early self does not need to reward the later self for good outcomes that are independent of effort. So when asset stocks are high, consumption responds modestly (according to the wishes of the early self). Hence on the equilibrium path, the marginal propensity to consume out of assets is low when compared to the marginal propensity to consume out of labor income. This pattern replicates some of the mental accounting behavior documented by Richard Thaler (1990).

The body of the chapter formalizes these claims. Section 2 lays out the model. Equilibrium outcomes are characterized in Section 3. A numerical example is illustrated in Section 4. Section 5 concludes with a critical evaluation and a discussion of ongoing work.

3.2 An Intra-personal Principal-Agent Model

The model has three critical components. First, the individual is assumed to have a hyperbolic discount function. This sets up an intra-personal conflict. Second, the individual has access to a costless binding automaton. This implies that the initial self can perfectly commit the observable actions of future selves. These commitments take the form of contingent rules, where the contingencies are based on observable states. Third, I assume that effort is not observable, but effort is positively correlated with good outcomes. The details of the model follow.

The individual makes decisions over three periods, $t \in \{1,2,3\}$. I adopt the semantic convention used in my previous chapters, and refer to self t as the self in control at time t. During period 1 the binding automaton is built by self 1 to maximize the welfare of self 1. In period 2, self 2 chooses effort, receives labor income (a stochastic function of effort), and consumes according to the instructions of the binding automaton. In period 3, self 3 consumes whatever assets are left.[2]

The individual is assumed to have a hyperbolic discount function. I capture the properties of the hyperbolic discount function by assuming that the one period discount factor is $\beta\delta$, $(0 < \beta < 1)$, and assuming that the two-period discount factor is $\beta\delta^2$. This implies that the discount rate is falling. The utility functions of the three selves are given by,

$$\begin{aligned}
\text{Self 1:} \quad & u_1 + \beta\delta[u(c_2) - e] + \beta\delta^2 u(c_3) \\
\text{Self 2:} \quad & [u(c_2) - e] + \beta\delta u(c_3) \\
\text{Self 3:} \quad & u(c_3)
\end{aligned}$$

where u_1 is a constant, $u' > 0$, $u'' < 0$. The production technologies in the economy are very simple. Effort (chosen in period 2) takes one of two values, $e \in \{\underline{e}, \overline{e}\}$, $\underline{e} < \overline{e}$. Income (realized in period 1) is also drawn from a two-point set: $y \in \{y_L, y_H\}$, $y_L <$

[2]It is possible to also let self 3 consume according to rules of the binding automaton, but those rules will generally dictate that self 3 consume all remaining assets anyway. The only exception to this pattern arises when the optimal contingent rules in period 2 constitute a border solution. In that case it may be *ex ante* optimal to instruct self 3 to consume less than the remaining asset stock contingent on a low income realization in period 2.

y_H. The distribution of income is a function of the chosen effort level. Specifically,

$$Prob(y_H|\overline{e}) = \overline{p} > \underline{p} = Prob(y_H|\underline{e}),$$

which implies that $Prob(y_L|\overline{e}) = 1 - \overline{p}$, and $Prob(y_L|\underline{e}) = 1 - \underline{p}$. Finally, the gross interest rate, R characterizes the storage technology. For simplicity, I set $R = 1$.

We are now in a position to discuss explicitly the construction of the binding automaton. The only observable state in period 2 will be the realized income level, y. So the consumption rule for period 2 can only be a function of y. Since y can only take on two values, it is possible to represent the consumption rule as an ordered pair: $\{c_L, c_H\}$, where c_L is the consumption level when y_L is realized, and c_H is the consumption level when y_H is realized.

3.3 Equilibrium

Self 2's problem must be solved first. Let, $W(c_L, c_H) \equiv$

$$(\overline{p} - \underline{p})\left[u(c_H) + \beta\delta u(y_H - c_H) - u(c_L) - \beta\delta u(y_L - c_L)\right] - (\overline{e} - \underline{e}),$$

which is the difference between self 2's payoff given $e = \overline{e}$ and self 2's payoff given $e = \underline{e}$. Self 2 must select an effort level e^*. She sets $e^* = \overline{e}$ if $W > 0$, $e^* = \underline{e}$ if $W < 0$, and is assumed to choose whatever effort level is preferred by self 1 if $W = 0$.

Self 1's decision problem is less simple. Let $U(c_L, c_H|\overline{e}) \equiv$

$$\overline{p}\left[u(c_H) + \beta\delta u(y_H - c_H)\right] + (1 - \overline{p})\left[u(c_L) + \beta\delta u(y_L - c_L)\right] - \overline{e},$$

and let $U(c_L, c_H|\underline{e})$ represent the same object with $\overline{p}$ replaced by $\underline{p}$ and $\overline{e}$ replaced by $\underline{e}$. Hence, $U(c_L, c_H|\overline{e})$ is the payoff to self 1, given $e = \overline{e}$, and $U(c_L, c_H|\underline{e})$ is the payoff to self 1, given $e = \underline{e}$.

Self 1's decision process requires self 1 to solve two subproblems and compare the solutions.

$$\text{I.} \quad \max_{\{c_L, c_H\}} \; U(c_L, c_H|\overline{e})$$
$$\text{s.t.} \quad W(c_L, c_H) \geq 0$$

$$\text{II.} \quad \max_{\{c_L, c_H\}} \; U(c_L, c_H|\underline{e})$$
$$\text{s.t.} \quad W(c_L, c_H) \leq 0$$

Let C^I (C^{II}) represent the set of ordered pairs which are the solutions to program I (II). Let C^* represent the set of ordered pairs which are the solutions to self 1's global problem.

$$C^* = \begin{cases} C^I & \text{if } U(C^I|\overline{e}) \geq U(C^{II}|\underline{e}) \\ C^{II} & \text{otherwise} \end{cases}$$

3.3.1 Self 1's first-best solution

The first goal of this section is to simplify self 1's decision problem. To do this it is helpful to discuss a benchmark case: the first-best solution from the perspective of self 1. Consider the problem in which self 1 chooses the ordered pair $\{c_L, c_H\}$ and also directly chooses the effort level, e. It is trivial to show that there exists a unique ordered pair, $\hat{C} = \{\hat{c}_L, \hat{c}_H\}$ which solves this first-best problem. This ordered pair is determined by the first-order conditions:

$$u'(\hat{c}_H) = \delta u'(y_H - \hat{c}_H),$$

$$u'(\hat{c}_L) = \delta u'(y_L - \hat{c}_L).$$

The effort level, $\hat{e}$, which solves self 1's first-best problem is given by,

$$\hat{e} = \begin{cases} \overline{e} & \text{if } U(\hat{C}|\overline{e}) \geq U(\hat{C}|\underline{e}) \\ \underline{e} & \text{otherwise} \end{cases}$$

The following lemma will be used in several of the results which follow.

Lemma 1: $y_H - \hat{c}_H > y_L - \hat{c}_L$.

Proof: Suppose $y_H - \hat{c}_H \leq y_L - \hat{c}_L$, and look for a contradiction. By FOC's of first-best problem,

$$u'(\hat{c}_H) = \delta u'(y_H - \hat{c}_H) \geq \delta u'(y_L - \hat{c}_L) = u'(\hat{c}_L),$$

where the inequality follows from the assumption, $y_H - \hat{c}_H \leq y_L - \hat{c}_L$. The inequality $u'(\hat{c}_H) \geq u'(\hat{c}_L)$ implies $\hat{c}_H \leq \hat{c}_L$. So,

$$\hat{c}_H + (y_H - \hat{c}_H) \leq \hat{c}_L + (y_L - \hat{c}_L),$$

implying, $y_H \leq y_L$, which is a contradiction. □

3.3.2 Simplifying self 1's problem

We are now ready to simplify self 1's decision rule. In particular, it is possible to solve self 1's problem without solving program II. The following lemma is used to prove this result.

Lemma 2: *If* $U(C^I|\overline{e}) < U(C^{II}|\underline{e})$, *then* $C^* = \hat{C}$.

Proof: Suppose $U(C^I|\overline{e}) < U(C^{II}|\underline{e})$, and $C^* \neq \hat{C}$, and look for contradiction. Recall that in general C^* is a solution *set.* It is easy to show that $C^* \neq \hat{C}$ implies that $C^* \not\ni \hat{C}$. Continuing with the proof, note that $W(\hat{C})$ must be greater than zero, (since $C^* = C^{II}$, C^{II} is the solution set of program II, $C^* \not\ni \hat{C}$, and $U(\hat{C}|\underline{e}) \geq U(C^*|\underline{e})$). $W(\hat{C}) > 0$ implies that,

$$U(\hat{C}|\overline{e}) \leq U(C^I|\overline{e}) \leq U(C^*|\underline{e}) \leq U(\hat{C}|\underline{e}).$$

Combining, $W(\hat{C}) > 0$ with $U(\hat{C}|\overline{e}) \leq U(\hat{C}|\underline{e})$ yields,

$$u(y_L - c_L) \geq u(y_H - c_H),$$

which contradicts lemma 1. □

The following proposition shows that self 1's optimal decision can be calculated without solving program II.

Proposition 1: *Let*

$$C^{**} = \begin{cases} C^I & \text{if } U(C^I|\overline{e}) \geq U(\hat{C}|\underline{e}) \\ \hat{C} & \text{otherwise} \end{cases}$$

Then $C^{**} = C^*$.

Proof: First, suppose $U(C^I|\overline{e}) < U(C^{II}|\underline{e})$. Then $C^* = C^{II}$, and by Lemma 2, $C^{II} = \hat{C}$. Hence, for this case the proposition is true. Now WLOG, assume, $U(C^I|\overline{e}) \geq U(C^{II}|\underline{e})$. If $C^{II} = \hat{C}$, the proposition is true. So WLOG, assume, $C^{II} \neq \hat{C}$. This implies, $W(\hat{C}) > 0$, which implies that $C^I = \hat{C}$. Hence,

$$U(C^I|\overline{e}) = U(\hat{C}|\overline{e}) > U(\hat{C}|\underline{e}),$$

where the last inequality is derived by combining Lemma 1 with $W(\hat{C}) > 0$. Hence, $C^{**} = C^I$, completing the proof. □

3.3.3 Characterizing the solution

This subsection characterizes the set of solutions to self 1's problem. This characterization is useful for two reasons. First, it helps to develop intuition about the solutions. Second, it enables me to prove that second order conditions are satisfied. Program 1 is not concave over the entire solution space. However, the solutions of the problem can be shown to exist in a subspace in which sufficient conditions are

satisfied. I will return to these issues in the next subsection. I use the following notation in the claims below:

$$U_H \equiv \frac{\partial U}{\partial c_H}, \quad U_L \equiv \frac{\partial U}{\partial c_L}, \quad W_H \equiv \frac{\partial W}{\partial c_H}, \quad W_L \equiv \frac{\partial W}{\partial c_L}.$$

Proposition 2: *In equilibrium* $u'(c_H) \geq \beta\delta u'(y_H - c_H)$.

Proof: If $C^* = \hat{C}$, the proposition follows immediately, (since $\beta < 1$). So, WLOG, assume $C^* \neq \hat{C}$. This implies (by Proposition 1) that $C^* = C^I$. The Kuhn-Tucker necessary conditions associated with program I are given below:

$$U_H + \lambda W_H = 0,$$

$$U_L + \lambda W_L = 0,$$

$$\lambda \geq 0, \quad W \geq 0, \quad \lambda W = 0.$$

Note that the proposition is equivalent to the claim that in equilibrium $W_H \geq 0$. Assume that in equilibrium $W_H < 0$, and look for a contradiction. Note that $U_H = W_H - (1-\beta)\delta u'(y_H - c_H)$, so $W_H < 0$ implies $U_H < 0$, which implies $\lambda < 0$, which contradicts the necessary conditions. □

Proposition 3: *In equilibrium* $u'(c_H) \leq \delta u'(y_H - c_H)$.

Proof: Using the same argument as above, assume WLOG $C^* \neq \hat{C}$, and hence, $C^* = C^I$. Recall the Kuhn-Tucker conditions associated with program I. Note that $W_H \geq 0$ (by Proposition 2), and $\lambda \geq 0$ implies $U_H \leq 0$, which completes the proof. □

Lemma 3: *In equilibrium,* $u'(c_L) \geq \delta u'(y_L - c_L)$ *or* $u'(c_L) < \beta\delta u'(y_L - c_L)$.

Proof: Using the same argument as above, assume WLOG $C^* \neq \hat{C}$, and hence, $C^* = C^I$. Recall the Kuhn-Tucker conditions associated with program I. Note that $u'(c_L) < \delta u'(y_L - c_L)$ implies $U_L < 0$, which implies, $W_L > 0$, (since $\lambda \geq 0$). □

Lemma 4: *In equilibrium,* $u'(c_L) \geq \beta\delta u'(y_L - c_L)$.

Proof: Using the same argument as above, assume WLOG $C^* \neq \hat{C}$, and hence, $C^* = C^I$. Fix any equilibrium c_L. Suppose, $u'(c_L) < \beta\delta u'(y_L - c_L)$, and look for a contradiction. Note that $u(x) + \beta\delta u(y_L - x)$ is concave in x with a maximum at x : $u'(x) = \beta\delta u'(y_L - x)$. Note that $u(0) + \beta\delta u(y_L) \leq u(y_L) + \beta\delta u(0)$. So there exists a $\tilde{c}_L < c_L$ s.t.

$$u(\tilde{c}_L) + \beta\delta u(y_L - \tilde{c}_L) = u(c_L) + \beta\delta u(y_L - c_L).$$

Note that $u(y_L - \tilde{c}_L) > u(y_L - c_L)$. Combining this observation with the previous line yields,

$$u(\tilde{c}_L) + \delta u(y_L - \tilde{c}_L) > u(c_L) + \beta\delta u(y_L - c_L),$$

which implies that self 1 is made strictly better off by switching from c_L to $\tilde{c}_L$, which violates the original equilibrium assumption. □

Proposition 4: *In equilibrium* $u'(c_L) \geq \delta u'(y_H - c_L)$.

Proof: The proposition follows from Lemma 3 and Lemma 4. □

Proposition 5: *In equilibrium,*

$$\frac{c_H - c_L}{y_H - y_L} \leq 1.$$

Proof: If $C^* = \hat{C}$ then the Proposition follows from Lemma 1. WLOG assume $C^* = C^I \neq \hat{C}$. This implies that $W(C^*) = 0$. Note that any perturbation of C^* must make self 1 no better off. Consider perturbations to C^* which lead self 2 to choose $\underline{e}$ instead of $\overline{e}$. Such perturbations are possible since $W(C^*) = 0$. Optimality of C^* requires that,

$$U(C^*|\overline{e}) - U(C^*|\underline{e}) \geq 0.$$

Subtracting $W(C^*) = 0$ from the LHS of this expression, yields,

$$\delta(1-\beta)(\overline{p} - \underline{p})\left[u(y_H - c_H) - u(y_L - c_L)\right] \geq 0.$$

This implies that $(y_H - c_H) \geq (y_L - c_L)$ which completes the proof. □

3.3.4 Sufficient Conditions

It is now possible to derive a sufficiency theorem.

Proposition 6: *There exists at most one solution to program I. If a solution exists, it is in the region described by Propositions 2-4, and it is the only point in this region which satisfies the Kuhn-Tucker conditions of program I.*

Proof: The proposition is trivial to confirm if $W(\hat{C}) \geq 0$. WLOG assume $W(\hat{C}) < 0$. Then at any solution, $W = 0$. So the solution set of program I (under these assumptions) is equivalent to the solution set of the following program, (in which the constraint binds).

$$\text{IB.} \quad \max_{\{c_L, c_H\}} \quad U(c_L, c_H|\overline{e})$$
$$\text{s.t.} \quad W(c_L, c_H) = 0$$

Define a subspace S of the non-negative orthant of $\Re^3$, such that elements of S are ordered triplets, $\{c_L, c_H, \lambda\}$ which satisfy the properties,

$$\delta u'(y_L - c_L) \leq u'(c_L),$$

$$\beta\delta u'(y_H - c_H) \leq u'(c_H) \leq \delta u'(y_H - c_H),$$

$$\lambda \leq \frac{1-\overline{p}}{\overline{p}-\underline{p}}.$$

Note that by Propositions 2-4 and the Kuhn-Tucker conditions of program I, any solution to program I must be an element of S. Moreover, since program I and program IB have the same solution set and the same Kuhn-Tucker conditions, all solutions of program IB must also be in S.

The next step of the proof is to show that the bordered Hessian associated with program IB has a positive determinant in S. Let,

$$U_{HH} \equiv \frac{\partial^2 U}{\partial c_H^2}, \quad U_{LL} \equiv \frac{\partial^2 U}{\partial c_L^2}, \quad U_{HL} \equiv \frac{\partial^2 U}{\partial c_H \partial c_L}.$$

and represent the second derivatives of W in an analogous way. Let λ be the Lagrange multiplier associated with the constraint in program IB. Then the bordered Hessian of program IB is,

$$H \equiv \begin{bmatrix} 0 & W_H & W_L \\ W_H & U_{HH} + \lambda W_{HH} & U_{HL} + \lambda W_{HL} \\ W_L & U_{HL} + \lambda W_{HL} & U_{LL} + \lambda W_{LL} \end{bmatrix}$$

Note that $U_{LH} = 0$, and $W_{HL} = 0$. So the determinant of the bordered Hessian is,

$$|H| = -\left[W_H^2(U_{LL} + \lambda W_{LL}) + W_L^2(U_{HH} + \lambda W_{HH})\right].$$

Note that $(U_{HH} + \lambda W_{HH}) < 0$ in S, (since $U_{HH} < 0$, $\lambda \geq 0$, and $W_{HH} < 0$ in S). Hence, to show $|H| > 0$, it is sufficient to show $U_{LL} + \lambda W_{LL} =$

$$(1 - \overline{p})\,[u''(c_L) + \delta u''(y_L - c_L)] - \lambda(\overline{p} - \underline{p})\,[u''(c_L) + \beta\delta u''(y_L - c_L)] < 0,$$

in S. This inequality follows from the properties, $\beta < 1$, and $\lambda \leq \frac{1-\overline{p}}{\overline{p}-\underline{p}}$.

Hence, the determinant of the bordered Hessian is positive in S, so there exists a unique point in S which satisfies the Kuhn-Tucker conditions of program IB. (Note, there must exist at least one point in S which satisfies the Kuhn-Tucker conditions, since S contains all solutions of program IB.) Since all solutions of program IB satisfy the Kuhn-Tucker conditions, and a unique point in S satisfies the Kuhn-Tucker conditions, and all solutions of program IB are in S, there exists a unique solution of program IB. Hence, program I must also have a unique solution.

Note that the solution to program IB satisfies the Kuhn-Tucker conditions of program I. Now it only remains to show that the Kuhn-Tucker conditions of program I admit no other solutions in S. Let C be the unique maximum of the two programs. Let C' be any point in S which satisfies the Kuhn-Tucker conditions of program I. If $W(C') = 0$ then C' also satisfies the Kuhn-Tucker conditions of program IB, implying that $C' = C$ (since I have shown that C is the only point in S which satisfies the Kuhn-Tucker conditions of program IB). So WLOG assume that $W(C') > 0$. Then $\lambda = 0$, and $C' = \hat{C}$, contradicting the assumption $W(\hat{C}) < 0$. □

3.3.5 Comparative Statics

This model suggests a natural measure of the marginal propensity to consume out of labor income:

$$MPC^y = \frac{c_H - c_L}{y_H - y_L}.$$

Note that β measures the degree of congruence between the interests of self 1 and self 2. When $\beta = 1$ those interests are perfectly aligned, and the principal-agent problem collapses to the first-best problem of self 1. As β goes to zero the interests

of self 1 and self 2 are maximally unaligned. At the limit, self 2 desires complete instantaneous gratification.

Proposition 7: *Let $C(\beta)$ be the (unique) solution to self 1's problem, given a particular value of β. If $C(\beta) \neq \hat{C}$,*

$$\frac{\partial MPC^y(\beta)}{\partial \beta} < 0.$$

Proof: $C(\beta) \neq \hat{C}$ implies that $W(C(\beta)) = 0$, so the inequality constraint in program I binds. Hence, in equilibrium the following equations hold (derived from the Kuhn-Tucker conditions): $U_H W_L = U_L W_H$, and $W = 0$. Applying the implicit differentiation theorem and eliminating zero terms yields,

$$\frac{\partial c_H}{\partial \beta} = \frac{W_L[U_H W_{L\beta} - U_L W_{H\beta}] - W_\beta[U_H W_{LL} - U_{LL} W_H]}{|H|},$$

$$\frac{\partial c_L}{\partial \beta} = \frac{-W_H[U_H W_{L\beta} - U_L W_{H\beta}] + W_\beta[U_{HH} W_L - U_L W_{HH}]}{|H|}.$$

Recall (from previous proof) that $|H|$ is the determinant of the bordered Hessian. Combining the previous two equations yields, $\frac{\partial(c_H - c_L)}{\partial \beta} =$

$$\frac{(W_L + W_H)[U_H W_{L\beta} - U_L W_{H\beta}] - W_\beta[U_H W_{LL} - U_{LL} W_H + U_{HH} W_L - U_L W_{HH}]}{|H|}$$

I've already shown (see previous proof) that $|H|$ is positive at all optima. Hence, it is sufficient to sign the numerator of this expression. I proceed term by term.

Note that,

$$\begin{aligned}
\frac{-W_L}{W_H} &\geq \frac{-W_L}{W_H} \cdot \frac{u'(y_H - c_H)}{u'(y_L - c_L)} \\
&= \frac{\frac{u'(c_L) - \beta\delta u'(y_L - c_L)}{u'(y_L - c_L)}}{\frac{u'(c_H) - \beta\delta u'(y_H - c_H)}{u'(y_H - c_H)}} \\
&= \frac{\frac{u'(c_L)}{u'(y_L - c_L)} - \beta\delta}{\frac{u'(c_H)}{u'(y_H - c_H)} - \beta\delta} \\
&> 1
\end{aligned}$$

where the first inequality follows from Proposition 5, and the last inequality follows from Proposition 3 and Proposition 4. Note that $W_H > 0$ by Proposition 2. Multiplying the last line by W_H yields,

$$W_H + W_L < 0.$$

Note that,

$$\begin{aligned}
\frac{-U_L W_{H\beta}}{U_H W_{L\beta}} &= \frac{-W_L W_{H\beta}}{W_H W_{L\beta}} \\
&= \frac{-\frac{u'(c_L) - \beta\delta u'(y_L - c_L)}{u'(y_L - c_L)}}{\frac{u'(c_H) - \beta\delta u'(y_H - c_H)}{u'(y_H - c_H)}} \\
&< -1
\end{aligned}$$

The first inequality follows from substitution of the Kuhn-Tucker conditions, and the last inequality follows from the arguments made in the previous derivation. Note that

$U_H W_{L\beta} < 0$ by Proposition 3. Multiplying the last line by $U_H W_{L\beta}$ yields,

$$U_H W_{L\beta} - U_L W_{H\beta} > 0.$$

Note that,

$$\begin{aligned}
\frac{U_H W_{LL}}{U_{LL} W_H} &= \frac{-U_L W_{LL}}{W_L U_{LL}} \\
&= \frac{u'(c_L) - \delta u'(y_L - c_L)}{u'(c_L) - \beta\delta u'(y_L - c_L)} \cdot \frac{u''(c_L) + \beta\delta u''(y_L - c_L)}{u''(c_L) + \delta u''(y_L - c_L)} \\
&> 1
\end{aligned}$$

The first inequality follows from substitution of the Kuhn-Tucker conditions, and the last inequality follows from $\beta < 1$, and Proposition 4. Note that $U_{LL} W_H < 0$, by Proposition 2. Multiplying the last line by $U_{LL} W_H$ yields,

$$U_H W_{LL} - U_{LL} W_H > 0.$$

The remaining terms can be signed directly. Note that $W_\beta > 0$ by Proposition 5, $U_L > 0$ by Proposition 4, $W_L < 0$ by Proposition 4, and $U_{HH}, W_{HH} < 0$. So,

$$U_{HH} W_L - U_L W_{HH} > 0.$$

Together these observations imply that $MPC^y(\beta)$ is falling in β. □

This long proof establishes a very simple result. As self 2's interests move closer to self 1's interests, self 1 needs to sanction less instantaneous gratification to motivate self 2. Put differently, self-reward is used less and less for self-motivation as β goes to unity.

3.3.6 Mental Accounts

I am now in a position to return to the discussion of mental accounts. The goal of this subsection is to propose a meaningful way of comparing the MPC out of labor income—MPC^y—with the MPC out of assets, henceforth represented as MPC^A. The first order of business is to define MPC^A.

To simplify analysis I will propose a definition of MPC^A which is independent of the effort incentive problem. Modify the earlier intra-personal principal-agent problem by setting $\overline{p} = \underline{p}$. This modified problem is a limiting case of the original problem. Note that the modified problem has no meaningful effort decision. Income in the modified problem is uncorrelated with the effort decision. So the agent trivially chooses low effort. In this setting income can be interpreted as asset windfalls (*e.g.* capital gains). Define MPC^A to be the MPC which arises in the modified problem. Specifically,

$$MPC^A = \frac{c_H - c_L}{y_H - y_L},$$

where c_H and c_L are the equilibrium contingent consumption levels associated with the modified problem.[3]

Proposition 8: *Let MPC^A be the equilibrium MPC of the limiting principal-agent problem with $\overline{p} = \underline{p}$. Let MPC^y_{FB} be the MPC associated with the first-best solution of the original principal-agent problem. Let $MPC^y(1)$ be the equilibrium MPC of the original principal-agent problem with $\beta = 1$. Let $MPC^y(\beta)$ be the equilibrium MPC of the original principal-agent problem with $\beta < 1$. Then, if C^* is the equilibrium in the original principal-agent problem, and $C^* \neq \hat{C}$,*

$$MPC^A = MPC^y_{FB} = MPC^y(1) < MPC^y(\beta).$$

[3]Other sensible definitions are possible. In particular, I originally worked with the definition

$$MPC^A = \omega \frac{\partial c_H}{\partial \Delta} + (1 - \omega) \frac{\partial c_L}{\partial \Delta},$$

where ω is a weighting function (*e.g.* $\omega = \overline{p}$, or $\omega = \frac{1}{2}$), and Δ represents an income component which is in both y_H and y_L. This approach generated the same qualitative results as the one pursued in the chapter, but with far less clarity and simplicity.

Proof: It is straightforward to confirm that the agency problems associated with MPC^A, MPC^y_{FB}, and $MPC^y(1)$, all have solution $\{c_H, c_L\} = \hat{C}$. The first two inequalities follow immediately from this observation. The last inequality is implied by Propositions 3 and 4, and $C^* \neq \hat{C}$. □.

3.4 An illustration

I have shown that the MPC out of labor income (*i.e.* income which is positively correlated with effort) is higher than the MPC out of asset income (*i.e.* income which is uncorrelated with effort). But I haven't discussed the magnitude of this difference. The following example provides an arbitrary illustration of the magnitude of these effects. The example is calibrated by setting $u(\cdot) = \ln(\cdot)$, $\delta = 1$, $y_H = 6$, $y_L = 5$, $\overline{p} = \frac{2}{3}$, $\underline{p} = \frac{1}{3}$, and $\overline{e} - \underline{e} = .116$. All of these parameters and specifications were chosen independently, except $\overline{e} - \underline{e}$, which was chosen so that there would be a range of β values over which self 1 would induce self 2 to select the high effort level.

Figure 1 graphs c_H and c_L as functions of β. The unit interval of β values can be broken down into three subintervals of interest: $\beta \in [0, .1002]$, $\beta \in [.1002, .9087]$, and $\beta \in [.9087, 1]$. I will refer to these as intervals A, B, and C. In interval A, β is close to zero, and the interests of self 1 and self 2 are highly divergent. Self 1 would like self 2 to exert high effort, but setting up an incentive scheme to induce self 2 to do so is too costly from the perspective of self 1. So self 1 choose $C^* = \hat{C}$ and self 2 chooses $e^* = \underline{e}$. In interval B, β takes on intermediate values. Now, from self 1's perspective it is desirable to set up an incentive scheme to induce self 2 to exert high effort. Because self 2's interests are still sufficiently divergent from self 1's interest, self 2 must be rewarded to motivate a high effort choice. Hence in interval B, $c^*_H > \hat{c}_H$, and $c^*_L < \hat{c}_L$. In interval C, β is sufficiently close to one that self 1 does not need to create any (extra) incentive to motivate self 2 to choose $\overline{e}$. So in this region, $C^* = \hat{C}$, and $e^* = \overline{e}$.

Figure 2 graphs, $MPC^y(\beta) = \frac{c_H(\beta) - c_L(\beta)}{y_H - y_L}$. MPC^y should be contrasted with MPC^A. The latter is equal to .5 for all β values. Finally, note that the drop in

Figure 1: cH and cL as a function of Beta

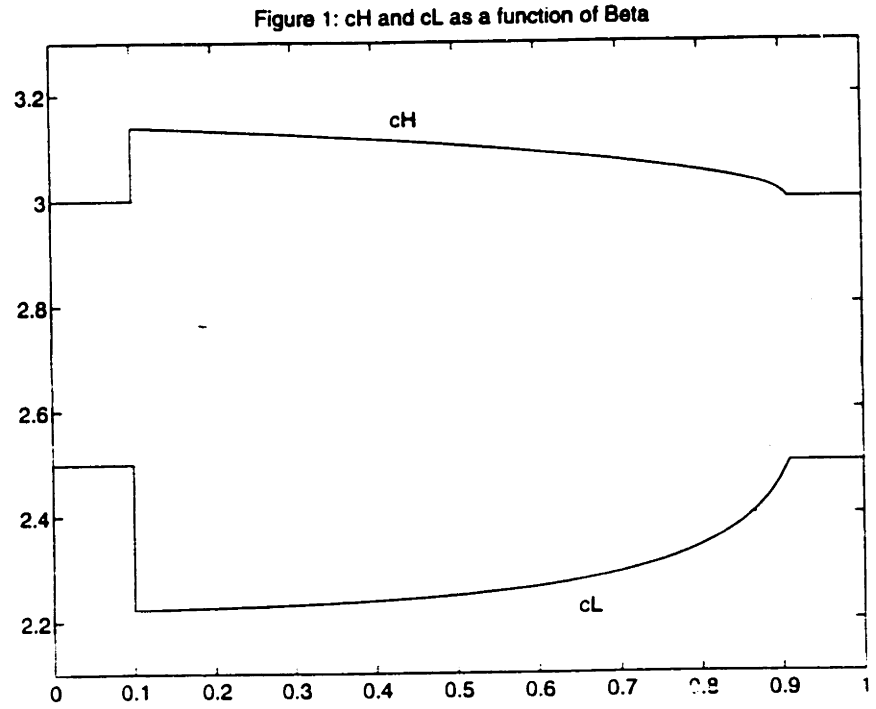
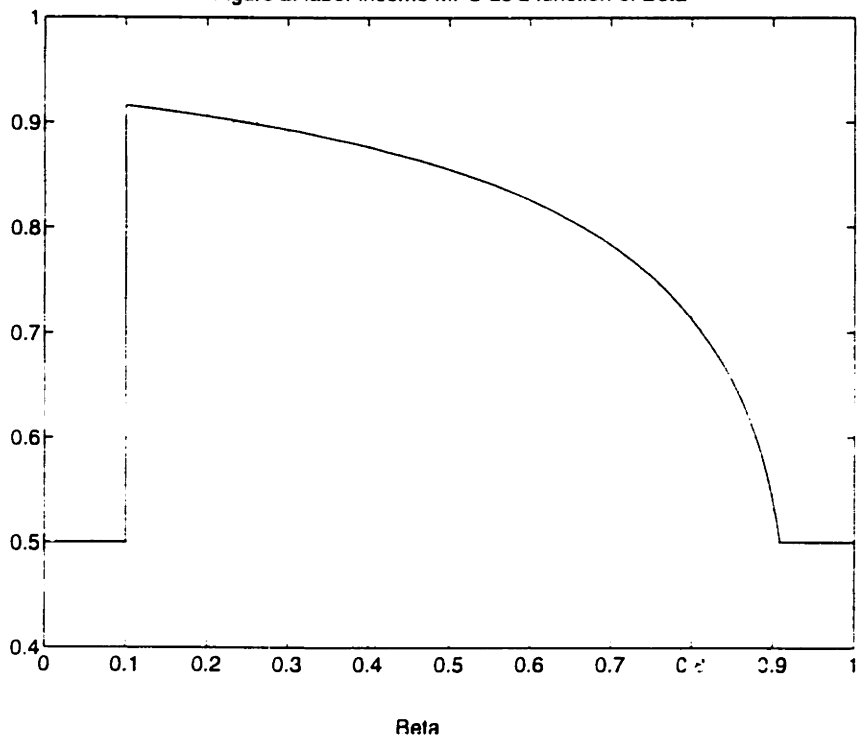


Figure 2: labor income MPC as a function of Beta



MPC^y at low β values is a consequence of the discrete effort assumption. I conjecture that a model with continuous effort choice will generate an equilibrium MPC^y which rises everywhere as β falls. Preliminary work with such a model supports this observation.

3.5 Evaluation

The intrapersonal principal agent problem in these notes explains mental accounts in a novel way. In my story, mental accounts represent a sophisticated tradeoff between the desire for consumption smoothing and the need to motivate effort with instantaneous gratification/punishment. I acknowledge that not all mental account phenomena fit naturally into this paradigm. In particular, the most problematic mental account phenomenon for my intra-personal agency model is the high measured MPC out of exogenous, liquid wealth windfalls (see Thaler, 1990). Such behavior can be shoehorned into my intra-personal agency model in either of two ways. First, one can assume that no liquid wealth shocks are truly independent of effort, or at least that if such exogenous shocks do exist, they are sufficiently rare or difficult to identify that we do not even bother creating norms to handle them. Second, one can assume that the high measured MPC out of liquid wealth shocks reflects an incentive scheme that was historically highly successful but has ceased to be useful due to a weakening of the connection between recent effort and liquid wealth. There may have been a time when all liquid wealth shocks were effort-related—*e.g.*, in a hunter-gatherer society—and that is when these norms evolved.

While I find both of these stories intriguing, I believe that my intra-personal agency model does not satisfactorily explain the high measured MPC out of exogenous liquid wealth windfalls. However, any model with liquidity constraints (endogenous or exogenous), can explain the liquid wealth anomaly (*e.g.* see the second chapter of this thesis). Hence, I am comfortable separating the liquid wealth anomaly from the body of other mental accounting phenomena that my intra-personal agency model does readily explain: a high MPC out of income/wealth that is effort related (*e.g.*

labor income) and a relatively low MPC out of income/wealth that is independent of effort (*e.g.* capital gains). For example, my agency model explains why passive investors have a lower MPC out of capital gains than active investors. The agency model explains why students reward themselves when they get back a successful exam. The agency model explains why shoppers reward themselves for finding a needed item on sale (by splurging the "saved" money on something frivolous). The agency model explains why many people punish themselves when something "bad" happens, like losing an airplane ticket or getting fined for speeding.

Future work on my intra-personal agency model of mental accounts will focus in two areas. First, I hope to extend the model to a multi-period framework. Such an extension will make the model more closely map observed mental account phenomena by increasing the gap between the MPC out of effort-independent wealth and the MPC out of effort-related wealth. Second, I hope to weaken my assumption about the effectiveness of the binding automaton. I have assumed that the binding automaton can be used to perfectly commit the agent to *any* contingent rule linking observable actions to observable states. Weakening this assumption is analogous to weakening the complete contracts assumption in the standard principal-agent literature. Preliminary work suggests that such a weakening will pave the way for a rich theory of organizations and norms based on the value of commitment.

3.6 References

Ainslie, George W. (1992) *Picoeconomics,* Cambridge: Cambridge University Press.

Grossman, Sanford, and Oliver Hart (1983) "An Analysis of the Principal-Agent Problem." *Econometrica,* 51, 7-45.

Holmström, B. (1979) "Moral Hazard and Observability." *Bell Journal of Economics,* 10, 74-91.

Shavell, Steve (1979) "Risk Sharing and Incentives in the Principal and Agent Relationship." *Bell Journal of Economics,* 10, 55-73.

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

Thaler, Richard H. (1990) "Saving, Fungibility, and Mental Accounts." *Journal of Economic Perspectives,* 4:1, 193-205.