Commitment and Self-Control*

Jawwad Noor

Department of Economics,
Harkness Hall,
University of Rochester,
Rochester, NY 14627.
Email: jwdb@troi.cc.rochester.edu.

April 24, 2005

*I am greatly indebted to Larry Epstein for his guidance and many suggestions. I have also benefitted from discussions with John Duggan, Faruk Gul, John Leahy, Mark Machina, Jianjun Miao and participants at the Canadian Economic Theory Conference 2004. Any errors are my responsibility.

Abstract

The literature on self-control problems has typically concentrated on immediate temptations. This paper studies a Gul and Pesendorfer [13, 14] style model in which decision-makers are affected by temptations that lie in the future. While temptation is commonly understood to give rise to a demand for commitment, it is shown that 'temptation by future consumption' can induce its absence. The model also exhibits procrastination, provides an alternative to projection bias as an explanation for some experimental results, and can simultaneously account for myopic and hyperopic behavior. The evidence on preference reversals supports temptation by future consumption, and suggests that it may not be restricted to short time horizons.

Keywords: Self-Control, Temptation, Commitment, Preference Reversals, Procrastination.

JEL classification number: D11

1. Introduction

The phenomenon of preference reversals is well documented in the psychology literature (see Ainslie [1] for references). Preference reversals occur when subjects prefer, say, \$30 in two months to \$20 in one month, but prefer \$20 now to \$30 in one month. That is, they reverse preferences in favor of the smaller/earlier reward when it is close to the present. This suggests that subjects are tempted by opportunities of immediate gratification. This, in turn, suggests the existence of self-control problems, or the inability to fully resist temptations.

Models that incorporate self-control problems have been developed and studied by Strotz [33], Ainslie [1], Laibson [21], O'Donoghue and Rabin [27] and Gul and Pesendorfer [13, 14]. An implication of these models is that decision-makers who are aware of their self-control problems seek commitment opportunities. To see this, consider a smoker who is deciding whether or not to commit to quitting. Denoting smoking by s and not smoking by n, his problem is to choose which of the choice sets

$$\{n\} \text{ or } \{n, s\}$$

to face tomorrow. In order to avoid the temptation of s in $\{n, s\}$, he chooses $\{n\}$. That is, anticipation of future temptation leads to a preference for commitment.

Given the presumption that self-control problems are common, the models predict a 'high' value for commitment. For instance, Laibson, Repetto and Tobacman [22] estimate that the value of perfect financial commitment mechanisms is worth 36% of consumption at age twenty. However, several researchers (for instance, Gale [4], Kocherlakota [18]) have argued that commitment may not be so highly valued in the market. Their observations are collected below:

- We observe compulsive behavior. If compulsive behavior is the result of self-control problems, we should observe addicts taking advantage of commitment opportunities.[1] Yet, in practice, there is a significant problem of non-compliance with commitment-based treatment procedures among addicts (Ainslie [1], Fuller and Roth [11], Goldstein [12]). For instance, disulfiram-based treatments for alcoholics and naltexrone-based treatments for heroin and morphine addicts are known to be of limited effectiveness, primarily because patients do not comply with the

[1] Examples of commitment opportunities include treatment involving disulfiram for alcoholics. Disulfiram leads to a reaction to ingestion of even small quantities of alcohol.

disulfirum or naltexrone regimen despite exhibiting a desire for the treatment. Fuller and Roth [11] state that "willingness to take [disulfirum] may not in itself be sufficient to achieve abstinence - receiving the drug is probably necessary". Thus, agents with self-control problems do not necessarily demand commitment, in contrast with the prediction of the above models.

- Self-control problems have been put forward as an explanation for the apparent undersaving in the U.S. The earlier noted models imply that undersavers do not need added incentives to participate in saving schemes such as 401(k) and IRAs which provide a means to commit to saving for retirement. Yet such saving schemes have substantial tax benefits associated with them: all contributions are tax deductible. Furthermore, participation in these schemes is closely related to the tax benefits. For instance, IRA contributions fell by 62% when the Tax Reform Act of 1986 excluded higher-income groups from tax benefits (Venti and Wise [35], Poterba, Venti and Wise [30]).[2] The fall in participation took place although there was no change in the commitment aspect of IRAs (early withdrawal penalties). This suggests that the appeal of such saving vehicles is primarily the tax benefits, not their commitment value (Gale [4]).

- Since agents who value commitment are willing to pay for it, commitment assets should cost more than the present value of their returns. However, IRAs, Christmas Clubs, etc. offer competitive rates of return comparable with those available in the market (Kocherlakota [18]).

- It is hard to find examples of perfect commitment devices. For instance, 401(k)s and illiquid assets can be used as collateral for borrowing. More to the point, if agents value perfect commitment then firms and workers would write contracts that commit workers to save. Such contracts are not hard to create and yet are rarely observed (Gale [4]).

These observations cast doubt on the claim that there is a significant demand for commitment. This may be regarded as reason not to take self-control problems seriously. However, it is difficult to ignore the evidence on preference reversals and the normative implications of models that incorporate temptation.[3] Hence, it is worth asking the question: *is an insignificant demand for commitment compatible*

[2] Prior to the decline, IRA contributions accounted for one-fifth of aggregate personal savings in the U.S.

[3] See Laibson [21] and O'Donoghue and Rabin [27] for a discussion of normative implications.

with the existence of self-control problems? This paper provides a positive answer – we suggest that commitment requires self-control.

To understand why commitment may require self-control, consider again the smoker who is choosing between

$$\{n\} \text{ and } \{n,s\}.$$

As before, anticipation of future temptation in $\{n,s\}$ leads to a preference for $\{n\}$. However, *he may be tempted by* $\{n,s\}$ *because it provides the opportunity to smoke tomorrow*. That is, he may be tempted by menus that contain tempting items. In such a case, choosing $\{n\}$ requires him to exert self-control and resist the temptation of $\{n,s\}$. Indeed, if the self-control cost is too high, he chooses not to commit. Put differently, the agent's problem is to decide whether to exert self-control today by choosing $\{n\}$, or to choose $\{n,s\}$ and (perhaps unsuccessfully) exert self-control tomorrow.

The literature has typically assumed that agents experience only immediate temptations. Under such an assumption, tomorrow's s (contained in the menu $\{n,s\}$) cannot tempt an agent today. Therefore, to model a smoker that does not commit, it is necessary to depart from the literature and allow for temptation by future consumption.

The remainder of the paper proceeds as follows. Section 2 presents our model, which is based on Gul and Pesendorfer [13, 14]. Section 3 provides axioms and a representation result, and Section 4 formally defines commitment in our model and contrasts it with the notion in [13, 14] . Section 5 demonstrates that the model unifies various behaviors, namely, a demand for commitment, preference reversals and a preference for early choice. Several applications are considered in Section 6: Normative implications of the model are shown differ from other models in the temptation literature. The evidence on hyperbolic discounting and preference reversals is shown to demonstrate the idea of temptation by future consumption and also to demonstrate that such temptation is not restricted to short time horizons. Behavioral evidence on projection bias is shown to be consistent with temptation by future consumption. The model is used to rationalize procrastination, and it is also shown to be able to simultaneously exhibit myopia and hyperopia. Section 7 concludes. Proofs are collected in appendices.

2. The Model

Our model is based on Gul and Pesendorfer [13, 14], who formalize the notions of temptation and self-control. Before introducing our model it is necessary to first introduce theirs.

Gul and Pesendorfer (henceforth GP) study an infinite horizon model in [14]. The domain of preference $\succsim$ is the set of infinite horizon choice problems Z. A choice problem $x \in Z$ is a set of lotteries, where each lottery μ is a measure over present consumption $c \in C$ and a continuation choice problem $y \in Z$; a budget set is an example of an infinite horizon choice problem, since a choice from a budget set yields immediate consumption and another budget set for tomorrow. Consider the following time-line:

$$\underset{x \succsim y}{\overset{t=0}{\bullet}} \text{ ―――――― } \underset{(c,z)\in x}{\overset{t=1}{\bullet}} \text{ ―――――― } \underset{(c',z')\in z}{\overset{t=2}{\bullet}} \text{ ―― }$$

At time 0 (the ex-ante stage), the agent makes a choice of menu. The chosen menu, say x, is faced at time 1, and a choice from x is made. Assuming for simplicity that objects in x are degenerate lotteries, the choice (c, z) from x yields some current consumption c, and a continuation menu z that is faced at time 2. At time 2, a choice from z is made, and the process is repeated ad infinitum. Thus, in period 0 is there a choice *of* menu, and in each subsequent period, there is a choice *from* a menu. Period 0 choice of menu is dictated by the primitive preference $\succsim$.

GP axiomatize 'Dynamic Self-control' (DSC) preferences which describe an agent who experiences temptation by immediate consumption. DSC preferences are represented by a recursive function $W(\cdot)$ defined, for all $x \in Z$, by

$$\begin{aligned} W(x) &= \max_{\mu \in x}\{U(\mu) + \left(V(\mu) - \max_{\eta \in x} V(\eta)\right)\}, \qquad (2.1) \\ \text{such that } U(\mu) &= \int_{C\times Z} (u(c) + \delta W(y))\, d\mu(c, y), \\ V(\mu) &= \int_{C\times Z} v(c) d\mu(c, y). \end{aligned}$$

To understand the representation, first interpret (2.1). The function U is called commitment utility since it represents the agent's utility if he committed to a singleton menu x.[4] Commitment utility is interpreted as capturing the agent's view of his long-run best-interest. On the other hand, the function V is interpreted as capturing his urges. It is referred to as temptation utility. Resisting temptation, that is, not choosing the V-maximizing choice, leads to a self-control cost $|V(\mu) - \max_{\eta \in x} V(\eta)|$. Therefore, (2.1) states that utility $W(x)$ is the maximum over x of commitment utility net of self-control cost. Observe that while $W(\cdot)$ is just a representation of the period 0 preference over menus, it is suggestive of how the agent chooses *from* a menu in subsequent periods: it suggests that when choosing from x in any period $t > 0$, the agent maximizes commitment utility net of self-control cost.

According to the functional forms, U and V are expected utilities. The commitment utility index is the sum of the utility from immediate consumption, and the utility W of the continuation menu discounted by δ. The temptation utility index is a function of immediate consumption only. This captures the feature that only immediate consumption tempts the DSC agent in any period $t > 0$.

This paper studies agents who are tempted by future consumption as well. 'DSC preferences with Future Temptation', or simply, Future Temptation (FT) preferences, have the same representation as above, except that temptation utility V is modelled as a recursive function,

$$\begin{aligned} V(\mu) &= \int_{C \times Z} \left(v(c) + \gamma \overline{V}(x)\right) d\mu(c, x), \\ \text{where } \overline{V}(x) &= \max_{\mu \in x} V(\mu). \end{aligned}$$

Temptation utility from current consumption is captured by v and that from continuation menus is captured by $\overline{V}$. The main feature of the model is the structure placed on $\overline{V}$: continuation menus are ranked by $\overline{V}$ according to the most tempting item contained in them. Thus, future temptations (that is, tempting items contained in a continuation menu) affect the agent in the form of a temptation by continuation menus. The discount factor γ parametrizes the strength of this temptation. DSC preferences obtain if $\gamma = 0$.

Comments

An important feature of GP's model (which is inherited by ours) is that period 0 is the period prior to the experience of temptation (see GP [14]). That is,

[4] If $x = \{\mu\}$ then $V(\mu) - \max_{\mu \in x} V(\mu) = 0$ and so $W(\{\mu\}) = U(\mu)$.

while choice *from* a menu in each period $t > 0$ is subject to temptation, ex-ante choice *between* menus is not. Thus, $\succsim$ reflects the agent's views when he is in a detached, 'cool' state (Lowenstein [23]) and is anticipating being in a 'hot state' when choosing from menus in each period $t > 0$. We emphasize that period 0 should be interpreted as a *hypothetical* period – the preference $\succsim$ captures how the agent would behave *if* he was not affected by temptation. The agent's *observed* choices are those in each period $t > 0$ when temptation is potentially experienced. The role of the period 0 preference $\succsim$ is to allow us to provide foundations for the FT model (Section 3).

To see that the model can capture the behavior of an agent with self-control problems who does not commit, consider some binary menu

$$x = \{(c, y_1), (c, y_2)\},$$

where immediate consumption is fixed at c for simplicity, and where the continuation menus y_1 and y_2 satisfy $y_1 \subset y_2$. Since there is no choice of immediate consumption, the choice from x is essentially a choice of what menu to face tomorrow. The menu y_1 offers commitment since $y_1 \subset y_2$; imagine that y is a singleton menu that contains only the option to abstain from smoking, whereas z is a binary menu that also contains the option of smoking.

What determines the agent's choice from x? Recall from our earlier description of the FT model that the choice from x (in some period $t > 0$) solves

$$\max_{(c,y_i)\in x} \{U(c, y_i) + \left(V(c, y_i) - \max_{(c,y_j)\in x} V(c, y_j)\right)\}.$$

The functional forms of U and V imply

$$\begin{aligned} U(c, y_i) &= u(c) + \delta W(y_i), \\ V(c, y_i) &= v(c) + \gamma \overline{V}(y_i), \end{aligned}$$

where

$$W(y_i) = \max_{\mu\in y_i}\{U(\mu) + \left(V(\mu) - \max_{\eta\in y_i} V(\eta)\right)\} \text{ and } \overline{V}(y_i) = \max_{\mu\in y_i} V(\mu).$$

For simplicity, let $u(c) = v(c) = 0$. Thus, the choice from x solves[5]

$$\max_{(c,y_i)\in x} \{\delta W(y_i) + \left(\gamma \overline{V}(y_i) - \max_{(c,y_j)\in x} \gamma \overline{V}(y_j)\right)\}. \tag{2.2}$$

[5] Observe that since the term '$\max_{(c,y_j)\in x} \gamma \overline{V}(y_j)$' is a constant when x is given, the choice essentially maximizes $\delta W(\cdot) + \gamma \overline{V}(\cdot)$.

The functions W and $\overline{V}$ evaluate a continuation menu y_i very differently. Indeed, $y_1 \subset y_2$ implies

$$W(y_1) \geq W(y_2) \text{ and } \overline{V}(y_1) \leq \overline{V}(y_2).$$

This is because temptations in y_2 make y_2 attractive according to $\overline{V}$, but unattractive according to W; observe that temptation utility enters W only in the form of a cost whereas it enters $\overline{V}$ as a benefit. Thus, the agent is tempted to choose y_2 but believes that choosing y_1 is in his best interest. His eventual choice is determined by (2.2). If he resists temptation and chooses commitment, he incurs the self-control cost $\left|\gamma\overline{V}(y_i) - \max_{(c,y_j)\in x} \gamma\overline{V}(y_j)\right|$, that is, commitment comes at the cost of self-control. If this cost is too high, he chooses not to commit.

3. Axioms and Representation Result

For any compact metric space X, $\Delta(X)$ denotes the set of all probability measures on the Borel σ-algebra of X, endowed with the weak convergence topology; $\Delta(X)$ is compact and metrizable [29]. Let $\mathcal{K}(X)$ denote the set of all nonempty compact subsets of X. When endowed with the Hausdorff topology, $\mathcal{K}(X)$ is a compact metric space [8].

The set C is a compact metric space that denotes possible consumption levels. The domain of preferences $\succsim$ is the set of choice problems Z. Each choice problem $z \in Z$ is a compact set of lotteries, where each lottery is a measure over current consumption and a continuation menu. Thus Z can be identified with $\mathcal{K}(\Delta(C \times Z))$. See GP [14] for the formal definition of Z and the homeomorphism between Z and $\mathcal{K}(\Delta(C \times Z))$. In particular, Z is compact metric.

For convenience, $\Delta(C \times Z)$ is written as Δ. Generic elements of Z are x, y, z whereas generic elements of Δ are μ, η, ν. For $\alpha \in [0,1]$, $\alpha\mu + (1-\alpha)\eta \in \Delta$ is the measure that assigns $\alpha\mu(A) + (1-\alpha)\eta(A)$ to each A in the Borel $\sigma-$algebra of $C \times Z$. Similarly, $\alpha x + (1-\alpha)y \equiv \{\alpha\mu + (1-\alpha)\eta : \mu \in x, \eta \in y\} \in Z$ is a mixture of the choice problems x and y.

The axioms imposed on $\succsim$ are related to those in GP [14] and are presented in three groups to facilitate comparison. The first set is identical to corresponding axioms in [14] and the second set strengthens corresponding axioms. The last axiom is the major point of departure.

Axiom 1 (Order). *$\succsim$ is a complete and transitive binary relation on Z.*

Axiom 2 (Continuity). *The sets $\{x : x \succsim y\}$ and $\{x : y \succsim x\}$ are closed.*

Axiom 3 (Independence). *For any $\alpha \in (0, 1)$,*

$$\{\mu\} \succ \{\eta\} \Longrightarrow \{\alpha\mu + (1 - \alpha)\nu\} \succ \{\alpha\eta + (1 - \alpha)\nu\}.$$

Axiom 4 (Set-Betweenness). $x \succsim y \Longrightarrow x \succsim x \cup y \succsim y$.

Axiom 5 (Stationarity). $z \succsim z' \Longleftrightarrow \{(c, z)\} \succsim \{(c, z')\}$.

The first two axioms are standard and the third is a version of the Independence axiom applied to singleton menus. The motivation for Independence is precisely as in GP [13, 14]. Roughly, Stationarity states that the ranking of choice problems is unchanged if the choice problems are pushed one period into the future.

To understand Set-Betweenness, begin with the stronger assumption

$$x \succsim y \Longrightarrow x \sim x \cup y \succsim y.$$

This axiom describes a standard decision-maker who evaluates a menu by its best element: if the best item in x is better than the best item in y, then $x \sim x \cup y$ since the best item in x and in $x \cup y$ are the same. The decision-maker is said to be *strategically rational* (Kreps [19]). Such a decision-maker is not worse off with larger menus and thus does not experience temptation. Set-Betweenness allows for the experience of temptation by permitting a preference for commitment, $x \succ x \cup y$.

Set-Betweenness also allows the decision-maker to resist temptation. Let $\mu, \eta \in \Delta$ be such that $\{\mu\} \succ \{\mu, \eta\}$, that is, η is tempting. When $\{\mu, \eta\} \sim \{\eta\}$ holds, the indifference suggests that the agent would choose the same item when faced with $\{\mu, \eta\}$ or $\{\eta\}$. That is, choice from $\{\mu, \eta\}$ is η and so, the decision-maker succumbs to temptation. On the other hand, the ranking $\{\mu, \eta\} \succ \{\eta\}$ suggests that μ is chosen from $\{\mu, \eta\}$ and so, the decision-maker resists temptation.

The next axiom strengthens two axioms in [14]. For any lottery $\mu \in \Delta(C \times Z)$, μ^1 denotes the marginal distribution over C and μ^2 the marginal distribution over Z. Let $\Delta_s \subset \Delta$ be the set of lotteries on $C \times Z$ with finite support and $\Delta_s(Z)$ the set of lotteries on Z with finite support. Let δ_z denote the lottery degenerate at menu z.

Define $\varphi : \Delta_s(Z) \longrightarrow Z$ by

$$\varphi(\sum p(x)\delta_x) = \sum p(x)x.$$

Two lotteries on Z that have the same $\varphi-$value induce the same uncertainty over continuation menus in the sense that the probability of *ultimately* choosing from a given continuation menu $z \in Z$ tomorrow is the same. However, the timing of the resolution of this uncertainty can be different. For example, consider lotteries μ, π such that

$$\mu^2 = \alpha\delta_z + (1-\alpha)\delta_{z'} \text{ and } \pi^2 = \delta_{\alpha z+(1-\alpha)z'}.$$

Then,

$$\varphi(\mu^2) = \varphi(\alpha\delta_z + (1-\alpha)\delta_{z'}) = \alpha z + (1-\alpha)z' = \varphi(\delta_{\alpha z+(1-\alpha)z'}) = \varphi(\pi^2).$$

Under both μ and π, the uncertainty about the continuation menu is the same: the agent chooses from z tomorrow with probability α and from z' with probability $(1-\alpha)$. However, there is early resolution of uncertainty in the former since, under μ, the uncertainty is played out today, but under π the uncertainty is resolved tomorrow.

Axiom 6 (Indifference to Timing). *For any $\mu, \eta, \pi, \nu \in \Delta_s$, if $\mu^1 = \pi^1, \eta^1 = \nu^1, \varphi(\mu^2) = \varphi(\pi^2)$ and $\varphi(\eta^2) = \varphi(\nu^2)$, then,*

$$\{\mu, \eta\} \backsim \{\pi, \nu\}.$$

According to the hypothesis, μ and π have the same first marginal and the same $\varphi-$value. That is, the two lotteries are similar, except that they may differ in how uncertainty about the continuation menu is resolved. The same is true for η and ν. Consequently, the choices available in the menus $\{\mu, \eta\}$ and $\{\pi, \nu\}$ may differ only in how uncertainty is resolved. Hence, indifference between the menus amounts to an indifference to the timing of resolution of uncertainty. GP formulate a weaker axiom restricted to singleton menus.

The axiom implicitly imposes a form of separability on preferences as well. Consider $\mu, \eta, \pi, \nu \in \Delta_s$ such that $\mu^1 = \pi^1, \eta^1 = \nu^1, \mu^2 = \pi^2$ and $\eta^2 = \nu^2$. Since $\varphi(\mu^2) = \varphi(\pi^2)$ and $\varphi(\eta^2) = \varphi(\nu^2)$ holds trivially, Indifference to Timing implies

$$\{\mu, \eta\} \backsim \{\pi, \nu\}.$$

Observe that the set of marginals on C induced by both these menus are same, and so are the induced sets of marginals on Z. Thus, the indifference between $\{\mu, \eta\}$ and $\{\pi, \nu\}$ suggests that 'only marginals matter' and hence, preferences are insensitive to correlations between current consumption and continuation menus. GP's axiom 'Separability' is implied by our Indifference to Timing axiom – their axiom adopts the special case where

$$\begin{aligned} \mu &= \eta = \frac{1}{2}(c, z) + \frac{1}{2}(c', z') \\ \pi &= \nu = \frac{1}{2}(c, z') + \frac{1}{2}(c', z). \end{aligned}$$

Our final axiom departs from GP's in a more fundamental way. GP's final axiom states

Axiom (Temptation by Immediate Consumption). *For any $\mu, \eta, \nu \in \Delta$ such that $\eta^1 = \nu^1$, if $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ and $\{\mu\} \succ \{\mu, \nu\} \succ \{\nu\}$, then*

$$\{\mu, \eta\} \sim \{\mu, \nu\}.$$

According to the axiom, if the marginal distributions on present consumption implied by two lotteries are the same, then the lotteries are equally tempting, implying that continuation problems do not tempt the agent. Our axiom is formulated so that continuation problems may tempt the agent.

Axiom 7 (Temptation Stationarity). $x \succ x \cup y \Longleftrightarrow \{(c, x)\} \succ \{(c, x), (c, y)\}$.

The ranking $x \succ x \cup y$ reveals that y tempts x, that is, the most tempting item in y is more tempting than that in x. Similarly, the ranking $\{(c, x)\} \succ \{(c, x), (c, y)\}$ reveals that (c, y) is more tempting than (c, x). Thus, the axiom states that y tempts x if and only if (c, y) tempts (c, x), that is, if and only if the continuation menu y tempts the continuation menu x. Note that this axiom provides a way of distinguishing an agent who experiences temptation by future consumption from one who does not. When y contains temptations, the former type of agent strictly prefers, ex-ante, not to have the option to choose (c, y), as reflected in the ranking $\{(c, x)\} \succ \{(c, x), (c, y)\}$. An agent who does not experience temptation by future consumption would not care whether or not he has (c, y) and hence, he expresses the ranking $\{(c, x)\} \sim \{(c, x), (c, y)\}$.

As mentioned in Section 2, the preference $\succsim$ is meant to capture how the agent would behave in a special period 0 where he does not experience temptation. This assumption is crucial for Set-Betweenness and Temptation Stationarity to make sense. As we argued in the Introduction, temptation by future consumption can lead to the absence of a preference for commitment. Thus, if such temptation is experienced in period 0 (that is, if $\succsim$ does not represent behavior in the absence of temptation), then the absence of a preference for commitment may not imply the absence of temptation within a menu. That is, it is no longer true that y contains tempting items if *and only if* $x \succ x \cup y$. The intuition underlying Set-Betweenness and supporting Temptation Stationarity rests on this equivalence holding.

Representation Result

The preference $\succsim$ is said to be *nondegenerate* if there exists $x, y \in Z$ such that $x \supset y$ and $x \succ y$.

Theorem 3.1. *If the nondegenerate preference $\succsim$ satisfies Axioms 1-7 then there exist $\delta \in (0,1)$, $\gamma \in (0,1)$, functions $u, v : C \longrightarrow \mathbb{R}$, $\overline{V} : Z \longrightarrow \mathbb{R}$, and $W : Z \longrightarrow \mathbb{R}$ that represents $\succsim$ such that for all $z \in Z$,*

$$\begin{aligned} W(x) &= \max_{\mu \in x} \{U(\mu) + \Big(V(\mu) - \max_{\eta \in x} V(\eta)\Big)\}, \qquad (3.1) \\ \textit{where } U(\mu) &= \int_{C \times Z} (u(c) + \delta W(y))\, d\mu(c,y), \\ V(\mu) &= \int_{C \times Z} \big(v(c) + \gamma \overline{V}(y)\big)\, d\mu(c,y), \\ \overline{V}(x) &= \max_{\eta \in x} V(\eta). \end{aligned}$$

Each of the functions $u, v, \overline{V}$ and W is continuous and $\overline{V}$ is linear.

Conversely, for any $\delta \in (0,1)$, $\gamma \in (0,1)$, continuous $u, v : C \longrightarrow \mathbb{R}$, there is a unique continuous function W satisfying (3.1), and the preference it represents satisfies Axioms 1-7.

The preferences in Theorem 3.1 are referred to as DSC preferences with Future Temptation, or simply Future Temptation (FT) preferences. Note that DSC preferences are not a special case of FT preferences. In the case of DSC preferences, temptation utility V has the form

$$V(\mu) = \int_C v(c) d\mu^1(c),$$

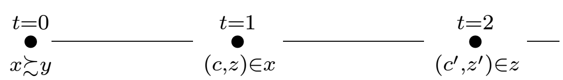
which amounts to our representation with $\gamma = 0$. For FT preferences, $\gamma > 0$.

For any W as in (3.1) that represents $\succsim$, the corresponding tuple (u, v, δ, γ) is referred to as a representation of $\succsim$. Theorem 3.2 establishes uniqueness properties of the representation. Say that $\succsim$ *exhibits a preference for commitment at* y if there exists $x \subset y$ such that $x \succ y$.

Theorem 3.2. *Suppose $\succsim$ exhibits a preference for commitment at some y, and let (u, v, δ, γ) be a representation of $\succsim$. Then $(u', v', \delta', \gamma')$ also represents $\succsim$ if and only if $\delta = \delta'$, $\gamma = \gamma'$ and there exist $\alpha > 0$, and $\beta_u, \beta_v \in \mathbb{R}$ such that $u' = \alpha u + \beta_u$ and $v' = \alpha v + \beta_v$.*

4. Demand for Commitment

The FT model permits agents to have a preference for commitment, and yet not demand commitment. This section elucidates the meaning of this statement. Recall the time-line of the FT model:



As discussed in Section 2, the ex-ante preference $\succsim$ describes how the agent would rank menus in a hypothetical period 0 where the agent experiences no temptation. The observable choices of this agent – the choices that are subject to temptation – are those in each subsequent period $t > 0$ where temptation is experienced. When defining 'commitment', GP concentrate on commitment in period 0:

Definition 4.1. *$\succsim$ exhibits a preference for commitment at y if there exists $x \subset y$ such that $x \succ y$. Say that $\succsim$ exhibits a preference for commitment if there exists $y \in Z$ such that $\succsim$ exhibits a preference for commitment at y.*

Thus, an agent is said to prefer commitment at y if he commits to a subset in an ex-ante stage, when he is not experiencing temptation. However, we are interested in the question of whether he would commit to a subset of y if he were not in the ex-ante stage, and did experience temptation. If he commits in such a scenario, we say that he has a *demand for commitment at* y. Since only choice *from* a menu is subject to temptation, we define a demand for commitment in terms of choice of a continuation menu.

Definition 4.2. *$\succsim$ exhibits a demand for commitment at y if there exists $x \subset y$ such that*

$$\{(c,x),(c,y)\} \succ \{(c,y)\},$$

for some $c \in C$. Say that $\succsim$ exhibits a demand for commitment if it exhibits a demand for commitment at some y.

The ranking $\{(c,x),(c,y)\} \succ \{(c,y)\}$ implies that (c,x) is chosen from $\{(c,x),(c,y)\}$. Therefore the definition states that when there is a choice between continuation menus x and y, the agent is said to demand commitment if he chooses the smaller continuation menu x, that is, if he chooses (c,x) from $\{(c,x),(c,y)\}$. An assumption about the timing of commitment is implicit in Definitions 4.1 and 4.2: commitment is chosen in the period before it is received. This implies that there is a passage of time between when commitment is chosen and when it is received; for example, the time between joining a rehabilitation clinic and actually starting treatment.

It is important to understand the relationship between the demand and preference for commitment. A demand for commitment implies a preference for commitment.[6] In what follows, we show that the converse is true for DSC preferences, but not necessarily for FT preferences.

For a DSC agent, a preference for commitment at y implies a demand for commitment at y: when Temptation by Immediate Consumption is satisfied, continuation menus are not tempting, and hence, for all x, y,

$$\{(c,x)\} \sim \{(c,x),(c,y)\}.$$

[6] To see this, let $x \subset y$ and $\{(c,x),(c,y)\} \succ \{(c,y)\}$. Then Set-Betweenness implies $\{(c,x)\} \succ \{(c,y)\}$, which, by Stationarity, implies $x \succ y$.

By Stationarity, $x \succ y$ implies $\{(c,x)\} \succ \{(c,y)\}$. Therefore, if $x \subset y$ and $x \succ y$, then

$$\{(c,x)\} \sim \{(c,x),(c,y)\} \succ \{(c,y)\},$$

That is, (c,x) is chosen from $\{(c,x),(c,y)\}$.

In the FT model, however, a preference for commitment at y may not imply a demand for commitment at y. To see this, observe that if $x \subset y$, then $x \succ y$ is equivalent to $x \succ x \cup y$, which by Temptation Stationarity is equivalent to

$$\{(c,x)\} \succ \{(c,x),(c,y)\}.$$

Hence a preference for commitment at y implies a temptation to choose (c,y). By Definition 4.2, the agent demands commitment when

$$\{(c,x)\} \succ \{(c,x),(c,y)\} \succ \{(c,y)\},$$

that is, when he resists the temptation to choose (c,y). However, if the self-control cost of resisting the temptation is too high, the agent submits to the temptation of (c,y),

$$\{(c,x)\} \succ \{(c,x),(c,y)\} \sim \{(c,y)\}.$$

In such a case, there is a preference for commitment at y, but no demand.

Hence temptation by future consumption severs the link between the preference and demand for commitment that exists in the DSC model. The result that a preference for commitment does not imply a choice to commit has the same paradoxical nature as the result in GP [14], where dynamically consistent preferences may produce choices that appear dynamically inconsistent. This wedge between preferences and choices arises because temptations can make choices appear inconsistent with preferences. For instance, one may prefer $\{\mu\}$ over $\{\eta\}$, but choose η from $\{\mu,\eta\}$ due to a temptation to choose η. Temptation by future consumption serves the purpose of making certain choices tempting, namely making (c,y) tempting in $\{(c,x),(c,y)\}$. This introduces the appropriate wedge, since now a preference for commitment (which is equivalent to $\{(c,x)\} \succ \{(c,y)\}$) does not necessarily imply that (c,x) is chosen from $\{(c,x),(c,y)\}$.

We close with a comment about the ex-ante preference $\succsim$. Since period 0 choice between menus (captured by $\succsim$) is presumably in the absence of temptation, the FT agent has no reason not to commit in period 0 if he has a preference for doing so. That is, although we set out to model a smoker that does not commit, our FT smoker would always commit in period 0! This would appear to be a shortcoming of the model, since the model is meant to describe agents who have self-control problems and yet do not commit. However, we remind the reader that period 0 is a hypothetical construct (Section 2); the period 0 preference $\succsim$ describes how the agent would behave *if* he did not experience temptation. Only the choices in each period $t > 0$ are subject to temptation, and thus, correspond to actual choice.

5. Temptation Discounting

Two primitives of the FT model are δ and γ, the commitment and temptation discount factors respectively. In this section we define some behaviors and discuss how they depend on the relative magnitudes of δ and γ. In particular, we discuss how the demand for commitment depends on δ and γ.

5.1. Preference Reversals

Subjects in psychology experiments on 'preference reversals' typically prefer a small immediate reward to a large delayed reward, but reverse preferences in favor of the latter when both rewards are pushed into the future by a sufficient number of periods (see, for instance, Kirby and Hernnstein [16]). For instance, they may choose (\$100, now) over (\$110, one month), but after both rewards are delayed by two months, they switch preferences and choose (\$110, three months) over (\$100, two months). Preference reversals have been interpreted in terms of temptation by the immediately available reward, which disappears when both rewards are in the future.

Preference reversals can be defined in our model as follows. For some $c \in C$ and any $\mu \in \Delta$ let $\mu^{+0} = \mu$ and inductively, $\mu^{+t} = (c, \{\mu^{+(t-1)}\})$. That is, μ^{+t} represents getting the reward μ with a delay of t periods, where in each of the t periods the agents receives some fixed consumption c.

Definition 5.1. $\succsim$ *exhibits preference reversals if* $\{\mu\} \succ \{\mu, \eta\} \sim \{\eta\}$ *implies the existence of* t^* *such that*

$$\begin{aligned} \{\mu^{+t}, \eta^{+t}\} &\sim \{\eta^{+t}\} \text{ for all } t < t^*, \\ \{\mu^{+t}, \eta^{+t}\} &\succ \{\eta^{+t}\} \text{ for all } t \geq t^*. \end{aligned}$$

Thus, choices from menus of the form $\{\mu^{+t}, \eta^{+t}\}$ with different t determine whether the agent exhibits preference reversals. The rankings

$$\{\mu^{+t}, \eta^{+t}\} \sim \{\eta^{+t}\} \text{ for } t < t^*$$

state that η^{+t} is chosen from $\{\mu^{+t}, \eta^{+t}\}$ when $t < t^*$. The rankings

$$\{\mu^{+t}, \eta^{+t}\} \succ \{\eta^{+t}\} \text{ for } t \geq t^*$$

state that η^{+t} is no longer chosen from $\{\mu^{+t}, \eta^{+t}\}$ when $t \geq t^*$, and thus a reversal is observed. Note that the hypothesis of the Definition requires that the choice of η from $\{\mu, \eta\}$ be subject to overwhelming temptation, thereby suggesting that temptation is the underlying cause of the preference reversal – $\succsim$ is said to exhibit preference reversals if an irresistible temptation in the menu $\{\mu, \eta\}$ is resisted when both rewards μ and η are pushed sufficiently far into the future.

Definition 5.1 is not restricted to preference reversals involving earlier rewards and later rewards. It allows for preference reversals between rewards available at the same date. An examples of such a reversal can be found in Trope and Liberman [34] who find that subjects tend to prefer watching a noneducational but entertaining movie now to an educational but unentertaining movie now, yet they switch preferences when the movie is to be watched at some point in the future.[7] Both rewards are available at the *same* date.

5.2. Preference for Early Choice

The literature on choice under risk has studied attitudes toward the timing of resolution of risk, that is, the agent's preference regarding whether uncertainty regarding tomorrow's consumption should resolve today or tomorrow. We consider a parallel attitude in our framework: the agent's attitude towards the timing of self-control, that is, the agent's preference regarding whether to exert self-control today when deciding tomorrow's consumption, or to postpone exerting self-control until tomorrow.

[7]Such rewards can be modelled as follows. Let the consumption set C be a subset of $\mathbb{R}^2$, where the first coordinate of $(c_1, c_2) \in C$ represents the 'education rating' and the second represents the 'entertainment rating'. The movies can thus be represented by elements of C. To get 'temptation by entertaining movies', let v be a function of c_2 only, u a function of c_1 only, and u, v strictly increasing.

Definition 5.2. $\succsim$ *exhibits preference for early choice if*

$$\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\} \text{ implies } \{(c, \mu), (c, \eta)\} \succ \{(c, \{\mu, \eta\})\}.$$

The hypothesis $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ implies that when the agent faces the menu $\{\mu, \eta\}$, he is tempted by η but is able to exert self-control and resist it. Furthermore, by Temptation Stationarity, $\{\mu\} \succ \{\mu, \eta\}$ implies $\{(c, \mu)\} \succ \{(c, \mu), (c, \eta)\}$, that is, when the agent faces $\{(c, \mu), (c, \eta)\}$, he is tempted by (c, η). These two observations imply the following: Under the menu $\{(c, \mu), (c, \eta)\}$, the decision-maker has the opportunity to exert self-control today and choose to consume μ tomorrow. Under $\{(c, \{\mu, \eta\})\}$, tomorrow's consumption has to be chosen tomorrow, and tomorrow he will exert self-control and choose μ. Thus the agent's ranking of the menus $\{(c, \mu), (c, \eta)\}$ and $\{(c, \{\mu, \eta\})\}$ reveals his preference for when to exert self-control. The axiom states that he has a preference for exerting self-control today rather than tomorrow.

5.3. A Characterization

Theorem 5.3 reveals that demand for commitment, preference reversals and preference for early choice are all behavioral manifestations of large temptation discounting.

Theorem 5.3. *Let the nondegenerate FT preference* $\succsim$ *exhibit a preference for commitment. Then the following statements are equivalent.*
(a) $\gamma < \delta$.
(b) $\succsim$ *exhibits demand for commitment.*
(c) $\succsim$ *exhibits preference reversals.*
(d) $\succsim$ *exhibits preference for early choice.*

The intuition behind the result is that sufficient discounting of the temptation utility of future consumption makes it possible to resist temptation by future consumption. There exists a demand for commitment when the agent can resist the temptation by some menu y, and this is possible only when the temptation utility of the menu is discounted sufficiently. Large discounting is also responsible for preference reversals because the latter occurs when an irresistible immediate temptation is resisted when it is pushed into the future. Finally, $\{(c, \mu), (c, \eta)\}$ is strictly preferred to $\{(c, \{\mu, \eta\})\}$ (and so there is preference for early choice) if it is easier to resist the future temptation of η in $\{(c,\mu),(c,\eta)\}$ than the immediate temptation of η in $\{\mu,\eta\}$.

The restriction $\gamma<\delta$ is desirable for two reasons. Firstly, there is evidence in favor of a demand for commitment (Elster [10], Schelling [32]) and preference reversals (see references in Ainslie [1]). Secondly, $\gamma\geq\delta$ produces behavior that is not intuitive. For instance, when $\gamma>\delta$, the following behavior arises: for any $\mu,\eta\in\Delta$, $\{\mu\}\succ\{\mu,\eta\}\succ\{\eta\}$ implies the existence of t^* such that

$$\begin{aligned}\{\mu^{+t},\eta^{+t}\} &\succ \{\eta^{+t}\} \textit{ for all } t<t^*,\\ \{\mu^{+t},\eta^{+t}\} &\sim \{\eta^{+t}\} \textit{ for all } t\geq t^*.\end{aligned}$$

That is, temptations that can be resisted when they are immediate become irresistible when they are pushed into the future. We normally expect immediate temptations to be harder to resist.

The main motivation for the FT model was to explain why agents with self-control problems may not commit. By Theorem 5.3 it would appear that we need to assume $\gamma\geq\delta$. But we just argued that $\gamma<\delta$ is the desirable restriction! Fortunately, Theorem 5.3 states that when $\gamma<\delta$, there exists *at least* one y where there is a demand for commitment, and therefore does not rule out the possibility that there are other y where there is a preference for commitment but no demand. The following table gives conditions under which a demand for commitment does or does not exist: for x,y such that $x\subset y$ and $x\succ y$,

	commits	does not commit
$\gamma=0$	always	–
$0<\gamma<\delta$	if $\frac{W(x)-W(y)}{\overline{V}(y)-\overline{V}(x)}>\frac{\gamma}{\delta}$	if $\frac{W(x)-W(y)}{\overline{V}(y)-\overline{V}(x)}\leq\frac{\gamma}{\delta}$
$\delta\leq\gamma$	–	always

See Appendix D for the derivation of the table.

6. Applications

6.1. Normative Implications

The normative implications of temptation models of addiction are different from those of models stemming from Becker and Murphy [6]. These latter models suggest that addiction is a rational choice, and so, the government has no role in regulating the addict's choices, except to the extent that his choices impose an externality on others. On the other hand, models that incorporate self-control problems, such as GP [15], imply that, along with externalities on others, an addict imposes negative internalities on himself and hence there is a role for the government to intervene.

However, the temptation literature has typically hypothesized that agents are tempted only by immediate consumption. A consequence of this assumption is that addicts that don't seek treatment in fact have no self-control problems. That is, they are 'happy addicts' in that they are addicted because it is optimal for them, and thus require no help. In contrast, the FT model suggests that an addict that does not seek commitment may be a very unhappy addict.[8] He does not seek commitment because of a lack of self-control. Hence, there is a role for intervention.

Furthermore, the welfare policy prescription of models that do not permit temptation by future consumption is simple: introduce commitment mechanisms into the market. In this way, agents with self-control problems take advantage of commitment opportunities and thus improve their welfare, whereas agents who do not have self-control problems are unaffected. But such a prescription may not be effective if agents experience temptation by future consumption. Such agents postpone commitment, and thus may never take advantage of the commitment opportunities. For such agents, narcotics control may be a superior policy.

6.2. Time of Preference Reversals

A large amount of experimental evidence in psychology reveals that agents discount future rewards by using a hyperbolic discount function (see Ainslie [1] for an

[8] In its current form, the FT model is not a model of addiction since it does not possess the nonseparability that is presumably characteristic of addiction. However, the model can be extended appropriately in a manner analogous to [15] in order to accommodate this.

overview of the literature). In what follows, we restrict attention to Mazur's [26] version of the hyperbolic discounting functional form, which has fit experimental evidence particularly well. According to his formulation, a reward that is delayed by d periods is discounted by

$$\frac{1}{1+kd},$$

where k parameterizes the subject's sensitivity to delay.

Let s^{+0} denote a small immediate reward s, and let l^{+d} denote a large reward l available with a delay of d periods. If a subject chooses s^{+0} over l^{+d}, then the hyperbolic discount function implies that in order to induce a preference reversal, both rewards must be delayed by $\widehat{\tau}(s^{+0}, l^{+d})$ periods, where

$$\widehat{\tau}(s^{+0}, l^{+d}) = \frac{s(1+kd)-l}{k(l-s)},$$

and where k is a parameter that captures the subject's 'sensitivity' to delay. That is, $\widehat{\tau}(s^{+0}, l^{+d})$ captures the time at which a preference reversal takes place for the rewards s^{+0} and l^{+d}. Observe that $\widehat{\tau}(s^{+0}, l^{+d})$ is increasing in s and d, and decreasing in l. The empirical evidence on hyperbolic discounting serves as indirect empirical evidence in favor of these properties of $\widehat{\tau}$. Some direct evidence may be found in Ainslie and Haendel [2].

How does the time-of-reversal function τ of DSC and FT (with $\gamma < \delta$) agents compare with the $\widehat{\tau}$ function above? In order to answer this question, we first need to define a time-of-reversal function $(\mu, \eta) \mapsto \tau(\mu, \eta)$ for these models. Definition 5.1 guides us. Take any μ, η such that $\{\mu\} \succsim \{\eta\}$. A preference reversal is observed for the pair of rewards (μ, η) if the hypothesis[9]

$$\{\mu\} \succ \{\mu, \eta\} \sim \{\eta\}$$

in Definition 5.1 holds, in which case we can set $\tau(\mu, \eta) = t^*$, where $t^* > 0$ is as in the Definition. If the hypothesis does not hold, then no preference reversal is observed for (μ, η), and we set $\tau(\mu, \eta) = 0$. To cover all remaining cases, let $\tau(\mu, \eta) = \tau(\eta, \mu)$ for all μ, η.

[9] This 'overwhelming temptation' condition is equivalent to $\frac{U(\mu)-U(\eta)}{V(\eta)-V(\mu)} \in (0, 1]$, where U and V are as in the DSC or FT representation.

Suppose a preference reversal is observed for the rewards μ, η. Then DSC agents exhibit,[10]

$$\tau^{DSC}(\mu, \eta) = 1,$$

whereas FT agents exhibit,[11]

$$\tau^{FT}(\mu, \eta) = \frac{\ln \frac{U(\mu) - U(\eta)}{V(\eta) - V(\mu)}}{\ln \frac{\gamma}{\delta}},$$

where U and V are as in the FT representation. That is, for DSC agents, a single period delay suffices to induce a reversal – DSC agents are only tempted by immediate consumption, and so a single period delay leads all temptation to be discounted fully. For FT agents, the required delay depends on the rewards and on the agent's discount factors.

First, we inquire whether the τ^{FT} function shares the same properties as $\widehat{\tau}$. For $r = s, l$, define r^{+t} as in Section 5.1 and for expositional simplicity, suppose $u(c) = v(c) = 0$. If the agent chooses s^{+0} from $\{s^{+0}, l^{+d}\}$ and a preference reversal is observed, then

$$\tau^{FT}(s^{+0}, l^{+d}) = \frac{\ln \frac{\delta^d u(l) - u(s)}{v(s) - \gamma^d v(l)}}{\ln \frac{\gamma}{\delta}}.$$

If u and v are strictly increasing functions, then the τ^{FT} function makes the same qualitative predictions as $\widehat{\tau}$, that is, τ^{FT} is increasing in s and d and decreasing in l. According to the model, the greater the value of s or d, or the smaller the value of l, the more tempting the small reward is, and so, the greater the number of periods of delay required in order to induce a preference reversal. Observe that the role of k in $\widehat{\tau}$ is played by $\frac{\gamma}{\delta}$ in τ^{FT}. Apparently, the 'sensitivity' to delay corresponds to how fast temptation utility is discounted (relative to commitment utility) when rewards are pushed into the future.

Thus, τ^{FT} shares the same features as $\widehat{\tau}$, while τ^{DSC} clearly does not. This supports the idea that agents are tempted by future consumption.

The evidence on preference reversals also reveals that temptation by future consumption is not restricted to a short time horizon. For instance, Ainslie and

[10] Since future temptations do not affect DSC agents, delaying rewards by one period suffices to induce a preference reversal [14].

[11] See the proof of "$(a) \Longrightarrow (c)$" in Appendix C. Strictly speaking, since time is discrete, τ^{FT} should be rounded off to the next integer. The same is true for $\widehat{\tau}$.

Haendel [2] and Kirby and Hernnstein [16, Experiment 3] find that the time of a reversal may be in years – the latter uncovers a preference reversal at almost 3.5 years delay. That is, it may take a delay of 3.5 years before the temptation of a reward is resistible, suggesting that rewards that are that far in the future may tempt.

6.3. Projection Bias vs. Temptation by Future Consumption

Various studies suggest that agents exhibit a 'projection bias' when predicting their future tastes: agents tend to systematically exaggerate the extent to which their future tastes will resemble their current tastes (Lowenstein and Schkade [24], Lowenstein, O'Donoghue and Rabin [25]). An experiment by Read and van Leeuwen [31] is often cited as evidence. Read and van Leeuwen give their subjects a choice between a healthy and unhealthy snack. The snacks are to be received after one week, either at a time when the subjects are in a hungry state H_1 (just before lunch) or a satiated state S_1 (immediately after lunch). Subjects are approached either when they are currently in a hungry state H_0 or satiated state S_0. The following table depicts the proportion of subjects who choose the unhealthy snack in each of the four scenarios:

	H_1	S_1
H_0	78%	42%
S_0	56%	26%

Thus, for any given current state, subjects choose the unhealthy snack more often when they anticipate being hungry. This presumably reflects the greater desire for the unhealthy snack in a hungry state. More interestingly, for any given future state, subjects choose the unhealthy snack more often when they are currently hungry. This is taken as evidence of projection bias since subjects who are currently hungry (and thus desire the unhealthy snack) act as if their future taste for the unhealthy snack will be similar to their current taste for it.

The choices admit an alternative interpretation: subjects are more sensitive to future cravings when they are currently experiencing a craving. The unhealthy snack may tempt subjects – more so when they are hungry – and the idea of having a tempting snack later is more attractive to them when they are currently hungry.[12] The FT model can capture such a story, once nonstationarity of temp-

[12] The healthy and unhealthy food items used in the experiment were among those ranked by a subset of the subjects according to how healthy or unhealthy they were.

tation preference is allowed for. For a simple illustration, consider the following generalization of the FT utility function:

$$\begin{aligned} W_t(x) &= \max_{\mu \in x}\{U_t(\mu) + \Big(V_t(\mu) - \max_{\eta \in x} V_t(\eta)\Big)\}, \\ \text{where } U_t(\mu) &= \int_{C\times Z} (u(c) + \delta W_{t+1}(y))\, d\mu(c,y), \\ V_t(\mu) &= \lambda_t \int_{C\times Z} \big(v(c) + \gamma \overline{V}_{t+1}(y)\big)\, d\mu(c,y), \\ \overline{V}_t(x) &= \max_{\eta \in x} V_t(\eta). \end{aligned}$$

The FT representation is the special case where $\lambda_t = 1$ for all t. Thus, the generalization permits the strength of temptation, parametrized by λ_t, to vary from period to period (exogenously).

Let a period be a week long. Thus, the choice problem faced by Read and van Leeuwen's subjects is

$$x = \{b^{+1}, h^{+1}\},$$

that is, the set containing the option to commit to the unhealthy snack b or healthy snack h for the next period; under either option, immediate consumption and the menu to be received in two periods is the same.[13] The choice from x at time t maximizes $U_t(\mu) + (V_t(\mu) - \max_{\eta \in x} V_t(\eta))$, which is equivalent to maximizing

$$U_t(\cdot) + V_t(\cdot)$$

over x. Given the structure on $U_t(\cdot)$ and $V_t(\cdot)$, the agent chooses b^{+1} if and only if

$$\delta u(b) + \lambda_t \gamma \lambda_{t+1} v(b) \geq \delta u(h) + \lambda_t \gamma \lambda_{t+1} v(h).$$

Assume that b is tempting, so that $u(h) > u(b)$ and $v(b) > v(h)$. Also, suppose that λ_t (resp. λ_{t+1}) captures the subject's state in period t (resp. $t+1$), so that higher levels of hunger correspond to higher values of λ_t (resp. λ_{t+1}). Under these assumptions, the model captures the choices of Read and van Leeuwen's subjects: higher the value of current and/or future hunger, the more likely the subjects are to choose the unhealthy snack.

[13] Formally, for $c, b, h \in C$ and $z \in Z$, let $b^{+1} \equiv (c, \{(b,z)\})$ and similarly, $h^{+1} \equiv (c, \{(h,z)\})$.

6.4. Procrastination

Procrastination occurs when an agent delays doing a task which, in some sense, he finds preferable to do now rather than later (Akerlof [3]). O'Donoghue and Rabin [27, 28] develop a theory of procrastination: given some task that is worth doing now rather than later, they suggest that a decision-maker procrastinates because, firstly, he incorrectly believes that he will do the task tomorrow if he delays today, and secondly, given the incorrect belief, it is optimal to delay. The naivete in expectations is an important ingredient in their theory.

We suggest another explanation: procrastination occurs because delaying is tempting. Thus, it is possible for a fully self-aware, rational agent to find it optimal to delay doing a task, despite preferring to commit to doing the task now. We illustrate how the FT model can be used to produce procrastination in the context of saving for retirement through a commitment asset.

There are T periods. A finite horizon choice problem x_t is an element of Z_t, where $Z_T = \mathcal{K}(\Delta(C))$ and for $t < T$, $Z_t = \mathcal{K}(\Delta(C \times Z_{t+1}))$. The representation W_0 for the preference $\succsim$ is defined inductively as follows. The utility W_T of a period T menu x_T is defined by

$$W_T(x_T) = \max_{c_T \in x_T} \{u(c) + v(c) - \max_{c'_T \in x_T} v(c')\},$$

and inductively, for all $t < T$,

$$\begin{aligned} W_t(x_t) \quad = \quad & \max_{(c_t, x_{t+1}) \in x_t} \{u(c) + \delta W_{t+1}(x_{t+1}) + v(c) + \gamma \overline{V}_{t+1}(x_{t+1})\} \\ & - \max_{(c'_t, x'_{t+1}) \in x_t} \{v(c') + \gamma \overline{V}_{t+1}(x'_{t+1})\}. \end{aligned}$$

Let $0 < \gamma < \delta$, $u = v$ and let u be increasing in c. The agent receives an endowment $\omega_t > 0$ every period $t < T$, but in the last period he has no endowment. In each period, besides choosing consumption, he decides whether or not to commit d units of consumption to a 401(k) plan. Contributions to the 401(k) can be consumed only in the last period, and there are no other saving opportunities. Formally, the choice problem is[14]

$$
\begin{aligned}
x_0(w_0) &= \{(c_0, x_1(w_1 - d_1)) : c_0 \leq w_0 \text{ and } d_1 = 0 \text{ or } d\}, \\
x_t(w_t - d_t) &= \{(c_t, x_{t+1}(w_{t+1} - d_{t+1})) : c_t \leq w_t - d_t \text{ and } d_{t+1} = 0 \text{ or } d\},\ 0 < t < T, \\
x_T(w_T) &= \{c_T : c_T \leq w_T \text{ and } w_T = \sum_{t=0}^{T-1} d_t\}.
\end{aligned}
$$

He enters period 0 with resources w_0, chooses current consumption $c_0 \leq w_0$, and the continuation menu $x_1(w_1 - d_1)$. The choice of continuation menu is made by choosing $d_1 \in \{0, d\}$, that is, by deciding whether or not to contribute d in the asset. The same kind of choices are made in every period $t < T$. In the final period T, his resources are $\sum_{t=0}^{T-1} d_t$, the quantity contributed to the 401(k) over his life, and his only choice is to decide how much of this to consume.

Assume that life-cycle commitment utility is maximized if he begins saving for retirement from the first period and that life-cycle temptation utility is maximized if he begins saving by some period $t^* > 0$. That is, commitment utility is maximized when $d_t = d$ for all $t < T$, and temptation utility is maximized when $d_t = 0$ for all $t < t^*$ and $d_t = d$ for all $t \geq t^*$. Given $\gamma < \delta$, this assumption is satisfied if the time horizon is sufficiently long. Note that there is a temptation to delay until t^*, so that any commitment before t^* is at the cost of self-control. For appropriate parameter values (for instance, a small d), the benefit of committing (that is, the discounted utility of gaining d units of consumption in the last period) is smaller than the self-control cost for all $t < t^*$. For $t \geq t^*$, there is no self-control cost of committing, and so he commits. That is, the agent rationally procrastinates on saving for retirement until t^*. The reason he eventually starts saving is that delaying ceases to be tempting once the deadline is close enough.

This demonstrates that the FT model can produce procrastination in using a commitment device. Turn to the question of how DSC preferences may be used to produce such behavior. Observe that we assumed that the commitment decision is made in the period prior to when commitment is received, that is, the agent specifies in advance the proportion of his income to save in an 401(k). In such an environment, DSC preferences cannot exhibit procrastination in commitment.

[14] Recall from Section 4 that commitment is chosen in the period prior to when it is received. This is evident in the way the choice problems are defined. For instance, at $t = 0$ the agent chooses d_1.

A DSC agent is tempted to maximize utility from *immediate* consumption. Since the 401(k) contribution is not coming out of current income (it is coming out of tomorrow's income), the DSC agent would not be tempted to postpone saving – he would take advantage of the opportunity to commit in advance. If one changes the specification of the environment so that first, the 401(k) contribution is made out of immediate income and second, it is not possible to pre-specify a contribution rate, then the DSC model can produce procrastination as well. However, it should be noted that the opportunity to pre-specify a contribution rate exists in real environments.

6.5. Temptation to Save

While there is evidence that people under-save, procrastinate, over-indulge in food, tobacco, alcohol or narcotics at the cost of future health problems, etc., there is also evidence that some people over-save, over-work, under-indulge etc. That is, alongside evidence of myopic behavior there is also evidence of hyperopic behavior. Evidence of under-indulgence is discussed in Kivetz and Simonson [17], and over-saving lends itself as a possible explanation for why retired people do not dissave as much as predicted by the life-cycle model and its variants (see Browning and Crossley [9] and Krusell, Kurusçu and Smith [20]). While myopia is typically understood in terms of temptation by immediate consumption, hyperopia can be understood in terms of temptation by consumption in the future. It can be explained by the FT model, and that too without departing from the restriction $\gamma < \delta$, that is, without departing from the assumption that delayed temptations are easier to resist than immediate consumption. In particular, the model can simultaneously account for myopia and hyperopia in an agent.

To illustrate this in the context of a temptation to save, assume that v is linear so that temptation preferences do not care about intertemporal consumption smoothing, and also that commitment preferences desire some positive consumption in the current period. Then, for any rate of interest r such that

$$1 + r > \frac{1}{\gamma},$$

the agent is tempted to save his entire endowment. This holds even though $\gamma < \delta$. Note that the agent switches between being tempted to save and tempted to spend as $1 + r$ fluctuates around $\frac{1}{\gamma}$.

DSC agents are tempted to maximize temptation utility from immediate consumption. In order to produce a temptation to over-save, one would have to assume that temptation utility is maximized by reducing immediate consumption. Therefore, DSC preferences cannot account for hyperopic behavior, unless one makes the counter-intuitive assumption that, given fixed future consumption, less immediate consumption is more tempting.

6.6. Save More Tomorrow

The results of Benartzi and Thaler [7] suggest that people find it easier to make delayed commitments than immediate commitments. The authors introduce a saving-enhancement plan, called the 'Save More Tomorrow^{TM} (SMT) plan'. Subjects in a firm are given the opportunity to commit in advance to allocating a portion of their future salary increases towards a defined-contributions plan.[15] Opting for this plan implies a preference for delayed commitment since subjects would rather commit portions of future salaries (future budget sets) than portions of the next salary. The authors implemented the SMT plan in several firms, and found a significant demand for it. Across the implementations, the percentage of employees that opted for the plan ranged between 27% (216 of 816) and 78% (162 of 207).[16]

Benartzi and Thaler suggest several possible stories that would rationalize these results. The FT model provides an additional explanation: commitment requires self-control, and it is easier to exert self-control when making delayed commitments. To illustrate, consider x, y such that $x \subset y$ and

$$\{(c, x)\} \succ \{(c, x), (c, y)\} \sim \{(c, y)\}.$$

By Stationarity, $\{(c, x)\} \succ \{(c, y)\}$ implies $x \succ y$, and since $x \subset y$, it follows that the agent has a preference for commitment at y, in the sense of Definition 4.1. By Definition 4.2, this agent does not have a demand for commitment at y. Assuming

[15] Any withdrawals one makes from a defined-contributions plan before the age of $59\frac{1}{2}$ is at a cost. Thus, such plans provide a means of committing funds for retirement.

[16] A feature of the SMT plan that is problematic for our purposes is that the allocation decisions made by participants are not binding. Thus, SMT falls short of providing commitment. However, the authors point out that only a small proportion of subjects drop out of the plan, and possible reasons for this include inertia, procrastination, etc. If subjects are aware that such psychological factors would stop them from dropping out of the plan in the future, then participating in the plan does serve as a means of commitment.

$\gamma < \delta$, Theorem 5.3 implies the existence of a preference reversal (Definition 5.1): there exists t^* such that

$$\{(c,x)^{+t},(c,y)^{+t}\} \succ \{(c,y)^{+t}\} \text{ for all } t \geq t^*.$$

That is, if an FT agent has a preference for commitment at y but no demand for commitment at y, he nevertheless has a demand for delayed commitment at y – he chooses $(c,x)^{+t}$ over $(c,y)^{+t}$ for sufficiently large t.

A comment is in order. The motivation of the FT model was to explain why a market for commitment mechanisms may be absent. However, the model predicts that these agents would nevertheless demand delayed commitment, and this raises the question why a market for mechanisms providing delayed commitment appears to be absent. There are two possible answers. First, since the benefit of choosing delayed commitment lies in the future, the discounting of future utility may lead FT agents to not value delayed commitment significantly enough to attract a supply of such mechanisms. Second, uncertainty regarding factors such as future tastes, which we have abstracted from in this paper, may lead to an off-setting desire to retain flexibility.

7. Conclusion

While the literature on self-control problems has concentrated on the implications of immediate temptation, this paper explores the implications of temptation by future consumption. The model provides an explanation for why agents who are aware of their self-control problem may not take advantage of commitment opportunities: the possibility of indulging temptations in the future is itself a source of temptation, and strategies such as commitment require this temptation to be resisted. That is, commitment requires self-control.

An alternative explanation for why agents with self-control problems may not commit is provided by O'Donoghue and Rabin [27], who suggest that agents may not be fully aware of their self-control problems. They may not recognize that their future choices will deviate from what is optimal in their current view, and thus may find no reason to employ commitment devices.[17] Such a view suggests

[17] This explanation is most appealing when coupled with the hypothesis that the agent does not learn about his self-control problem even though he is continually surprised by the fact that he never sticks to his plan.

that agents can be made better-off by educating them about self-control problems. However, if agents are tempted to postpone commitment, then they may not seek commitment despite having full knowledge about their self-control problems. In such a case, efforts to improve agents' welfare through information and education may not be effective.

A. Appendix: Proof of Theorem 3.1

$\Longleftarrow$: Given a representation W, necessity of the axioms is straightforward to establish. Necessity of Indifference to Timing follows from the additive separability of U and V, and the linearity of W and $\overline{V}$. To show that W is unique and well defined, consider the mappings S and T, from the space of all continuous, real valued functions on Z (endowed with the sup norm) to itself, defined below:

$$\begin{aligned} S\overline{V}(z) &= \max_{\eta\in z}\int v(c)+\gamma\overline{V}(y)d\eta, \\ TW(z) &= \max_{\mu\in z}\{\int u(c)+\delta W(x)+v(c)+\gamma\overline{V}(x)d\mu-\max_{\eta\in z}\int v(c)+\gamma\overline{V}(y)d\eta\}. \end{aligned}$$

For any v and γ as defined in Theorem 3.1, Blackwell's Theorem (see Aliprantis and Border [5]) gives a well-defined and unique $\overline{V}$ that is a fixed point of S. Using this $\overline{V}$ in the definition of T and invoking Blackwell's Theorem again yields that for any u, v, δ and γ as defined in Theorem 3.1, there is a well-defined and unique W that is the fixed point of T.

$\Longrightarrow$: By Independence, Stationarity and Indifference to Timing, $\succsim$ satisfies the stronger version of Independence:

$$x \succ y, \alpha\in(0,1) \Longrightarrow \alpha x+(1-\alpha)z \succ \alpha y+(1-\alpha)z. \tag{A.1}$$

By GP [13, Theorem 1], $\succsim$ satisfies Order, Continuity, Set-Betweenness and (A.1) if and only if there exist linear and continuous functions $U, V : \Delta \longrightarrow \mathbb{R}$ and the function $W : Z \longrightarrow \mathbb{R}$ defined by

$$W(z)=\max_{\mu\in z}\{U(\mu)+V(\mu)-\max_{\eta\in z}V(\eta)\}, \tag{A.2}$$

that represents $\succsim$. Lemma A.1 establishes some facts about (A.2) that will be used throughout the Appendix. Define $\overline{V} : Z \longrightarrow \mathbb{R}$ and $\overline{U+V} : Z \longrightarrow \mathbb{R}$ by

$$\begin{aligned} \overline{V}(x) &= \max_{\eta\in x}V(\eta), \\ (\overline{U+V})(x) &= \max_{\eta\in x}U(\eta)+V(\eta). \end{aligned}$$

Lemma A.1. *For all x, y,*

(a) $x \succ x \cup y \iff \overline{V}(y) > \overline{V}(x)$ and $W(x) > W(y)$.

(b) $x \cup y \succ y \iff \overline{(U+V)}(x) > \overline{(U+V)}(y)$ and $W(x) > W(y)$.

(c) $x \succ x \cup y \succ y \iff \overline{(U+V)}(x) > \overline{(U+V)}(y)$ and $\overline{V}(y) > \overline{V}(x)$.

Proof. (a) $\Longrightarrow$: Set-Betweenness implies that $x \succ y$, and hence $W(x) > W(y)$. Suppose by way of contradiction, $\overline{V}(y) \leq \overline{V}(x)$. Then

$$\begin{aligned} W(x) &= \overline{(U+V)}(x) - \overline{V}(x) \\ W(x \cup y) &= \overline{(U+V)}(x \cup y) - \overline{V}(x). \end{aligned}$$

Because $x \subset x \cup y$, $\overline{(U+V)}(x \cup y) \geq \overline{(U+V)}(x)$, and so $W(x \cup y) \geq W(x)$, a contradiction.

$\Longleftarrow$: Let $\overline{V}(y) > \overline{V}(x)$ and $W(x) > W(y)$. There are two cases to consider. If $x \cup y \sim y$, then $W(x) > W(y)$ implies that $x \succ x \cup y$. If $x \cup y \succ y$, then

$$\begin{aligned} W(x \cup y) &= \overline{(U+V)}(x \cup y) - \overline{V}(y) \\ &> \overline{(U+V)}(y) - \overline{V}(y) = W(y), \end{aligned}$$

and so, $\overline{(U+V)}(x \cup y) > \overline{(U+V)}(y)$. It follows that $\overline{(U+V)}(x) = \overline{(U+V)}(x \cup y)$, and so,

$$\begin{aligned} W(x) &= \overline{(U+V)}(x) - \overline{V}(x) \\ &> \overline{(U+V)}(x \cup y) - \overline{V}(y) = W(x \cup y). \end{aligned}$$

That is, $x \succ x \cup y$.

(b) $\Longrightarrow$: Set-Betweenness implies that $x \succ y$, and hence $W(x) > W(y)$. Suppose by way of contradiction that $\overline{(U+V)}(x) \leq \overline{(U+V)}(y)$. Then $\overline{(U+V)}(y) = \overline{(U+V)}(x \cup y)$, and

$$\begin{aligned} W(x \cup y) &= \overline{(U+V)}(y) - \overline{V}(x \cup y) \\ W(y) &= \overline{(U+V)}(y) - \overline{V}(y). \end{aligned}$$

Because $y \subset x \cup y$, $\overline{V}(x \cup y) \geq \overline{V}(y)$, and so $W(y) \geq W(x \cup y)$, a contradiction.

$\Longleftarrow$: There are two cases to consider. First, $\overline{V}(x) \geq \overline{V}(y)$. Then

$$W(x) = \overline{(U+V)}(x) - \overline{V}(x) = W(x \cup y)$$

and since by hypothesis $W(x) > W(y)$, Set-Betweenness implies $x \cup y \succ y$. Second, $\overline{V}(x) < \overline{V}(y)$. Then

$$\begin{aligned} W(x \cup y) &= \overline{(U+V)}(x) - \overline{V}(y) \\ &> \overline{(U+V)}(y) - \overline{V}(y) = W(y), \end{aligned}$$

as desired.

(c) This follows from (a) and (b). ■

Indifference to Timing implies GP's axioms 5 and 7, respectively. Hence our axioms 1-6 imply GP's axioms 1-7. The proof of GP [14, Theorem 1] yields that $U(\cdot)$ can be written as

$$U(\mu) = \int_{C\times Z} (u(c) + \delta W(z))\, d\mu(c,z),$$

for some $u : C \longrightarrow \mathbb{R}$ and $\delta \in (0,1)$. We want to show that

$$V(\mu) = \int_{C\times Z} \big(v(c) + \gamma \overline{V}(y)\big)\, d\mu(c,z),$$

where $\overline{V}(x) = \max_{\mu\in x} \int_{C\times Z} v(c) + \gamma\overline{V}(y) d\mu(c,y)$. Consider two possibilities:

Case (1) V is constant or U is a positive affine transformations of V.

In either case, we can take $U' = (U + V)$ and $V' = 0$ and U', V' yield the representation with $v(\cdot) = 0$ and any γ. In particular, the representation holds for $0 < \gamma < 1$.

Case (2) V is not constant and U is not a positive affine transformations of V.

It is not possible for there to exist $\alpha \leq -1$ such that $V = \alpha U + \beta$, $\beta \in \mathbb{R}$, since that would contradict nondegeneracy. Therefore consider the case that $\alpha \in (-1, 0)$ or U is not an affine transformation of V. The remainder of the proof will establish the result in this case. Let $\Delta_s \subset \Delta$ represent the set of measures on $C \times Z$ with finite support.

Lemma A.2. *Under Case (2), there exists $\overline{\mu}, \underline{\mu} \in \Delta_s$ such that*

$$\{\overline{\mu}\} \succ \{\overline{\mu}, \underline{\mu}\} \succ \{\underline{\mu}\}.$$

Furthermore, for any finite $L \subset \Delta$, there exists $\alpha \in (0,1]$ such that for all $\nu \in L$,

$$\{\overline{\mu}\} \succ \{\overline{\mu}, \nu\alpha\underline{\mu}\} \succ \{\nu\alpha\underline{\mu}\}.$$

Proof. It is established in the proof of GP [14, Theorem 1] that under the conditions of Case (2), there exist $\overline{\nu}, \underline{\nu}$ such that $U(\overline{\nu}) + V(\overline{\nu}) - U(\underline{\nu}) - V(\underline{\nu}) > 0 > V(\overline{\nu}) - V(\underline{\nu})$. By Lemma A.1(c),

$$\{\overline{\nu}\} \succ \{\overline{\nu}, \underline{\nu}\} \succ \{\underline{\nu}\}.$$

Since Δ_s is dense in Δ, and U, V are continuous, there exist $\overline{\mu}, \underline{\mu} \in \Delta_s$ such that $U(\overline{\mu}) + V(\overline{\mu}) - U(\underline{\mu}) - V(\underline{\mu}) > 0 > V(\overline{\mu}) - V(\underline{\mu})$, and so,

$$\{\overline{\mu}\} \succ \{\overline{\mu}, \underline{\mu}\} \succ \{\underline{\mu}\}.$$

To prove the second part of the Lemma, take any finite $L \subset \Delta$. By continuity of $\succsim$, for every $\eta \in L$ there exists some $\alpha_\eta \in (0,1)$ such that for all $\alpha'_\eta \in (0, \alpha_\eta]$,

$$\{\overline{\mu}\} \succ \{\overline{\mu}, \eta\alpha'_\eta\underline{\mu}\} \succ \{\eta\alpha'_\eta\underline{\mu}\}.$$

Taking $\alpha = \min\{\alpha_\eta\}_{\eta \in L}$ establishes the result. ■

The next two Lemmas establish Separability of V.

Lemma A.3.

$$V(\frac{1}{2}(c,z) + \frac{1}{2}(c',z')) = V(\frac{1}{2}(c,z') + \frac{1}{2}(c',z)).$$

Proof. Take $\nu_1 = \frac{1}{2}(c,z) + \frac{1}{2}(c',z')$ and $\nu_2 = \frac{1}{2}(c,z') + \frac{1}{2}(c',z)$. By Lemma A.2, there is $\overline{\mu}, \underline{\mu}$ and α such that

$$\{\overline{\mu}\} \succ \{\overline{\mu}, \nu_1\alpha\underline{\mu}\} \succ \{\nu_1\alpha\underline{\mu}\} \text{ and } \{\overline{\mu}\} \succ \{\overline{\mu}, \nu_2\alpha\underline{\mu}\} \succ \{\nu_2\alpha\underline{\mu}\}.$$

Observe that all the above measures have finite support, the first marginals of $\nu_1\alpha\underline{\mu}$ and $\nu_2\alpha\underline{\mu}$ are the same, and the second marginals $(\nu_1\alpha\underline{\mu})^2$ and $(\nu_2\alpha\underline{\mu})^2$ satisfy $\varphi((\nu_1\alpha\underline{\mu})^2) = \varphi((\nu_2\alpha\underline{\mu})^2)$. Hence by Indifference to Timing,

$$\{\overline{\mu}, \nu_1\alpha\underline{\mu}\} \sim \{\overline{\mu}, \nu_2\alpha\underline{\mu}\}.$$

By the representation (A.2), $\{\overline{\mu}, \nu_1\alpha\underline{\mu}\} \sim \{\overline{\mu}, \nu_2\alpha\underline{\mu}\}$. But

$\{\overline{\mu}, \nu_1\alpha\underline{\mu}\} \sim \{\overline{\mu}, \nu_2\alpha\underline{\mu}\}$

$\Longleftrightarrow U(\overline{\mu}) + V(\overline{\mu}) - V(\alpha\nu + (1-\alpha)\mu') = U(\overline{\mu}) + V(\overline{\mu}) - V(\alpha\eta + (1-\alpha)\mu')$

$\Longleftrightarrow V(\nu_1\alpha\underline{\mu}) = V(\nu_2\alpha\underline{\mu})$

$\Longleftrightarrow \alpha V(\nu_1) + (1-\alpha)V(\underline{\mu}) = \alpha V(\nu_2) + (1-\alpha)V(\underline{\mu})$ by linearity of V

$\Longleftrightarrow V(\nu_1) = V(\nu_2)$.

That is, $V(\frac{1}{2}(c,z) + \frac{1}{2}(c',z')) = V(\frac{1}{2}(c,z') + \frac{1}{2}(c',z))$. ■

Lemma A.4. *There exists continuous functions $v : C \longrightarrow \mathbb{R}$, $\widehat{V} : Z \longrightarrow \mathbb{R}$ such that $\forall\mu \in \Delta$,*

$$V(\mu) = \int_{C \times Z} v(c) + \widehat{V}(x)d\mu$$

Proof. Since V is linear and continuous, there exists continuous $\overline{v} : C \times Z \longrightarrow \mathbb{R}$ such that $V(\mu) = \int \overline{v}(c,x)d\mu$ for all $\mu \in \Delta$. By the previous Lemma,

$$V(\frac{1}{2}(c,x) + \frac{1}{2}(\overline{c},\overline{x})) = V(\frac{1}{2}(c,\overline{x}) + \frac{1}{2}(\overline{c},x)).$$

Then

$V(\frac{1}{2}(c,x) + \frac{1}{2}(\overline{c},\overline{x})) = V(\frac{1}{2}(c,\overline{x}) + \frac{1}{2}(\overline{c},x))$

$\Longrightarrow V(c,x) + V(\overline{c},\overline{x}) = V(c,\overline{x}) + V(\overline{c},x)$
$\Longrightarrow \overline{v}(c,x) + \overline{v}(\overline{c},\overline{x}) = \overline{v}(c,\overline{x}) + \overline{v}(\overline{c},x)$
$\Longrightarrow \overline{v}(c,x) = \overline{v}(c,\overline{x}) - \overline{v}(\overline{c},\overline{x}) + \overline{v}(\overline{c},x).$
Define $v(c) \equiv \overline{v}(c,\overline{x}) - \overline{v}(\overline{c},\overline{x})$ and $\widehat{V}(x) \equiv \overline{v}(\overline{c},x)$. We can then write $\overline{v}(c,x) = v(c) + \widehat{V}(x)$. Therefore $V(\mu) = \int v(c) + \widehat{V}(x) d\mu$ for all $\mu \in \Delta$. ■

The next two Lemmas establish the linearity of $\widehat{V}$.

Lemma A.5.

$$V(\alpha(c,z) + (1-\alpha)(c,z')) = V((c, \alpha z + (1-\alpha)z')).$$

Proof. Take $\nu_1 = \alpha(c,z) + (1-\alpha)(c,z')$ and $\nu_2 = (c, \alpha z + (1-\alpha)z')$. By Lemma A.2, there exists $\overline{\mu}, \underline{\mu} \in \Delta_s$ and β such that

$$\{\overline{\mu}\} \succ \{\overline{\mu}, \nu_1\beta\underline{\mu}\} \succ \{\nu_1\beta\underline{\mu}\} \text{ and } \{\overline{\mu}\} \succ \{\overline{\mu}, \nu_2\beta\underline{\mu}\} \succ \{\nu_2\beta\underline{\mu}\}.$$

Observe that all the above measures have finite support, the first marginals of $\nu_1\beta\underline{\mu}$ and $\nu_2\beta\underline{\mu}$ are the same and the second marginals satisfy $\varphi(\left(\nu_1\beta\underline{\mu}\right)^2) = \varphi(\left(\nu_2\beta\underline{\mu}\right)^2)$. Hence by Indifference to Timing,

$$\{\overline{\mu}, \nu_1\alpha\underline{\mu}\} \sim \{\overline{\mu}, \nu_2\alpha\underline{\mu}\}.$$

Arguing as in Lemma A.3,

$$V(\nu_1) = V(\nu_2),$$

that is, $V(\alpha(c,z) + (1-\alpha)(c,z')) = V((c, \alpha z + (1-\alpha)z'))$. ■

Lemma A.6. $\widehat{V}$ *is linear.*

Proof. By the previous lemma, $V((c, \alpha x + (1-\alpha)y)) = V(\alpha(c,x) + (1-\alpha)(c,y))$. But,
$V((c, \alpha x + (1-\alpha)y)) = V(\alpha(c,x) + (1-\alpha)(c,y))$
$\Longrightarrow V((c, \alpha x + (1-\alpha)y)) = \alpha V(c,x) + (1-\alpha)V(c,y)$ by linearity of V
$\Longrightarrow v(c) + \widehat{V}(\alpha x + (1-\alpha)y) = \alpha[v(c) + \widehat{V}(x)] + (1-\alpha)[v(c) + \widehat{V}(y)]$ by Lemma A.4
$\Longrightarrow v(c) + \widehat{V}(\alpha x + (1-\alpha)y) = v(c) + \alpha\widehat{V}(x) + (1-\alpha)\widehat{V}(y)$
$\Longrightarrow \widehat{V}(\alpha x + (1-\alpha)y) = \alpha\widehat{V}(x) + (1-\alpha)\widehat{V}(y)$
$\Longrightarrow \widehat{V}$ is linear. ■

Recall the function $\overline{V} : Z \longrightarrow \mathbb{R}$ defined by

$$\overline{V}(x) = \max_{\eta \in x} V(\eta).$$

The next two Lemmas establish linearity and continuity of $\overline{V}$.

Lemma A.7. $\overline{V}$ *is linear.*

Proof. The linearity of V and definition of the mixture $\alpha x+(1-\alpha)y$ implies $\overline{V}(\alpha x+(1-\alpha)y) = \alpha\overline{V}(x) + (1-\alpha)\overline{V}(y)$. ■

Lemma A.8. $\overline{V}$ *is continuous.*

Proof. Since $V : \Delta \longrightarrow \mathbb{R}$ is continuous, and the correspondence $\Phi : Z \rightsquigarrow \Delta$ defined by $\Phi(x) = x$ is compact-valued and continuous, the Maximum Theorem delivers continuous $\overline{V}(x) = \max_{\eta\in\Phi(x)} V(\eta)$. ■

The next two Lemmas establish the ordinal equivalence between $\overline{V}$ and $\widehat{V}$.

Lemma A.9. *If $x \succ y$, then,*

$$\overline{V}(y) > \overline{V}(x) \Longleftrightarrow \widehat{V}(y) > \widehat{V}(x).$$

Proof. If $x \succ y$, then $W(x) > W(y)$, and Stationarity implies $W(c,x) > W(c,y)$. Then by Lemma A.1(a), $x \succ y$ implies

$$\begin{aligned} x &\succ x \cup y \Longleftrightarrow \overline{V}(y) > \overline{V}(x), \\ \{(c,x)\} &\succ \{(c,x),(c,y)\} \Longleftrightarrow V(c,y) > V(c,x). \end{aligned}$$

By Temptation Stationarity,

$$x \succ x \cup y \Longleftrightarrow \{(c,x)\} \succ \{(c,x),(c,y)\},$$

and so $\overline{V}(y) > \overline{V}(x) \Longleftrightarrow V(c,y) > V(c,x)$. But, by Lemma A.4, $V(c,y) > V(c,x) \Longleftrightarrow \widehat{V}(x) > \widehat{V}(y)$ and we are done. ■

Lemma A.10. *For all x, y,*

$$\overline{V}(y) > \overline{V}(x) \Longleftrightarrow \widehat{V}(y) > \widehat{V}(x).$$

Proof. We show that the conclusion of Lemma A.9 also holds when its hypothesis is negated. So suppose $y \succsim x$. To establish '$\Longleftarrow$', suppose $\overline{V}(x) \geq \overline{V}(y)$. By Lemma A.2, there exists $w, z \in Z$ such that $w \succ z$ and $\overline{V}(w) < \overline{V}(z)$.[18] Linearity of W and V implies $y\alpha w \succ x\alpha z$ and $\overline{V}(x\alpha z) > \overline{V}(y\alpha w)$ for all $\alpha \in (0,1)$. Lemma A.9 implies $\widehat{V}(x\alpha z) > \widehat{V}(y\alpha w)$ for all $\alpha \in (0,1)$ and continuity of $\widehat{V}$ implies $\widehat{V}(x) \geq \widehat{V}(y)$, as desired.

To establish '$\Longrightarrow$', let $\widehat{V}(x) \geq \widehat{V}(y)$. As above, there is $w, z \in Z$ such that $w \succ z$ and $\overline{V}(w) < \overline{V}(z)$. By Lemma A.9, $\widehat{V}(w) < \widehat{V}(z)$. Therefore, $y\alpha w \succ x\alpha z$ and $\widehat{V}(x\alpha z) > \widehat{V}(y\alpha w)$ for all $\alpha \in (0,1)$. Lemma A.9 implies $\overline{V}(x\alpha z) > \overline{V}(y\alpha w)$ for all $\alpha \in (0,1)$ and continuity of $\overline{V}$ implies $\overline{V}(x) \geq \overline{V}(y)$, as desired. ■

---

[18] Take $w = \{\overline{\mu}\}$ and $z = \{\overline{\mu}, \underline{\mu}\}$.

Lemma A.11. *There exists $\gamma > 0$ and $\theta \in \mathbb{R}$ such that $\widehat{V}(x) = \gamma \overline{V}(x) + \theta$ for all $x \in Z$.*

Proof. By Lemma A.10, $\widehat{V}$ and $\overline{V}$ are ordinally equivalent. By Lemmas A.4, A.6, A.7 and A.8, they are also continuous and linear. Moreover, under the conditions of Case 2, both are nonconstant.[19] It follows that $\widehat{V}$ and $\overline{V}$ are cardinally equivalent (the proof is analogous to GP [14, Lemma 9, Step 2]). ■

Lemma A.12. $\gamma < 1$.

Proof. Define $z^c = \{(c, z^c)\}$. Since V is nonconstant (Case 2), we can find η such that $V(\eta) \neq V(z^c)$, where we have identified z^c with (c, z^c). Let $y^1 = \{(c, \eta)\}$ and define y^n inductively as $y^n = \{(c, y^{n-1})\}$. Note that $y^n \longrightarrow z^c$. We shall identify y^n with (c, y^{n-1}) below. By Lemma A.2 there exists α such that

$$\{\overline{\mu}\} \succ \{\overline{\mu}, \underline{\mu}\alpha z^c\} \succ \{\overline{\mu}\alpha z^c\}.$$

Since $y^n \longrightarrow z^c$ implies $\{\overline{\mu}, \underline{\mu}\alpha y^n\} \longrightarrow \{\overline{\mu}, \underline{\mu}\alpha z^c\}$ and $\{\underline{\mu}\alpha y^n\} \longrightarrow \{\underline{\mu}\alpha z^c\}$, Continuity of W implies $W(\{\overline{\mu}, \underline{\mu}\alpha y^n\}) \longrightarrow W(\{\overline{\mu}, \underline{\mu}\alpha z^c\})$ and $W(\{\underline{\mu}\alpha y^n\}) \longrightarrow W(\{\underline{\mu}\alpha z^c\})$. Then, there exists N^* such that for all $n \geq N^*$,

$$\{\overline{\mu}\} \succ \{\overline{\mu}, \underline{\mu}\alpha y^n\} \succ \{\overline{\mu}\alpha y^n\}.$$

Without loss of generality, $N^* = 1$. But then,

$$\begin{aligned}
&W(\{\overline{\mu}, \underline{\mu}\alpha y^n\}) - W(\{\overline{\mu}, \underline{\mu}\alpha z^c\}) \longrightarrow 0 \\
&\Longrightarrow V(\underline{\mu}\alpha y^n) - V(\underline{\mu}\alpha z^c) \longrightarrow 0 \qquad \text{by representation (A.2),} \\
&\Longrightarrow V(y^n) - V(z^c) \longrightarrow 0 \\
&\Longrightarrow \gamma^n[V(\eta) - V(z^c)] \longrightarrow 0.
\end{aligned}$$

Since $V(\eta) \neq V(z^c)$ by construction, it follows that $\gamma < 1$. ■

Lemmas A.4 and A.11 establish that $V(\mu) = \int (v(c) + \gamma \overline{V}(x) + \theta) d\mu$. By GP [13, Theorem 4], we can write $V(\mu) = \int v(c) + \gamma \overline{V}(x) d\mu$.[20] It is also established that $0 < \gamma < 1$. Hence, we are done.

B. Appendix: Proof of Theorem 3.2

First establish that for any (U, V) representing the preference $\succsim$, U and V are not constant, and U is not an affine transformation of V. By hypothesis, $\succsim$ exhibits a preference for commitment, and so, by Lemma A.1(a) it is clear that U and V are not constant and that U cannot be a positive affine transformation of V. Suppose by way of contradiction that U is a negative affine

[19] To see that $\widehat{V}$ is nonconstant, note that by Temptation Stationarity, $\{\overline{\mu}\} \succ \{\overline{\mu}, \underline{\mu}\} \Longleftrightarrow \{(c, \overline{\mu})\} \succ \{(c, \overline{\mu}), (c, \underline{\mu})\}$, where $\overline{\mu}, \underline{\mu}$ are as in Lemma A.2. By Lemma A.1(a), $V(c, \overline{\mu}) \neq V(c, \underline{\mu})$, which implies that $\widehat{V}$ is nonconstant.

[20] To see that U is not an affine transformation of V, argue as in Appendix B.

transformation of V. Then, $U(c,x) = -\alpha V(c,x) + \beta$ for some $\alpha > 0$, and so, by the functional forms of U and V,

$$W(x) = -\alpha \overline{V}(x) + \xi(c), \tag{B.1}$$

where $\xi(c)$ is some function of c. Observe that by definition of $\overline{V}$, for any x, y,

$$\overline{V}(x) \geq \overline{V}(y) \Longrightarrow \overline{V}(x) = \overline{V}(x \cup y) \geq \overline{V}(y). \tag{B.2}$$

But by Lemma A.2, there is w, z such that

$$W(w) > W(w \cup z) > W(z).$$

It follows by (B.1) that

$$\overline{V}(z) > \overline{V}(w) \text{ and } \overline{V}(z) > \overline{V}(z \cup w) > \overline{V}(w),$$

contradicting (B.2). Hence, for any (U,V) representing $\succsim$, U is not an affine transformation of V.

We prove Theorem 3.2 by exploiting GP [13, Thm 4], whose hypothesis holds given the above observation. Arguing as in the proof of GP [14, Thm 2] yields that $(u', v', \delta', \gamma')$ represents $\succsim$ if and only if $\delta = \delta'$ and there exist $\alpha > 0, \beta_u, \beta_v \in \mathbb{R}$ such that $u' = \alpha u + \beta_u$ and $V' = \alpha V + \beta_v$. We show that $\gamma = \gamma'$ and $v' = \alpha v + (1-\gamma)\beta_v$ if and only if $V' = \alpha V + \beta_v$. Begin by noting that, by hypothesis, there is a preference for commitment at some x and hence V is nonconstant. Therefore v is nonconstant as well. Also note that $V' = \alpha V + \beta_v$ if and only if $\overline{V}' = \alpha \overline{V} + \beta_v$. Now, observe that,

$$\begin{aligned}
& V'(c,x) = \alpha V(c,x) + \beta_v \\
& \Leftrightarrow v'(c) + \gamma' \overline{V}'(x) = \alpha[v(c) + \gamma \overline{V}(x)] + \beta_v \\
& \Leftrightarrow v'(c) + \gamma' \overline{V}'(x) = \alpha v(c) + (1-\gamma)\beta_v + \gamma[\alpha \overline{V}(x) + \beta_v] \\
& \Leftrightarrow v'(c) + \gamma' \overline{V}'(x) = \alpha v(c) + (1-\gamma)\beta_v + \gamma \overline{V}'(x) \\
& \Leftrightarrow v'(c) - [\alpha v(c) + (1-\gamma)\beta_v] = (\gamma - \gamma')\overline{V}'(x) \\
& \Leftrightarrow \gamma = \gamma' \text{ and } v' = \alpha v + (1-\gamma)\beta_v.
\end{aligned}$$

The last equivalence holds because v is nonconstant.

C. Appendix: Proof of Theorem 5.3

First, two preliminary lemmas.

Lemma C.1. *(a) $x \subset y \Longrightarrow \overline{V}(y) \geq \overline{V}(x)$ and $W(y) + \overline{V}(y) \geq W(x) + \overline{V}(x)$.*

(b) $x \subset y$ and $y \succsim x \Longrightarrow \{(c,x),(c,y)\} \sim \{(c,y)\}$.

Proof. (a) If $x \subset y$, the definition of $\overline{V}$ implies

$$\overline{V}(y) \geq \overline{V}(x).$$

To see $W(y) + \overline{V}(y) \geq W(x) + \overline{V}(x)$, note that $x \subset y$ implies

$$\max_{\mu \in y}\{U + V\} \geq \max_{\mu \in x}\{U + V\}. \tag{C.1}$$

The result then follows from the fact that for any $z \in Z$,

$$W(z) + \overline{V}(z) = \max_{\mu \in z}\{U + V\} - \overline{V}(z) + \overline{V}(z) = \max_{\mu \in z}\{U + V\}.$$

(b) By Stationarity and Set-Betweenness, if $y \sim x$ then $\{(c, x), (c, y)\} \sim \{(c, y)\}$. Next, if $y \succ x$, then by Stationarity, $\{(c, y)\} \succ \{(c, x)\}$ and so $U(c, y) > U(c, x)$. Furthermore, since $x \subset y$, $\overline{V}(y) \geq \overline{V}(x)$ and so $V(c, y) \geq V(c, x)$. But then by Lemma A.1(a), $\{(c, y)\} \precsim \{(c, x), (c, y)\}$. Since $\{(c, y)\} \succ \{(c, x)\}$, it follows from Set-Betweenness that $\{(c, x), (c, y)\} \sim \{(c, y)\}$. ■

Lemma C.2. *If $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ and $\{(c, \mu)\} \succ \{(c, \mu), (c, \eta)\} \succ \{(c, \eta)\}$, then*

$$\{(c, \mu), (c, \eta)\} \succsim \{(c, \{\mu, \eta\})\} \iff \gamma \leq \delta.$$

Proof. If the hypothesis holds, then

$\{(c, \mu), (c, \eta)\} \succsim \{(c, \{\mu, \eta\})\}$
$\iff W(\{(c, \mu), (c, \eta)\}) \geq W(\{(c, \{\mu, \eta\})$
$\iff u(c) + \delta W(\mu) + v(c) + \gamma \overline{V}(\mu) - v(c) - \gamma \overline{V}(\eta) \geq u(c) + \delta W(\{\mu, \eta\})$
$\iff u(c) + \delta U(\mu) + \gamma[V(\mu) - V(\eta)] \geq u(c) + \delta(U(\mu) + V(\mu) - V(\eta))$
$\iff u(c) + \delta U(\mu) + \gamma[V(\mu) - V(\eta)] \geq u(c) + \delta U(\mu) + \delta[V(\mu) - V(\eta)]$
$\iff \gamma[V(\mu) - V(\eta)] \geq \delta[V(\mu) - V(\eta)]$
$\iff \gamma \leq \delta$ since $V(\mu) - V(\eta) < 0$ by $\{\mu\} \succ \{\mu, \eta\}$ and Lemma A.1(a). ■

Proof of $(a) \iff (b)$:

$\Longleftarrow$: We prove the contrapositive, that is, $\gamma \geq \delta$ implies $\{(c, x), (c, y)\} \precsim \{(c, y)\}$ for all x, y such that $x \subset y$. By Lemma C.1(b), it suffices to prove that $\gamma \geq \delta$ implies

$$\{(c, x), (c, y)\} \precsim \{(c, y)\} \text{ for all } x, y \text{ such that } x \subset y \text{ and } x \succ y.$$

So let $\gamma \geq \delta$ and take any x, y such that $x \subset y$ and $x \succ y$. By Lemma C.1(a),

$$\begin{aligned} W(y) + \overline{V}(y) &\geq W(x) + \overline{V}(x), \\ \text{and } \overline{V}(y) &\geq \overline{V}(x). \end{aligned}$$

Since $\frac{\gamma}{\delta} \geq 1$, it follows that

$$W(y) + \frac{\gamma}{\delta}\overline{V}(y) \geq W(x) + \frac{\gamma}{\delta}\overline{V}(x),$$

and so

$$\delta W(y) + \gamma\overline{V}(y) \geq \delta W(x) + \gamma\overline{V}(x).$$

Adding $u(c) + v(c)$ to both sides gives $U(c, y) + V(c, y) \geq U(c, x) + V(c, x)$. By Stationarity, $x \succ y$ implies $U(c, x) > U(c, y)$, and so $V(c, y) > V(c, x)$. It follows from Lemma A.1(b) that

$$\{(c, x), (c, y)\} \precsim \{(c, y)\},$$

as desired.

$\Longrightarrow$: Before proving this, note that by hypothesis there is x, y such that $x \subset y$ and $x \succ y$, and that $x \succ x \cup y$ since $y = x \cup y$. By Lemma A.1(a), this implies that V is nonconstant and U is not a positive affine transformation of V. Hence, by Lemma A.2, there is $\mu, \eta \in \Delta$ such that $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$.

Now prove the contrapositive: Suppose that $\{(c, x), (c, y)\} \succsim \{(c, y)\}$ for all x, y such that $x \subset y$. Then $\{\mu\} \subset \{\mu, \eta\}$ and $\{\mu\} \succ \{\mu, \eta\}$ implies

$$\{(c, \mu), (c, \{\mu, \eta\})\} \sim \{(c, \{\mu, \eta\})\}. \tag{C.2}$$

By Stationarity, $\{\mu\} \succ \{\mu, \eta\}$ implies $\{(c, \mu)\} \succ \{(c, \{\mu, \eta\})$. Then, by Set-Betweenness and (C.2),

$$\{(c, \mu)\} \succ \{(c, \mu), (c, \{\mu, \eta\})\} \sim \{(c, \{\mu, \eta\})\}.$$

It follows from Lemma A.1 that $U(c, \{\mu, \eta\}) + V(c, \{\mu, \eta\}) \geq U(c, \mu) + V(c, \mu)$ and $V(c, \{\mu, \eta\}) > V(c, \mu)$. But,

$$\begin{aligned}
&U(c, \{\mu, \eta\}) + V(c, \{\mu, \eta\}) \geq U(c, \mu) + V(c, \mu)\\
&\Longrightarrow \delta W(\{\mu, \eta\}) + \gamma\overline{V}(\{\mu, \eta\}) \geq \delta W(\{\mu\}) + \gamma\overline{V}(\{\mu\})\\
&\Longrightarrow W(\{\mu, \eta\}) + \tfrac{\gamma}{\delta}\overline{V}(\{\mu, \eta\}) \geq W(\{\mu\}) + \tfrac{\gamma}{\delta}\overline{V}(\{\mu\})\\
&\Longrightarrow U(\mu) + V(\mu) - V(\eta) + \tfrac{\gamma}{\delta}V(\eta) \geq U(\mu) + \tfrac{\gamma}{\delta}V(\mu)\\
&\Longrightarrow \tfrac{\gamma}{\delta}\left(V(\eta) - V(\mu)\right) \geq V(\eta) - V(\mu).
\end{aligned}$$

Furthermore, $V(c, \{\mu, \eta\}) > V(c, \mu)$ implies that $\overline{V}(\{\mu, \eta\}) > \overline{V}(\mu)$, and thus $V(\eta) - V(\mu) > 0$. Conclude that $\frac{\gamma}{\delta} \geq 1$.

Proof of $(a) \Longleftrightarrow (c)$:

$\Longrightarrow$: Let $\gamma < \delta$. By Lemma A.1, the hypothesis $\{\mu\} \succ \{\mu, \eta\} \sim \{\eta\}$ implies,

$$V(\eta) - V(\mu) > 0 \tag{C.3}$$

$$U(\mu) - U(\eta) > 0 \tag{C.4}$$

$$\frac{U(\mu) - U(\eta)}{V(\eta) - V(\mu)} \leq 1 \tag{C.5}$$

By repeated application of Stationarity, $\{\mu\} \succ \{\eta\}$ implies $\{\mu^{+t}\} \succ \{\eta^{+t}\}$. Then by Lemma A.1(b) and the structure on $U+V$, we have $\{\mu^{+t}, \eta^{+t}\} \succ \{\eta^{+t}\}$ if and only if $U(\mu^{+t})+V(\mu^{+t}) > U(\eta^{+t}) + V(\eta^{+t})$ if and only if

$$(\frac{\gamma}{\delta})^t < \frac{U(\mu) - U(\eta)}{V(\eta) - V(\mu)}.$$

By (C.3), (C.4) and (C.5), we have $0 < \frac{U(\mu)-U(\eta)}{V(\eta)-V(\mu)} \leq 1$. Define $\phi(t) \equiv (\frac{\gamma}{\delta})^t$ and note that ϕ is a continuous, monotone decreasing function with $\phi(0) = 1$ and $\phi(\infty) = 0$. Hence,

$$\phi(\infty) < \frac{U(\mu) - U(\eta)}{V(\eta) - V(\mu)} \leq \phi(0).$$

By the the monotonicity of ϕ and the Intermediate Value Theorem, there exists a unique $t' > 0$, given by

$$t' = \frac{\ln \frac{U(\mu)-U(\eta)}{V(\eta)-V(\mu)}}{\ln \frac{\gamma}{\delta}},$$

such that $\phi(t') = \frac{U(\mu)-U(\eta)}{V(\eta)-V(\mu)}$. Let t^* be the smallest integer larger than t'. Then for all $t < t^*$, $\{\mu^{+t}, \eta^{+t}\} \sim \{\eta^{+t}\}$, and for all $t \geq t^*$, $\{\mu^{+t}, \eta^{+t}\} \succ \{\eta^{+t}\}$.

$\Longleftarrow$: Suppose $\{\mu\} \succ \{\mu, \eta\} \sim \{\eta\}$. Recall from the above that $\{\mu^{+t}, \eta^{+t}\} \succ \{\eta^{+t}\}$ if and only if

$$U(\mu) - U(\eta) > (\frac{\gamma}{\delta})^t (V(\eta) - V(\mu)), \tag{C.6}$$

and that for $t = 0$, $U(\mu) - U(\eta) < V(\eta) - V(\mu)$. Suppose by way of contradiction that $\gamma \geq \delta$. It follows that (C.6) never holds for any t, that is, there is no t^* such that for all $t > t^*$, $\{\mu^{+t}, \eta^{+t}\} \succ \{\eta^{+t}\}$, contradicting the hypothesis that $\succsim$ exhibits preference reversals. Hence $\gamma < \delta$.

Proof of $(a) \Longleftrightarrow (d)$:

$\Longleftarrow$: By hypothesis there exists a preference for commitment, and so we are in Case 2 (see proof of 3.1). Then by Lemma A.2, there exists μ, η such that $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$. We show that Preference for Early Choice implies

$$\{(c, \mu)\} \succ \{(c, \mu), (c, \eta)\} \succ \{(c, \eta)\}.$$

The result then follows from Lemma C.2.

By Stationarity and Preference for Early Choice, $\{\mu, \eta\} \succ \{\eta\}$ implies

$$\{(c, \mu), (c, \eta)\} \succsim (c, \{\mu, \eta\}) \succ \{(c, \eta)\},$$

that is, $\{(c, \mu), (c, \eta)\} \succ \{(c, \eta)\}$. Also, by Temptation Stationarity, $\{\mu\} \succ \{\mu, \eta\}$ implies

$$\{(c, \mu)\} \succ \{(c, \mu), (c, \eta)\}.$$

Therefore, $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ implies

$$\{(c, \mu)\} \succ \{(c, \mu), (c, \eta)\} \succ \{(c, \eta)\}.$$

$\Longrightarrow$: Take $\mu, \eta \in \Delta$ such that $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$. We show that $\gamma \leq \delta$ implies

$$\{(c, \mu)\} \succ \{(c, \mu), (c, \eta)\} \succ \{(c, \eta)\},$$

and then the result follows from Lemma C.2.

By Temptation Stationarity, $\{\mu\} \succ \{\mu, \eta\}$ implies

$$\{(c, \mu)\} \succ \{(c, \mu), (c, \eta)\}.$$

To show that $\{(c, \mu), (c, \eta)\} \succ \{(c, \eta)\}$, by Lemma A.1(b) it suffices to show that $U(c, \mu) + V(c, \mu) > U(c, \eta) + V(c, \eta)$.[21] By hypothesis, $\{\mu, \eta\} \succ \{\eta\}$, and so by Lemma A.1(b), $U(\mu) + V(\mu) > U(\eta) + V(\eta)$. Also, $\{\mu\} \succ \{\mu, \eta\}$ implies (by Lemma A.1(a)) that $V(\mu) < V(\eta)$. Observe that $U(\mu) + V(\mu) > U(\eta) + V(\eta)$ implies

$$U(\mu) - U(\eta) > V(\eta) - V(\mu),$$

and $V(\eta) > V(\mu)$ and $\gamma \leq \delta$ implies

$$V(\eta) - V(\mu) \geq \frac{\gamma}{\delta} \left(V(\eta) - V(\mu) \right).$$

Therefore, $U(\mu) - U(\eta) > \frac{\gamma}{\delta} (V(\eta) - V(\mu))$. But

$$\begin{aligned}
& U(\mu) - U(\eta) > \tfrac{\gamma}{\delta} (V(\eta) - V(\mu)) \\
& \Longrightarrow \delta U(\mu) - \delta U(\eta) > \gamma V(\eta) - \gamma V(\mu) \\
& \Longrightarrow \delta U(\mu) + \gamma V(\eta) > \delta U(\eta) + \gamma V(\mu) \\
& \Longrightarrow u(c) + \delta U(\mu) + v(c) + \gamma V(\eta) > u(c) + \delta U(\eta) + v(c) + \gamma V(\mu) \\
& \Longrightarrow U(c, \mu) + V(c, \mu) > U(c, \eta) + V(c, \eta),
\end{aligned}$$

and hence by Lemma A.1(b),

$$\{(c, \mu), (c, \eta)\} \succ \{(c, \eta)\},$$

as was to be shown.

D. Appendix: Derivation for Section 5.3

A choice between continuation menus x and y is a choice from

$$z = \{(c, x), (c, y)\},$$

[21] It suffices because by Set-Betweenness $\{(c, \mu)\} \succ \{(c, \mu), (c, \eta)\}$ implies $\{(c, \mu)\} \succ \{(c, \eta)\}$, and so $U(c, \mu) > U(c, \eta)$ already holds.

and choice from z is determined by $\max_{\mu \in z}\{U(\mu)+V(\mu)\}$. By the functional forms for U and V, choice of continuation menu is determined by

$$\max_{\{x,y\}}\{W(\cdot)+\frac{\gamma}{\delta}\overline{V}(\cdot)\}.$$

Therefore, the agent chooses to commit if $\frac{W(x)-W(y)}{\overline{V}(y)-\overline{V}(x)} > \frac{\gamma}{\delta}$. Note that by Lemma A.1(a), the hypotheses $x \subset y$ and $x \succ y$ imply $\frac{W(x)-W(y)}{\overline{V}(y)-\overline{V}(x)} > 0$. This explains the first two rows of the table.

To see why the agent never commits when $\gamma \geq \delta$, note that for any menu w,
$W(w)+\frac{\gamma}{\delta}\overline{V}(w)$
$= \max_{\mu \in w}\{U+V\} - \max_{\eta \in w} V + \frac{\gamma}{\delta}\overline{V}(w)$
$= \max_{\mu \in w}\{U+V\} + k \max_{\eta \in w} V$, where $k = \frac{\gamma}{\delta} - 1 > 0$.
Hence, $x \subset y$ implies that $\frac{W(x)-W(y)}{\overline{V}(y)-\overline{V}(x)} \leq \frac{\gamma}{\delta}$.

References

[1] Ainslie, G. (1992), *Picoeconomics*, Cambridge University Press.

[2] Ainslie, G. and V. Haendel (1983): 'The Motives of the Will', in Gotteheil et al (Eds), *Etiologic Aspects of Alcohol and Drug Abuse* , Springfield, Il: Thomas.

[3] Akerlof, G. (1991), "Procrastination and Obedience," *American Economic Review*, Papers and Proceedings, 81(2), pp 237-41.

[4] Akerlof, G., W. Gale and R. Hall (1998), "Comments on 'Self-control and Saving for Retirement'," *Brookings Papers on Economic Activity*, pp 196-207.

[5] Aliprantis, C., and K. Border (1994), *Infinite Dimensional Analysis: a Hitchhiker's Guide*, 2nd Edition, Springer Verlag.

[6] Becker, G., and K. Murphy (1988), "A Theory of Rational Addiction," *Journal of Political Economy* 96(a), pp 675-700.

[7] Benartzi, S., and R. Thaler (2004), "Save More Tomorrow: Using Behavioral Economics to Increase Employee Savings," *Journal of Political Economy* 112(1), pp 164-187.

[8] Brown, A., and C. Pearcy (1995), *An Introduction to Analysis*, Springer, New York.

[9] Browning M., and T. Crossley (2001), "The Life-Cycle Model of Consumption and Saving," *Journal of Economic Perspectives* 15(3).

[10] Elster, J. (2000), *Ulysses Unbound*, Cambridge University Press, Cambridge, UK.

[11] Fuller, R., and H. Roth (1979), "Disulfiram for the Treatment of Alcoholism," *Annals of Internal Medicine* 90, pp 901-904.

[12] Goldstein, A. (2001), *Addiction: From Biology to Drug policy*, Second Edition, New York: Oxford University Press.

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

[14] Gul, F., and W. Pesendorfer (2004), "Self-Control and the Theory of Consumption," *Econometrica* 72, 119-158.

[15] Gul, F., and W. Pesendorfer (2004), "Harmful Addiction," forthcoming in *Review of Economic Studies.*

[16] Kirby, K., and R. Hernnstein (1995), "Preference Reversals due to Myopic Discounting of Delayed Reward," *Psychological Science,* pp 83-89.

[17] Kivetz, R. and I. Simonson (2002), "Self-Control for the Righteous: Toward a Theory of Precommitment to Indulgence," *Journal of Consumer Research* 29, pp 199-217.

[18] Kocherlakota, N. (2001), "Looking for Evidence of Time-inconsistent Preferences in Asset Market Data," *Fed Reserve Bank of Minneapolis Quarterly Review* 25(3), pp 13-24.

[19] Kreps, D. (1998), *Notes on the Theory of Choice*, Westview.

[20] Krussel, P., B. Kuruscu and A. Smith, Jr. (2002), "Time Orientation and Asset Prices," *Journal of Monetary Economics* 49(1), pp 107-135.

[21] Laibson, D. (1997), "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics* 112, pp 443-77.

[22] Laibson, D., A. Repetto and J. Tobacman (1998), "Self-control and Saving for Retirement," *Brookings Papers on Economic Activity* 1, pp 91-196.

[23] Lowenstein, G. (1996), "Out of Control: Visceral Influences on Behavior," *Organizational Behavior and Human Decision Processes* 65, pp 272-292.

[24] Lowenstein, G. and D. Schkade (1999): "Wouldn't it be Nice? Predicting Future Feelings," in D. Kahneman, E. Diener and N. Schwarz (eds.), *Well-Being: The Foundations of Hedonic Psychology*, New York, NY: Russell Sage Foundation Press.

[25] Lowenstein, G. , T. O'Donoghue and M. Rabin (2003), "Projection Bias in Predicting Future Utility," *Quarterly Journal of Economics* 118(4), pp. 1209-1248.

[26] Mazur, J. E. (1987): 'An Adjustment Procedure for Studying Delayed Reinforcement' in M. L. Commons, J. E. Mazur, J. A. Nevin and H. Rachlin (Eds), Quantitative Analyses of Behavior: Vol. 5. The Effect of Delay and Intervening Events on Reinforcement Value, pp 55-73, Hillsdale, NJ: Erlbaum.

[27] O'Donoghue, T., and M. Rabin (1999), "Doing it Now or Later," *American Economic Review* 89, pp 103-24.

[28] O'Donoghue, T., and M. Rabin (2001), "Choice and Procrastination," *Quarterly Journal of Economics* 116, pp 121-60.

[29] Parthasarathy, K. (1970), *Probability Measures on Metric Spaces*, New York. Academic Press.

[30] Poterba, J., Venti, S., and D. Wise (2001), "The Transition to Personal Accounts and Increasing Retirement Wealth: Macro and Micro Evidence," NBER Working Paper No. 8610.

[31] Read, D. and B. van Leeuwen (1998), "Predicting Hunger: The Effects of Appetite and Delay on Choice," *Organizational Behavior and Human Decision Processes* 76(2), pp 189-205.

[32] Schelling, T.C. (1992), "Self-control" in G. Lowenstein and J. Elster (eds.), *Choice Over Time*, New York: Russell Sage, pp. 167-76.

[33] Strotz, R. (1955), "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies* XXIII, pp 165-180.

[34] Trope, Y., and N. Liberman (2000), "Temporal Construal and Time-dependent changes in Preference," *Journal of Personality and Social Psychology* 79, pp 876-889.

[35] Venti, S., and D. Wise (1987), "Have IRAs Increased U.S. Savings? Evidence from Consumer Expenditure Surveys," NBER Working Paper No. 2217.