Review of Economic Studies (2005) **72**, 429–448 0034-6527/05/00180429$02.00

The Revealed Preference Theory of Changing Tastes

FARUK GUL and WOLFGANG PESENDORFER
Princeton University

First version received June 2002; *final version accepted February* 2004 (*Eds.*)

We analyse preferences over finite decision problems in order to model decision-makers with "changing tastes". We provide conditions on these preferences that identify the Strotz model of consistent planning. Building on an example given by Peleg and Yaari (1973), we show that for problems with infinitely many choices, Strotz's representation of preferences may not be well defined. For that case, we propose a well-defined approximation which is empirically indistinguishable from the Strotz preference that is being approximated.

1. INTRODUCTION

In the canonical example of time-inconsistent behaviour an agent has to choose between a smaller period 1 reward and a larger period 2 reward. If the decision is made in period 1, the agent chooses the smaller period 1 reward. If the decision is made in period 0, the agent chooses the larger period 2 reward.

To study the behaviour described above, Strotz (1955) proposes a model of changing tastes. The agent has a distinct utility function for each period. If the period 0 utility function differs from the period 1 utility function the agent may make different choices depending on when he/she chooses. The behaviour in the example above can result if the period 0 utility function is more patient than the period 1 utility function with respect to intertemporal trade-offs between periods 1 and 2.

The utility functions in Strotz's model describe how the agent would choose among consumption paths under the assumption that each choice offers commitment. However, in typical economic settings, commitment to a single consumption path is not feasible. For example, in period 0 the agent may have to choose between different investments that affect the feasible choices for periods 1 and 2. To determine the behaviour of the agent in such a decision problem, the model must specify how the decision-maker expects to behave in future periods. Strotz (1955) proposed the "strategy of consistent planning", and argued that the decision-maker will choose the optimal plan among those plans that he/she is willing to carry out in the future. Peleg and Yaari (1973) treat each decision period as a distinct player and solve for Nash equilibria in the resulting dynamic game. O'Donoghue and Rabin (1999) argue that the decision-maker may naively believe that his/her future behaviour will maximize current preferences. In each version of the model, the primitives are (1) behaviour at each decision date under the assumption of commitment and (2) a rule that specifies how the agent forms expectations.

In this paper, we develop an alternative approach to and interpretation of Strotz's model. Following Kreps[1] (1979) and our earlier work on self-control (Gul and Pesendorfer, 2001) we

[1] Kreps' main axiom captures preference for flexibility by allowing a two-element set to be preferable to either singleton. We rule out preference for flexibility and instead insist that the two-element set be indifferent to one of the singletons.

define preferences over decision problems. These preferences represent the agent's behaviour in period 0 when he/she must choose among alternatives that constrain future choices. Hence our model takes as its starting point the behaviour of the agent at one decision date and requires no hypothesis on expectation formation. However, we require that the agent be able to rank all decision problems and not just those that offer commitment. The advantage of our approach is that preferences over decision problems are—at least in principle—observable. Rather than speculate about the appropriate model of expectation formation, we offer choice experiments that identify Strotz's model of behaviour.

As an example, consider a three-period problem ($t = 0, 1, 2$) with consumption in periods 1 and 2. A period 2 decision problem denoted as z_2 is a set of consumption choices for that period. Hence z_2 represents the set of options available to the decision-maker in period 2. In a simple consumption–savings model, the set z_2 is determined by the agent's wealth at the beginning of period 2. In period 1, a decision problem is a set z_1 where each element of z_1 specifies a consumption choice for period 1 (denoted c_1) and a period 2 decision problem z_2. In a consumption–savings model z_1 would be determined by the agent's wealth at the beginning of period 1. In period 0, there is no consumption and the agent chooses among period 1 decision problems. The period 0 choice can be interpreted as a choice among "assets". For example, the agent may choose between a liquid savings account and an illiquid asset that can be converted into consumption only in period 2.

A standard decision-maker is characterized by a utility function U and his/her ranking of decision problems is described by the value function W where

$$\begin{aligned}&W(z_1) = \max U(c_1, c_2)\\&\text{subject to: } (c_1, z_2) \in z_1 \text{ and } c_2 \in z_2.\end{aligned}$$

The decision-maker described by Strotz is characterized by a utility function for each decision period. Let U_1, U_2 denote the utility functions for periods 1 and 2. Consistent planning requires that the agent maximizes U_2 in period 2 and, given this period 2 behaviour, the agent maximizes U_1 in period 1. This iterative maximization results in a collection of consumption paths consistent with U_1, U_2 and z_1. The utility function U describes the period 0 ranking of consumption paths. The ranking of decision problems is described by the value function W where

$$\begin{aligned}&W(z_1) = \max U(c_1, c_2)\\&\text{subject to: } (c_1, c_2) \text{ is consistent with } U_1, U_2, z_1.\end{aligned}$$

The main result of the paper (Theorem 4) gives three axioms on period 0 preferences that imply this representation.

To understand the key axiom consider a decision problem z_1 and two subsets x_1 and y_1 with the property that $z_1 = x_1 \cup y_1$. If $x_1 \succeq y_1$ then for a standard decision-maker an optimal choice from z_1 must be in x_1 and therefore $x_1 \sim z_1$. To allow for a preference for commitment we relax this requirement. Our weaker axiom requires that

$$x_1 \sim z_1 \qquad \text{or} \qquad y_1 \sim z_1. \tag{NC}$$

Note that (NC) allows for the possibility that $x_1 \succ z_1$, that is, a strict preference for a smaller set of alternatives. However, in that case the period 1 choice must be from y_1 and therefore $y_1 \sim z_1$.

The agent's preferences over decision problems describe the agent's behaviour in period 0. From these preferences, we derive a representation that suggests a particular choice behaviour in all future periods. This implied choice behaviour can be interpreted as the agent's *expectation* of future behaviour. To make this connection precise, we must confront the possibility that there may exist multiple Strotz representations of the same preference. Hence the agent's expected future behaviour may not correspond to the implied behaviour associated with a particular representation. In Section 2 we demonstrate how to construct a *canonical representation* that has the property that if a choice is optimal for some Strotz representation it is optimal for the canonical representation. Hence, the canonical representation identifies a set of choices in each period that must contain the agent's expectations. To see how this is done, consider the preference

$$x_1 \nsim x_1 \cup y_1 \sim y_1.$$

The above preference implies that the decision-maker *does not* expect the period 1 choice from $x_1 \cup y_1$ to be an element of x_1. Hence, by asking the agent to choose between decision problems we can elicit the agent's expectation of his/her future behaviour.

Section 2 also considers a variant of the model where we observe choice behaviour not only in period 0 but in all periods. In that case, the Strotz model is identified by the following two assumptions. First, behaviour in each period must be rational, that is, maximize some objective function. Second, period 0 preferences must satisfy an "irrelevance of redundant alternatives" axiom that requires that the agent's welfare is unaffected if we eliminate options from the period t choice set that will not be chosen. If the period 0 ranking of decision problems is unaffected by the elimination of redundant alternatives then we can conclude that the agent is "sophisticated", that is, has correct expectations.

In Section 3, we study a general, possibly non-separable version of the model due originally to Phelps and Pollak (1968). Phelps and Pollak's $\beta-\delta$ preferences constitute the most widely utilized subclass of Strotz's model. An agent with $\beta-\delta$ preferences discounts utility associated with consumption τ periods later at the rate $\beta\delta^\tau$, where $\beta < 1$. Hence, the agent is willing to give up more consumption in period $t+1$ in exchange for consumption in period t if he/she makes the decision in period t rather than in some earlier period. That is, the agent's behaviour displays a "bias" towards current consumption. Theorem 4 is a representation theorem for our generalization of the $\beta-\delta$ model. The theorem provides a revealed preference condition that characterizes this bias towards current consumption.

Sections 2 and 3 rely on the assumption that there are finitely many possible consumption levels. Section 4 analyses a model with "continuous" consumption problems. In particular, we assume that consumption in a given period can be any number in the unit interval. We introduce an assumption which we call "local preference for commitment". This assumption ensures that after any history $(c_1, \dots, c_t)$, we can always find two close consumption paths $(c'_{t+1}, \dots, c'_T)$, $(c''_{t+1}, \dots, c''_T)$ such that the agent would prefer $(c'_{t+1}, \dots, c'_T)$ to $(c''_{t+1}, \dots, c''_T)$ if he/she could commit at time t but will end up choosing $(c''_{t+1}, \dots, c''_T)$ rather than $(c'_{t+1}, \dots, c'_T)$ if commitment is not feasible and he/she has to make the choice in period $t+1$. Theorem 5 shows that the $\beta-\delta$ model (and our generalization of it) is inconsistent with local preference for commitment unless the time horizon is three periods or fewer.

The difficulty of providing a well-behaved Strotz model with continuous choice has been noted by Peleg and Yaari (1973) and other researchers. Theorem 5 motivates our final result by showing that the same difficulties arise even if we allow for preferences that depend on the history of past consumption. Our final result (Theorem 6) shows that we can approximate Strotz preferences with a well-behaved representation taken from our earlier work on self-control. The approximation that we provide is perfect in the sense that it can rationalize any finite set of observed choices consistent with Strotz's theory. Hence, no finite data-set can distinguish between Strotz behaviour and the self-control preferences that we used to approximate such behaviour.

In addition to the time-inconsistency literature cited above, this paper is related to our earlier work on self-control (Gul and Pesendorfer, 2001, 2004). In both of these papers we analyse choice under uncertainty (*i.e.* choice over lotteries). Hence, both papers deal with continuous decision problems and take advantage of the linear structure associated with lottery spaces.

Gul and Pesendorfer (2004) offer a recursive, infinite horizon model that either rules out preference for commitment or NC, the main assumption of this paper. Hence, the only intersection of the model in that paper and the current one is the standard, time-consistent model.

Gul and Pesendorfer (2001) allow NC but have only two periods. Hence, the difficulties associated with continuous choice discussed above do not come up in their setting. The two-period setting renders Phelps–Pollak preferences indistinguishable from the general Strotz model. In addition to ruling out lotteries, the setting of Theorem 1 below differs from the one in the earlier paper by allowing for many periods but only a finite consumption set. Lemma 1 of the proof of Theorem 1 is taken from that paper.

1.1. *Why choice experiments?*

Our approach, both in the current paper and in earlier work, differs from that of Strotz and the subsequent literature on dynamic inconsistency in that we do not consider the agent's expectation of his/her future behaviour a primitive of our model.

Instead, we take as primitive the agent's choices from a larger domain, the collection of decision problems. In our theorems both the hypotheses (*i.e.* axioms) and the conclusions are statements about what choices the agent makes in various situations.

As in standard models of dynamic choice we view the decision-maker as expressing a preference at one point in time (period 0). The representation of these preferences suggests behaviour in future periods that can be interpreted as the agent's implicit expectations. Whether these expectations are correct or not (that is, whether the agent is sophisticated or not) can be treated as a separate question. That is, the representation is a valid description of period 0 behaviour whether or not the agent has correct expectations, as long as the axioms are satisfied.

To offer a testable hypothesis for "sophisticated" behaviour, we also examine a model that considers behaviour in all periods. There we find that a simple independence of redundant alternatives (IRA) condition is the only restriction on intertemporal choice behaviour implied by Strotz's preferences. IRA says that period 0 behaviour should be unaffected if unchosen alternatives are eliminated from choice sets.

1.2. *A time-consistent interpretation*

The time-inconsistency literature takes the view that agents have distinct and independent preferences in each period. This implies that for the purpose of welfare analysis there are as many agents as there are decision nodes. As a result, welfare statements are often ambiguous.

Consider the example of a consumer who must choose between a liquid and an illiquid asset. The liquid asset can be converted into high consumption in period 1 or period 2. The illiquid asset commits the consumer to low consumption in period 1 and delivers high consumption in period 2. If the consumer has a strict preference for the illiquid asset it must mean that the consumer would have chosen high consumption in the event that the liquid asset was chosen in period 0. The time-inconsistent interpretation must therefore conclude that the period 0 choice has ambiguous welfare implications: it increases the utility of the period 0 "self" but decreases the utility of the period 1 self. The time-inconsistent interpretation implies the following ranking in period 0:

$$\text{illiquid/low consumption} \sim \text{liquid/low consumption} \succ \text{liquid/high consumption}$$

whereas for period 1 the ranking is

$$\text{liquid/high consumption} \succ \text{liquid/low consumption} \sim \text{illiquid/low consumption.}$$

We propose an alternative time-consistent interpretation of the model. The time-consistent interpretation asserts that in *all periods* the ranking is given by

$$\text{illiquid/low consumption} \succ \text{liquid/high consumption} \succ \text{liquid/low consumption.}$$

Here, the agent is better off in *all* periods if he/she chooses the illiquid asset. If the liquid asset is chosen in period 0, the optimal behaviour in period 1 (from the perspective of all periods) is to choose high consumption. In this interpretation, the asset choice affects welfare. The liquid asset makes high consumption available in period 1 which reduces welfare in both periods. The interpretation is that high consumption constitutes a "temptation" that the decision-maker finds impossible to resist. If this temptation is available in period 1, high consumption is the optimal choice from the perspective of all periods because resisting the temptation would be too costly for the decision-maker.

Note that both interpretations are consistent with observed behaviour. The agent's behaviour reveals that in period 1

$$\text{liquid/high consumption} \succ \text{liquid/low consumption}$$

and in period 0

$$\text{illiquid/low consumption} \succ \text{liquid/high consumption.}$$

This is satisfied under either interpretation. The advantage of the time-consistent interpretation is that the observed choices are welfare maximizing from the perspective of every period. Therefore, to determine whether a policy (for example, a tax policy that discourages high consumption) improves the welfare of the agent it suffices to determine whether the agent would vote for the policy in the period in which it is introduced.

2. WEAK STROTZ PREFERENCES

We consider a model with T decision-making periods. Each period, the agent takes an action that results in a consumption for that period and a decision problem for the next period. Let C denote the set of possible consumptions. We assume that C is finite. For concreteness, we may think of C as a finite subset of $\mathbb{R}_+$, where each $c_t \in C$ denotes the level of consumption in period t of the single good.

For any non-empty, finite set X, let $K(X)$ denote the collection of all non-empty subsets of X. Let $Z_T := K(C)$ denote the set of one-period decision problems. For $1 \leq t < T$ we define inductively the set of $T - t + 1$-period decision problems as

$$Z_t := K(C \times Z_{t+1}).$$

Each $z_t \in Z_t$ is a finite menu of choices of the form (c, z_{t+1}) where c is the consumption in period t and z_{t+1} is the continuation problem in period $t + 1$. The set Z_1 denotes the collection of T-period decision problems.

Example 1. There are three consumption periods ($T = 3$) and consumption in each period is either high or low ($c_t \in \{h, l\}$). Consider three decision problems in which the agent can afford high consumption in exactly one period. Hence, the agent must choose a single t as his/her period of high consumption. First, consider the situation where the agent may choose high consumption in any period. Then, the decision problem is $z_1 = \{(l, z_2), (h, z_2')\}$ where $z_2 = \{(l, \{h\}), (h, \{l\})\}$, $z_2' = \{(l, \{l\})\}$. This decision problem can be represented by the decision tree z_1 in Figure 1.

The decision problem in which the agent is committed to low consumption in period 1 is $y_1 := \{(l, z_2)\}$ (with z_2 as defined above) and illustrated by the decision tree y_1 in Figure 1.

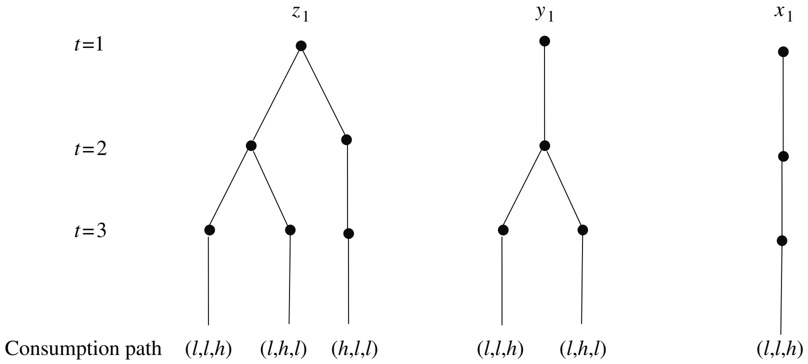

FIGURE 1
Decision problems

The decision problem in which the agent is committed to low consumption in periods 1 and 2 (and hence has no choice) is $x_1 := \{(l, x_2)\}$ with $x_2 = \{(l, \{h\})\}$ and represented by the decision tree x_1 in Figure 1.

Our model considers a decision-maker who has a preference defined on Z_1, the set of T-period decision problems. This preference describes the decision-maker's behaviour in period 0, prior to the first consumption period. To interpret this preference consider Example 1. The period 0 choice between x_1, y_1 and z_1 can be viewed as a choice between assets of varying degrees of liquidity. "Asset" z_1 can be converted into high consumption in any period and hence is the most liquid whereas x_1 is the least liquid since it can be converted into high consumption only in period 3. More generally, the period 0 decision can be thought of as a portfolio decision. Each portfolio defines a decision problem for subsequent periods. The preference $\succeq$ on Z_1 therefore represents a ranking of portfolios. A function representing the preference $\succeq$ is analogous to a value function in dynamic programming.

Strotz (1955) introduced a model of decision making with changing tastes. In that model a utility function (over consumption paths) for each decision date is specified. These utility functions may be inconsistent, that is, the utility functions in periods t and $t + 1$ may disagree in their ranking of consumption paths. To deal with this inconsistency, Strotz (1955) proposes the *strategy of consistent planning* according to which an individual who cannot commit to a consumption path "*rejects any plan which he will not follow through. His problem is then to find the best plan among those that he will actually follow*". An implication of time-inconsistent utilities and consistent planning is that the agent may have a *preference for commitment*, that is, he/she may express a strict preference for a situation where he/she has fewer choices. In Example 1, a strict preference for x_1 over y_1 or z_1 illustrates a preference for commitment.

The approach taken in this paper is to take as primitive the agent's period 0 behaviour as described by the preference $\succeq$. To allow for the type of behaviour described in the time-inconsistency literature, we let $\succeq$ have a preference for commitment. Our objective is to provide conditions on the preference $\succeq$ that identify the Strotz model of changing tastes. The key feature of this approach is that all *assumptions* are made in terms of choice experiments and therefore correspond to—in principle—observable behaviour. In contrast, it is difficult to see how assumptions that are made in terms of the agent's expectations, as in Strotz's definition of consistent planning, can be tested directly.

The following notation is needed to define consistent plans. For $t \geq 1$, a *t-period history* is a t-tuple of consumptions $(c_1, \ldots, c_t)$. Let H_t denote the set of all t-period histories and h_t denote a generic element of H_t. In order to avoid having to make separate statements for $t = 1$ and $t > 1$, we fix an arbitrary element $c_0 \in C$ and refer to $h_0 = c_0$ as *the* 0-period history and define $H_0 = \{h_0\}$. We also refer to H_T as the set of consumption paths. Let $H = \bigcup_{t=0}^{T} H_t$ be the set of all histories and let $h \in H$ denote a generic history.

A *node* at time t specifies a consumption history up to but not including time t and a $T - t + 1$-period decision problem z_t. Hence, a node at time t, $1 \leq t \leq T$, is a pair (h_{t-1}, z_t) where $h_{t-1} \in H_{t-1}$ and $z_t \in Z_t$. Let N_t be the set of all nodes at time t and $N = \bigcup_{t=1}^{T} N_t$. Generic elements of N_t and N are denoted as η_t and η, respectively.

A *plan* specifies the behaviour of a decision-maker at every decision node. Formally, a plan is a map $\phi : N \to N \cup C^T$. For $t < T$, a plan associates with each $\eta_t = (h_{t-1}, z_t) \in N_t$ a node $\eta_{t+1} = (h_{t-1}, c, z_{t+1}) \in N_{t+1}$ with history one period longer such that $(c, z_{t+1}) \in z_t$. For $t = T$ the plan ϕ associates with each node $\eta_T = (h_{T-1}, z_T)$ a consumption sequence (h_{T-1}, c) such that $c \in z_T$. Let Φ denote the set of plans.

For any plan $\phi \in \Phi$, the map ϕ^k specifies the outcome if the plan ϕ is applied k times. More precisely, for a given $\phi \in \Phi$ and $k = 0, 1, \ldots, T$, define $\phi^k : N \cup C^T \to N \cup C^T$ inductively as follows: $\phi^k(c_1, \ldots, c_T) = (c_1, \ldots, c_T)$ for all $(c_1, \ldots, c_T) \in C^T$ and all k. For all $\eta \in N$, $\phi^0(\eta) = \eta$, $\phi^1(\eta) = \phi(\eta)$ and $\phi^{k+1}(\eta) = \phi^k(\phi(\eta))$. Hence, $\phi^T(\eta)$ denotes the consumption path induced by ϕ given the node η.

Next, we define a generalized version of Strotz's model. The agent has a period 0 utility function U over consumption paths and objective functions V_t for all t. Consistent plans maximize V_t for each t. From the set of consistent plans the agent picks the one that maximizes the period 0 utility function U. The function $U : C^T \to \mathbb{R}$ assigns a utility to each consumption path. The function V_t assigns a utility to each period t decision and may depend on the consumption history. In period T the decision yields a consumption c_T and therefore $V_T : H_{T-1} \times C \to \mathbb{R}$. For period $t < T$ a decision yields a consumption c and a decision problem z_{t+1} for the following period. Hence, $V_t : H_{t-1} \times C \times Z_{t+1} \to \mathbb{R}$. Note that for $t < T$ the domain of V_t is the period $t + 1$ decision nodes N_{t+1}.

In Strotz's model, the functions V_t are derived from preferences over consumption streams through backward induction either with correct or naive expectations about future behaviour. In our revealed preference approach, we derive the V_t's from period 0 preferences. In Theorem 4, we impose a condition on the agent's preferences that ensures that the V_t's can be derived from some preferences over consumption streams through backward induction. Theorems 1 and 2 deal with the unrestricted case where V_t represents a general constraint on the consistent choices at each decision node.[2]

To illustrate the generality of the framework, consider the following example. There are two stores, denoted by A and B. The selection in store B is a strict subset of the selection in store A and the agent is given the choice of whether or not to enter one of the stores in period 1. When agents have the option of entering store A, they enter, while they do not enter when they have the option of entering store B. Nevertheless, from store A the agent makes a choice (in period 2) that is also available in store B. One interpretation of this is that the agent is "tempted" by the availability of options that he/she knows he/she will not take advantage of. The weak Strotz

[2] By setting $V_t(\eta) = U_t(\phi^T(\eta))$ we can associate a unique V_t with any utility function $U_t : C^T \to \mathbb{R}$ and plan ϕ. However, the converse is not true; given a V_t, it may not be possible to find any U_t, ϕ such that $V_t(\eta) = U_t(\phi^T(\eta))$. Observe that when the cardinality of C is greater than 1 and $t < T$, the cardinality of $H_{t-1} \times C \times Z_{t+1}$ is greater than the number of possible consumption paths. Therefore, for any V_t such that $V_t(\eta) \neq V_t(\eta')$ for $\eta \neq \eta'$, there is no U_t, ϕ such that $V_t(\eta) = U_t(\phi^T(\eta))$.

representation of Theorem 1 below permits this (the V_1 of store A is higher than the V_1 of store B) whereas the preferences of Theorem 4 do not.

A second way to interpret the store example would be to say that the agent has incorrect expectations and expects to make a choice from store A that is not available in store B. Theorem 3 identifies a condition that rules out the second interpretation. More generally, our approach enables us to distinguish between what can be tempting and whether or not the agent correctly anticipates his/her future behaviour.

Consider a decision node $\eta_t = (h, z_t) \in N_t$ and a plan ϕ with $\phi(\eta_t) = (h, c, z_{t+1})$. The plan ϕ is consistent with V_t at decision node η_t if

$$V_t(\phi(\eta_t)) = V_t(h, c, z_{t+1}) \geq V_t(h, c', z'_{t+1})$$

for all $(c', z'_{t+1}) \in z_t$ or, equivalently,

$$V_t(\phi(\eta_t)) \geq V_t(\psi(\eta_t))$$

for all plans $\psi \in \Phi$. Let $\bar{V} = (V_1, \dots, V_T)$. The set of plans consistent with $\bar{V}$ is denoted by $P^*(\bar{V})$ and defined as follows:

$$P^*(\bar{V}) := \{\phi \in \Phi \mid V_t(\phi(\eta_t)) \geq V_t(\psi(\eta_t)) \forall \psi \in \Phi \forall t\}.$$

In period 0, the agent evaluates decision problems $z \in Z_1$ by maximizing the period 0 utility $U : C^T \to \mathbb{R}$ among all plans in the set $P^*(\bar{V})$. Hence, the value of decision problem $z \in Z_1$ is given by

$$W(z) := \max_{\phi \in P^*(\bar{V})} U(\phi^T(z)).$$

Definition. The value function $W : Z_1 \to \mathbb{R}$ is weak Strotz if there exists $U, \bar{V}$ such that $W(z) := \max_{\phi \in P^*(\bar{V})} U(\phi^T(z))$.

The value function W represents the preference $\succeq$ if $W(z) \geq W(z')$ if and only if $z \succeq z'$. We say that $\succeq$ is a weak Strotz preference if there is a W that is weak Strotz and represents $\succeq$. We refer to $(U, \bar{V})$ as a representation of $\succeq$ or sometimes simply as the weak Strotz preference $(U, \bar{V})$.

Theorem 1 shows that the following two axioms are necessary and sufficient for $\succeq$ to be a weak Strotz preference.

Axiom P (Preference Relation). $\succeq$ is a complete and transitive binary relation.

For any history $h = (c_1, \dots, c_{t-t})$ and a decision problem z_t, we write $\{h, z_t\}$ to denote the decision problem in which the agent is committed to the consumption $(c_1, \dots, c_{t-1})$ in the first $t - 1$ periods and then is confronted with the decision problem z_t. Similarly, with some abuse of notation, we write $\{h, c, z_{t+1}\}$ to denote the situation where the agent is committed to $(c_1, \dots, c_{t-1}, c)$ in the first t periods and then is confronted with the decision problem z_{t+1}.

Consider a decision problem $z_t \cup z'_t$ and a consumption history $h = (c_1, \dots, c_{t-1})$. If the choice at node $(h, z_t \cup z'_t)$ is in z_t, then a preference that only cares about the choice must satisfy $\{h, z_t \cup z'_t\} \sim \{h, z_t\}$. If the choice at node $(h, z_t \cup z'_t)$ is in z'_t, then such a preference must satisfy $\{h, z_t \cup z'_t\} \sim \{h, z'_t\}$. This motivates the following axiom.

Axiom NC (No Compromise). $\{h, z_t \cup z'_t\} \sim \{h, z_t\}$ or $\{h, z_t \cup z'_t\} \sim \{h, z'_t\}$.

Theorem 1. *The preference $\succeq$ satisfies Axioms P and NC if and only if it is a weak Strotz preference.*

Proof. See Section 5.

In the proof of Theorem 1 we construct the utility functions U, V_t for $t = 1, \ldots, T$. The period 0 utility U represents the preference $\succeq$ restricted to decision problems that offer commitment. That is,

$$U(c_1, \ldots, c_T) \geq U(c'_1, \ldots, c'_T) \qquad \text{if and only if} \qquad \{c_1, \ldots, c_T\} \succeq \{c'_1, \ldots, c'_T\}.$$

The objective function $V_t(h, \cdot)$ represents the binary relation R_h defined below.

Definition. $(c, z_{t+1}) R_h (c', z'_{t+1})$ if and only if

(i) $\{h, c, z_{t+1}\} \nsim \{h, c', z'_{t+1}\}$ and $\{h, c, z_{t+1}\} \sim \{h, \{(c, z_{t+1}), (c', z'_{t+1})\}\}$ or
(ii) $\{h, c, z_{t+1}\} \sim \{h, c', z'_{t+1}\}$ and $[\{h, c', z'_{t+1}\} \sim \{h, \{(c', z'_{t+1}), (\hat{c}, \hat{z}_{t+1})\}\}$ implies $\{h, c, z_{t+1}\} \sim \{h, \{(c, z_{t+1}), (\hat{c}, \hat{z}_{t+1})\}\}]$.

To motivate this definition, consider a situation where $\{h, c, z_{t+1}\} \nsim \{h, c', z'_{t+1}\}$. In that case, $\{h, c, z_{t+1}\} \sim \{h, \{(c, z_{t+1}), (c', z'_{t+1})\}\}$ must mean that the agent expects (c, z_{t+1}) to be chosen after history h. Hence, it must be the case that $V_t(h, c, z_{t+1}) \geq V_t(h, c', z'_{t+1})$.

In the proof of Theorem 1 we establish that the relation R_h is a preference relation (complete and transitive) and therefore can be represented by a utility function $V_t(h, \cdot)$.

Note that the preference in the above theorem describes behaviour only in period 0. In contrast, the weak Strotz representation implies a behavioural rule for all periods. More precisely, the weak Strotz representation implies *expectations* for what will be chosen in subsequent periods.

Our next objective is to make precise the sense in which we can elicit expectations of future behaviour from the preference $\succeq$. Typically, there will be multiple Strotz representations for a single Strotz preference. However, the utility functions defined above are a *canonical representation* with the property that predicted behaviour from any other Strotz representation must be optimal for $(U, \bar{V})$. We say that $(U, \bar{V})$ is a canonical Strotz representation of $\succeq$ if U represents the commitment preference (as defined above) and $V_t(h, \cdot)$ represents R_h.

Theorem 2 shows that if a plan is optimal for any Strotz representation it must be optimal for the canonical representation. The canonical representation therefore allows us to make inferences about how the agent expects to choose in subsequent periods. More precisely, we know that the agent does not expect to choose alternatives that are suboptimal for the canonical representation.

Theorem 2. *Let $\succeq$ satisfy Axioms P and NC and let $(U, \bar{V})$ be a canonical representation of $\succeq$. Then*

$$\arg\max_{\phi \in P^*(\bar{V}')} U'(\phi^T(z)) \subset \arg\max_{\phi \in P^*(\bar{V})} U(\phi^T(z))$$

for any Strotz representation $(U', \bar{V}')$ of $\succeq$.

Proof. See Section 5.

To illustrate how our model allows inference about expectations consider Example 1. In that example, the agent must choose between three decision problems x_1, y_1, z_1 in period 0. Suppose the agent expresses the preference

$$x_1 \succ y_1 \succ z_1. \qquad (*)$$

Recall that in each decision problem the agent can enjoy high consumption in exactly one period and has low consumption in the remaining two periods. Recall also that x_1 commits the agent to high consumption in period 3 whereas y_1 commits the agent to low consumption in period 1 and offers a period 2 choice between high consumption in period 2 or high consumption in period 3. From $x_1 \succ y_1$ we conclude that the agent strictly prefers to commit to high consumption in period 3 over a decision problem where high consumption can be chosen either in period 2 or in period 3. This implies that he/she expects to choose the *excluded* alternative if commitment is not available. Hence, the agent expects to choose high consumption in period 2 if this choice is available. Similarly, the preference $y_1 \succ z_1$ implies that the agent prefers commitment to low consumption in period 1 and hence expects to choose high consumption in period 1 if that choice is available.

To this point our model is based entirely on period zero behaviour (as described by the preference $\succeq$) and therefore cannot address the question of whether the agent makes choices in periods $t > 0$ that are consistent with period 0 expectations. In other words, the model is silent on the question of whether expectations are correct. For example, naive agents as described in O'Donoghue and Rabin satisfy Axioms P and NC but their behaviour in periods $t > 0$ is not consistent with expectations. To see this, suppose in Example 1 the agent expresses the preference

$$x_1 \sim y_1 \sim z_1$$

and at the same time strictly prefers x_1 (commitment to low consumption in periods 1 and 2) to a situation where the agent is committed to low consumption in period 3. In other words, the agent has no preference for commitment and expresses a preference to delay high consumption until period 3. From this we can conclude that the agent expects to choose low consumption in periods 1 and 2. This is consistent with standard agents who are time consistent but also with naive agents who expect to be time consistent but contradict this expectation in their subsequent choice behaviour.

In order to identify agents whose behaviour is consistent with period 0 expectations ("sophisticated agents") we need to observe behaviour in all periods. Let $\mathcal{C}_1, \ldots, \mathcal{C}_T$ denote a collection of choice functions that describe behaviour in periods $t \geq 1$. Hence, $\mathcal{C}_t : H_{t-1} \times Z_t \to Z_t$, with $\mathcal{C}_t(h, z_t) \subset z_t$ and $\mathcal{C}_t(h, z_t) \neq \emptyset$. Our next objective is to characterize weak Strotz preferences in terms of behaviour in all periods.

The first axiom (Axiom H) says that choice behaviour in periods $t \geq 1$ satisfies the familiar Houthakker axiom below.

Axiom H (Houthakker's Axiom). $\mathcal{C}_t(h, z_t) \cap z'_t \neq \emptyset$ implies $\mathcal{C}_t(h, z'_t) \cap z_t \subset \mathcal{C}_t(h, z_t)$.

It is well known that Axiom H is equivalent to rational choice, that is, the choice function $\mathcal{C}_t$ maximizes some objective function.

As before, behaviour in period 0 is described by a preference $\succeq$ on Z_1. The following axiom relates period 0 behaviour to behaviour in later periods by assuming that the agent is indifferent between a decision problem and the option that he/she chooses from it.

Axiom IRA (Independence of Redundant Alternatives). $(c, z_{t+1}) \in \mathcal{C}_t(h, z_t)$ implies $\{h, c, z_{t+1}\} \sim \{h, z_t\}$.

One implication of Axiom IRA is that it rules out the possibility that unchosen alternatives could affect the agent's well-being, as in the model of self-control analysed in Gul and Pesendorfer (2001). Axiom IRA also rules out naive behaviour where period 0 expectations are inconsistent with period t choices. Naive agents expect to make choices that differ from their actual choices. If the expected choice is removed from a set, the agent's utility in period 0 changes while the actual choice may be unaffected hence leading to a violation of IRA.

The choice function $\mathcal{C}_t$ maximizes V_t if for all $(h, z_t) \in H_{t-1} \times Z_t$

$$\mathcal{C}_t(h, z_t) = \arg\max_{(c, z_{t+1}) \in z_t} V_t(h, c, z_{t+1}).$$

Theorem 3 shows that rational choice functions $\mathcal{C}_1, \ldots, \mathcal{C}_T$ and the preference $\succeq$ satisfy IRA if and only if the preference has a weak Strotz representation $(U, V_1, \ldots, V_T)$ such that $\mathcal{C}_t$ maximizes V_t for every $t \geq 1$.

Theorem 3. *Let $\succeq$ be a binary relation that satisfies Axiom P and let $\mathcal{C}_1, \ldots, \mathcal{C}_T$ satisfy Axioms H. Then, $\succeq, \mathcal{C}_1, \ldots, \mathcal{C}_T$ satisfy IRA if and only if there exists a weak Strotz representation $(U, V_1, \ldots, V_T)$ of $\succeq$ such that $\mathcal{C}_t$ maximizes V_t for all t.*

Proof. See Section 5.

Theorem 3 establishes that independence of redundant alternatives (IRA) is the only temporal revealed preference implication of weak Strotz behaviour. The theorem bridges the gap between the revealed preference approach adopted in this paper and the analysis based on expectation that is standard in the time-inconsistency literature. Unlike the previous two theorems, in Theorem 3, we consider as primitive not only period 0 behaviour but behaviour in each period. This enables us to compare the beliefs derived from period 0 preferences regarding behaviour in period t with the actual behaviour in period t. Suppose the choice behaviour of the decision-maker in every period is consistent with maximizing a preference relation. Then, Theorem 3 ensures that whenever IRA is satisfied, the observed behaviour in period t is consistent with the prediction of period t behaviour derived from the period 0 preferences. Hence, the theorem proves that the only observable implication of sophisticated behaviour is IRA. Conversely, the theorem ensures that if the decision-maker is sophisticated (*i.e.* if his/her period 0 preferences anticipated his/her future behaviour correctly), he/she will satisfy IRA.

3. PHELPS–POLLAK PREFERENCES

The time-inconsistency literature (see, for example, Laibson, 1997) specifies a utility function for each decision date. These utility functions are defined over consumption paths. In contrast, the weak Strotz model of the previous section allows for a general period t objective function with period t decisions as their domain.

A commonly used example of time-inconsistent utility functions is known as β–δ utility. These utility functions were introduced by Phelps and Pollak (1968) and were further analysed by Laibson (1997). Example 2 provides an illustration.

Example 2 (β–δ utility). Let $\beta, \delta \in (0, 1]$, $u : C \to \mathbb{R}$. Define

$$U_t(c_1, \ldots, c_T) = u(c_t) + \beta \sum_{k=1}^{T-t} \delta^k u(c_{t+k})$$

for all $t = 1, \ldots, T$ where $\sum_{k=1}^{0}(\cdot) \equiv 0$. Let

$$U(c_1, \ldots, c_T) = \sum_{t=1}^{T} \delta^t u(c_t)$$

be the period 0 utility function. (Recall that our model has no consumption in period 0.) If $\beta < 1$ the utility functions U_t and U may disagree in their rankings of consumption paths. However, the rankings of U and U_t agree if consumption in periods $t' \leq t$ is held fixed.

In this section, we analyse preferences that can be represented by a generalized version of β–δ utility. Let $\bar{U} = (U_1, \ldots, U_T)$ be a collection of utility functions where $U_t : C^T \to \mathbb{R}$.

We say that U and $\bar{U}$ agree on continuations if, for all $t = 1, \dots, T$,

$$U_t(c_1, \dots, c_T) \geq U_t(c'_1, \dots, c'_T) \qquad \text{iff} \qquad U(c_1, \dots, c_T) \geq U(c'_1, \dots, c'_T)$$

whenever $c_\tau = c'_\tau$ for all $\tau \leq t$.

Our next objective is to define consistent plans for a collection of utility functions $\bar{U}$. Unlike the objective functions $\bar{V}$ analysed in the previous section, the utility functions $\bar{U}$ assign utility to consumption sequences. Consistent plans for $\bar{U}$ are defined inductively. In period T, a consistent plan must be optimal for U_T. This gives us a set of period T consistent plans, $P_T(\bar{U})$. In period $T-1$, a consistent plan must maximize U_{T-1} among all plans in $P_T(\bar{U})$. This yields $P_{T-1}(\bar{U})$, the consistent plans for periods $T-1$ and T. We proceed inductively to define $P_1(\bar{U})$, the set of consistent plans for all periods.

Let $P_{T+1}(\bar{U}) = \Phi$. For $t \leq T$ we inductively define

$$P_t(\bar{U}) := \{\phi \in P_{t+1}(\bar{U}) \mid U_t(\phi^T(\eta_t)) \geq U_t(\psi^T(\eta_t)) \forall \psi \in P_{t+1}(\bar{U})\}.$$

(Recall that $\phi^T(\eta)$ denotes the consumption path generated by plan ϕ and node η.) The set $P_t(\bar{U})$ contains all plans that are consistent with $\bar{U}$ at times $t, t+1, \dots, T$. Let

$$P(\bar{U}) := P_1(\bar{U})$$

denote the set of consistent plans. Consistent planning requires that in period 0 the agent evaluates decision problems $z \in Z_1$ by maximizing the period 0 utility among all plans in the set $P(\bar{U})$. Hence, the value of decision problem $z \in Z_1$ is given by

$$W(z) := \max_{\phi \in P(\bar{U})} U(\phi^T(z)). \tag{S}$$

Definition. The value function $W : Z_1 \to \mathbb{R}$ is Phelps–Pollak (PP) if there is $(U, \bar{U})$ such that U and $\bar{U}$ agree on continuations and $W(z) := \max_{\phi \in P(\bar{U})} U(\phi^T(z))$.

We say that a preference $\succeq$ is a PP preference if it can be represented by a PP value function W. We refer to the corresponding $(U, \bar{U})$ as a PP representation of $\succeq$ or sometimes as the PP preference $(U, \bar{U})$. Theorem 4 shows that we get PP preferences if we impose the following axiom (Axiom TCC) in addition to Axioms P and NC.

Axiom TCC (Temptation by Current Consumption). If $\{h, c, z'_{t+1}\} \succeq \{h, c, z''_{t+1}\}$ and $(c, z'_{t+1}) \in z_t$ then $\{h, z_t \cup \{(c, z''_{t+1})\}\} \sim \{h, z_t\}$.

Axiom TCC considers situations where an alternative—(c, z''_{t+1})—is added to the choice set z_t. The period t consumption of the new alternative (c, z''_{t+1}) is the same as the period t consumption of an already existing alternative (c, z'_{t+1}). Moreover, commitment to the existing alternative is preferred to commitment to the new alternative (*i.e.* $\{h, c, z'_{t+1}\} \succeq \{h, c, z''_{t+1}\}$). In other words, the existing alternative has a better continuation than the added alternative. The axiom requires that the addition of (c, z''_{t+1}) to z_t has no effect on the agent's welfare. An agent whose period t utility agrees with his/her earlier utility functions when current consumption is unaffected would certainly satisfy this requirement. Theorem 4 shows that the converse is also true. Axioms P, NC and TCC imply that the preference has a PP representation.

Theorem 4. *The preference $\succeq$ satisfies P, NC and TCC if and only if it is a PP preference.*

Proof. See Section 5.

Theorem 4 characterizes agents with Phelps–Pollak preferences in terms of their period 0 preferences over decision problems. The axioms P, NC and TCC can be interpreted as the testable implications of the PP model.

As in the case of weak Strotz preferences, Theorem 4 characterizes period 0 behaviour but the representation implies expectations for the behaviour in subsequent periods. To illustrate further the role of Axiom TCC consider the three-period decision problems illustrated in Example 1. Suppose

$$x_1 \succ y_1 \succ z_1$$

and assume Axioms P, NC and TCC hold. Recall that x_1 represents the decision problem where the agent is committed to high consumption in period 3; y_1 represents the decision problem where the agent can choose high consumption either in period 2 or in period 3; and z_1 represents the decision problem where high consumption can be chosen in any one of the three periods. From $x_1 \succ y_1$ we conclude that a PP representation $(U, \bar{U})$ must satisfy $U(l, l, h) > U(l, h, l)$ and $U_2(l, h, l) > U_2(l, l, h)$. Together with $y_1 \succ z_1$ this in turn implies that $U(l, h, l) > U(h, l, l)$ and $U_1(h, l, l) > U_1(l, h, l)$. Consider a situation where in period 1 the agent is committed to low consumption and must choose between $x_2 = \{l, \{h\}\}$, $y_2 = \{h, \{l\}\}$ and $z_2 = \{(l, \{h\}), (h, \{l\})\}$. Consistent planning implies that the agent chooses x_2 and is indifferent between y_2 and z_2. The reason for this indifference is that the agent expects high consumption to be chosen from z_2 in period 2. The strict preference for x_2 follows because U and U_1 agree on continuations and therefore $U_1(l, l, h) > U_1(l, h, l)$. Note that if TCC is not assumed (and therefore we have a weak Strotz representation) the model places no restriction on the period 1 choice between (l, x_2), (l, y_2) and (l, z_2). For example, the agent may choose (l, y_2) over (l, z_2) in period 1 even though (l, y_2) and (l, z_2) ultimately lead to the same consumption path.

4. CONTINUITY AND APPROXIMATION

To this point we have assumed a finite set of possible consumptions in each period. This section analyses extensions of the model to a setting with a "continuous" choice of consumption levels.

For simplicity, let $D = [0, 1]$ denote the set of feasible consumptions in each period. For any subset X of a metric space, let $\tilde{K}(X)$ denote the non-empty compact subsets of X. Let $\tilde{Z}_T = \tilde{K}(D)$. For periods $t < T$ we then define $\tilde{Z}_t$ inductively as $\tilde{Z}_t := \tilde{K}(D \times \tilde{Z}_{t+1})$. The domain of preferences is $\tilde{Z}_1$. As before, N_t denotes the set of nodes at time t, Φ denotes the set of plans. The definition of nodes and plans are identical to the corresponding definitions given in Section 2 for finite decision problems.

We define a PP preferences $\succeq$ on $\tilde{Z}_1$ as in the discrete case: for any $(U, \bar{U})$, we say that U and $\bar{U}$ agree on continuations if, for all $t = 1, \dots, T$,

$$U_t(c_1, \dots, c_T) \geq U_t(c'_1, \dots, c'_T) \qquad \text{iff} \qquad U(c_1, \dots, c_T) \geq U(c'_1, \dots, c'_T)$$

whenever $c_\tau = c'_\tau$ for all $\tau \leq t$. The value function $W : Z_1 \to \mathbb{R}$ is defined by

$$W(z) := \max\nolimits_{\phi \in P(\bar{U})} U(\phi^T(z)). \tag{S*}$$

The value function W is Phelps–Pollak (PP) if there is $(U, \bar{U})$ such that U and $\bar{U}$ agree on continuations and $W(z) := \max_{\phi \in P(\bar{U})} U(\phi^T(z))$. Finally, we say that a preference $\succeq$ is a PP preference if it can be represented by a PP value function W and refer to $(U, \bar{U})$ as the PP representation of $\succeq$.

In this new setting, we consider preferences that have a monotone and continuous PP representation $(U, \bar{U})$. That is, we say that $\succeq$ has a continuous and monotone PP representation if there exist continuous, strictly increasing functions U and U_t, for $t = 1, \dots, T$, such that the function W defined by (S^*) represents $\succeq$.

Let $\|\cdot\|$ denote the Euclidean norm. We say that $\succeq$ has *local preference for commitment* if for all t such that $1 \leq t \leq T-2$, $h = (c_1, \dots, c_T) \in H$, and $\varepsilon > 0$, there exist $h' = (c'_1, \dots, c'_T)$, $h'' = (c''_1, \dots, c''_T) \in H$ satisfying $\|h - h'\| < \varepsilon$, $\|h - h''\| < \varepsilon$, $c_\tau = c'_\tau = c''_\tau$ for all $\tau \leq t$, $U_t(h') > U_t(h'')$ and $U_{t+1}(h') < U_{t+1}(h'')$.

Local preference for commitment implies that given any consumption history h, there are alternative consumption histories h', h'' such that both h', h'' are arbitrarily close to h, agree with h in every period up to t and the agent would strictly prefer committing to h' in period t to making the choice between h' and h'' in period $t+1$. Hence, local preference for commitment ensures that preference reversals arise even when the stakes are small. For $\beta{-}\delta$ preferences this difference between the agent's rankings at different times is captured by the difference between the discount rates δ and $\beta\delta$. For such preferences local preference for commitment is satisfied whenever the function u is strictly increasing and $\beta \neq 1$. In general, the local preference for commitment assumption does not require additive separability or history independence. The following theorem generalizes an example by Peleg and Yaari (1973) to provide a general impossibility theorem for multi-period Strotz preferences.

Theorem 5. *Suppose $\succeq$ has a continuous and monotone PP representation and has local preference for commitment. Then $T \leq 3$.*

Proof. See Section 5.

The main idea in the proof of Theorem 5 can be understood with the $\beta{-}\delta$ example below. The proof ensures that a similar example can be constructed whenever local preference for commitment is satisfied and U, U_t are all continuous, strictly increasing functions.

Example 3. Suppose $T = 4$, $\delta = 1$, $\beta = 0{\cdot}5$ and $u(c) = c$. Consider the corresponding $\beta{-}\delta$ preference on Z_1, that is $U(c_1, c_2, c_3, c_4) = c_1 + c_2 + c_3 + c_4$, $U_1(c_1, c_2, c_3, c_4) = c_1 + \frac{1}{2}(c_2 + c_3 + c_4)$, $U_2(c_1, c_2, c_3, c_4) = c_2 + \frac{1}{2}(c_3 + c_4)$, $U_3(c_1, c_2, c_3, c_4) = c_3 + \frac{1}{2}c_4$ and $U_4(c_1, c_2, c_3, c_4) = c_4$. Let $z_3(\gamma) = \{(\gamma, \{0\}), (0, \{1\})\}$. That is, $z_3(\gamma)$ denotes the decision problem in which the agent must choose between the consumption pair $(c_3, c_4) = (\gamma, 0)$ and $(c_3, c_4) = (0, 1)$ in period 3. (There is no choice in period 4.) Let $z_3^* = \{1, \{1\}\}$. Hence, z_3^* is the period 3 decision problem that guarantees the maximal consumption in the last two periods. Let $z_2(\gamma) = \{(1, z_3(\gamma)), (0{\cdot}4, z_3^*)\}$. Hence, in $z_2(\gamma)$ the agent faces a period 2 choice between 1 in the current period followed by $z_2(\gamma)$ or 0·4 in the current period followed by z_3^*. Finally, let $z_1 = \{(1 - \gamma, z_2(\gamma)) \mid \gamma \in [0{\cdot}4, 0{\cdot}8]\}$.

Suppose that the preference described above has a PP representation. Then, by definition, at any node $(1 - \gamma, 1, z_3(\gamma))$ the decision-maker chooses the consumption path $(1 - \gamma, 1, \gamma, 0)$ if $\gamma > 0{\cdot}5$ and chooses $(1 - \gamma, 1, 0, 1)$ if $\gamma \leq 0{\cdot}5$. Then it is easy to see that at any node $(1 - \gamma, z_2(\gamma))$, the decision-maker ends up with the consumption path $(1 - \gamma, 0{\cdot}4, 1, 1)$ if $0{\cdot}8 \geq \gamma > 0{\cdot}5$ and with $(1 - \gamma, 1, 0, 1)$ if $0{\cdot}4 \leq \gamma \leq 0{\cdot}5$. Note that $U_1(1 - \gamma, 0{\cdot}4, 1, 1) = 2{\cdot}2 - \gamma$ and $U_1(1 - \gamma, 1, 0, 1) = 2 - \gamma$. It follows that there is no optimal choice for the decision-maker confronting z_1 in period 1.

The example above is a version of the one presented by Peleg and Yaari (1973) who show that consistent plans (as defined in Section 3) are typically not well defined in a setting with a continuous consumption choice and $T \geq 4$.

Theorem 5 relies on two implicit assumptions. First, it utilizes the fact that the domain of decision problems is rich. Our theorem and the Peleg and Yaari example rely on being able to construct decision problems where the decision-maker in period $t - 1$ would like the period t indifference resolved one way while in period $t - 2$ he/she would like it resolved the other way. Second, our proof of non-existence takes advantage of the fact that in our definition of a consistent plan predicted behaviour in period t resolves ties in U_t in a manner that maximizes U_{t-1}. However, any alternative tie-breaking rule that depends only on the consumption history would lead to similar contradictions.

To resolve the issue of non-existence Peleg and Yaari use a model where the predicted behaviour of the decision-maker at time t depends not only on the consumption history but also on the history of decision problems that the decision-maker confronted up to time t.

While this approach does solve the existence problem, it often leads to discontinuous behaviour and preferences over decision problems. The setting with a continuous consumption choice loses much of its appeal once preferences fail to be continuous. After all, the typical motivation for working with a continuous setting is to facilitate the use of calculus. Continuity of the value function is often a necessary condition for the application of calculus based optimization techniques.[3]

We propose an extension of Strotz preferences that preserves continuity but allows for violations of NC. The representation is based on our earlier work on self-control preferences (Gul and Pesendorfer, 2001) and leads to continuous approximations without restricting the domain of decision problems.

Theorem 6 shows that self-control preferences can be used to approximate preferences with representations discussed in the previous sections of this paper. The theorem establishes that for any finite data-set (of observed choices), there exists a self-control preference over continuous consumption choices and a discrete grid of consumption choices such that the revealed Strotz preference and the restriction of the self-control preference to the grid are identical. Hence, the class of self-control preferences described below and the Strotz preferences they approximate are empirically indistinguishable and yet unlike the Strotz preferences the corresponding self-control preferences have the desirable continuity properties.

The preference $\succeq$ is a *self-control preference* if $\succeq$ can be represented by the value function W where

$$W(z_1) := \max_{\phi\in\Phi}\left\{U(\phi^T(z)) + \sum_{t=1}^{T} V_t(\phi^t(z_1)) - \sum_{t=1}^{T} \max_{\psi\in\Phi}\{V_t(\psi(\phi^{t-1}(z_1)))\}\right\} \tag{SC}$$

for some continuous functions $U : D^T \to \mathbb{R}$, $V_T : H_{T-1} \times D \to \mathbb{R}$, $V_t : H_t \times D \times \tilde{Z}_{t+1} \to \mathbb{R}$ for $t < T$. (Recall that $\phi^0(z_1) = z_1$ for all $z_1 \in \tilde{Z}_1$.) Below we refer to a preference that is represented by W satisfying (SC) as the self-control preference $(U, \bar{V})$, where $\bar{V} = (V_1, \ldots, V_T)$. Note that W is well defined and continuous since U and each V_t are continuous. The function V_t describes period t temptation. In contrast to the weak Strotz representation, the agent may choose alternatives in period t that do not maximize V_t. In that case, $V_t - \max V_t$ is the utility cost of self-control incurred in period t. Hence, $\sum_{t=1}^{T} V_t(\phi^t(z)) - \sum_{t=1}^{T} \max_{\psi\in\Phi}\left\{V_t(\psi(\phi^{t-1}(z)))\right\}$ denotes the total utility cost of self-control.

Consider a finite set of consumptions C such that $C \subset D$ and let $Z_t := K(C \times Z_{t+1})$ for $t \leq T$. Thus elements of Z_1 are the T-period decision problems when all choices are restricted to the finite set C. Let $\succeq$ be a preference defined on Z_1 and assume that $\succeq$ is a weak

[3] For this reason Harris and Laibson (2001) restrict the set of decision problems *and* utilize the strategic approach to identify a parametric class of problems in which well-behaved optimal plans exist.

Strotz preference. Theorem 6 shows that there exists a self-control preference $\succeq^*$ defined on $\tilde{Z}_1$ that coincides with $\succeq$ on Z_1.

Theorem 6. *Let $\succeq$ be a preference relation on Z_1. If $(U, \bar{V})$ is a weak Strotz representation of $\succeq$, then there exist $\alpha > 0$ and a self-control preference $(U', \bar{V}')$ on $\tilde{Z}_1$ such that $(U', \bar{V}')$ coincides with $\succeq$ on Z_1. Moreover, $U'(\bar{c}) = U(\bar{c})$ for all $\bar{c} \in C^T$ and $V_t'(\eta_{t+1}) = \alpha V_t(\eta_{t+1})$ for all η_{t+1} in the domain of V_t.*

Proof. See Section 5.

The argument for Theorem 6 is straightforward. Consider the weak Strotz preference $(U, \bar{V})$ and the self-control preference $(U, \alpha\bar{V})$ where $\bar{V} = (V_1, \dots, V_T)$ and $\alpha\bar{V} = (\alpha V_1, \dots, \alpha V_T)$. For a finite decision problem and α sufficiently large a plan that maximizes (SC) must be in $P^*(\bar{V})$, the set of consistent plans for $\bar{V}$. This follows because for α large enough a plan that maximizes (SC) must be optimal for each V_t. But in that case, the two representations yield the same preference on the finite choice set. Extending this preference to $\tilde{Z}_1$ yields the desired self-control preference.

To illustrate how self-control preferences can be used to approximate a PP preference, consider the standard β–δ utilities in Example 2:

$$U_t(c_1, \dots, c_T) = u(c_t) + \beta \sum\nolimits_{k=1}^{T-t} \delta^k u(c_{t+k})$$

for all $t = 1, \dots, T$ and

$$U(c_1, \dots, c_T) = \sum\nolimits_{t=1}^{T} \delta^t u(c_t).$$

Following Krusell, Kuruscu and Smith (2002) we can construct an approximating self-control preference as follows. Let

$$W_T(z_T) := \max\nolimits_{c \in z_T} u(c)$$

and for $t \leq T - 1$ let

$$\begin{aligned} W_{t-1}(z_t) := {} & \max\nolimits_{(c, z_{t+1}) \in z_t} \{(1+\alpha)u(c_t) + \delta(1+\alpha\beta)W_t(z_{t+1})\} \\ & - \max\nolimits_{(c, z_{t+1}) \in z_t} \alpha(u(c_t) + \beta\delta W_t(z_{t+1})). \end{aligned} \tag{SC*}$$

The value function W_0 represents a self-control preference that satisfies monotonicity (provided that u is increasing), TCC, and continuity.[4] However, it may not satisfy NC. Setting

$$V_t(c, z_{t+1}) = u(c) + \beta\delta W_t(z_{t+1})$$

and $W = W_0$ it is easily verified that (SC*) is an example of the preferences defined in (SC). Theorem 6 implies that as $\alpha \to \infty$ the self-control preference described in (SC*) approximates the β–δ preference described in Example 2.

5. PROOFS

5.1. *Theorems 1 and 2*

Let $\succeq_*$ be a complete and transitive binary relation defined on $K(X)$, the set of non-empty subsets of a finite set X. The preference $\succeq_*$ satisfies NC if for $A, B \in K(X)$, $A {\sim_*} A \cup B$ or $B {\sim_*} A \cup B$. Define the relation $R_{\succeq_*}$ on X as follows: $x R_{\succeq_*} x'$ if $\{x\} {\nsim_*} \{x'\}$ and $\{x, x'\} {\sim_*} \{x\}$ or if $\{x\} {\sim_*} \{x'\}$

[4] The parametrization (SC*) was first used by Krusell *et al.* (2002).

and $\{x', x''\} \sim_* \{x'\}$ implies $\{x, x''\} \sim_* \{x\}$. Lemma 1 below is taken from Gul and Pesendorfer (2001). It shows that $R_{\succeq_*}$ is a preference relation whenever $\succeq_*$ satisfies NC.

Lemma 1. *If $\succeq_*$ satisfies NC then $R_{\succeq_*}$ is a complete and transitive binary relation on X.*

Proof. In the proof of Lemma 1 we abbreviate the notation and write R instead of $R_{\succeq_*}$. First, we demonstrate that R is complete. If $\{x\} \nsim_* \{x'\}$ then xRx' or $x'Rx$ by NC. Suppose that $\{x\} \sim_* \{x'\}$, $\{x, \bar{x}\} \sim_* \{x\}$ and $\{x', \bar{x}\} \nsim_* \{x'\}$ for some $\bar{x}$. We need to show that $\{x', \hat{x}\} \sim_* \{x'\}$ implies $\{x, \hat{x}\} \sim_* \{x\}$. If $\{\hat{x}\} \sim_* \{x'\}$ then the result follows trivially from NC and transitivity of $\succeq$. Hence assume that $\{\hat{x}\} \nsim_* \{x'\}$. We know that $\{x, x', \bar{x}, \hat{x}\} \sim_* \{x'\}$ since both $\{x, \bar{x}\} \sim_* \{x'\}$ and $\{x', \hat{x}\} \sim_* \{x'\}$. But then it must be that either $\{x', \bar{x}\} \sim_* \{x'\}$ or $\{x, \hat{x}\} \sim_* \{x'\}$. Since the former indifference does not hold we have $\{x, \hat{x}\} \sim_* \{x'\} \sim_* \{x\}$ as desired.

To prove transitivity, let xRx' and $x'R\hat{x}$. Assume that $\{x\} \nsim_* \{x'\} \nsim_* \{\hat{x}\} \nsim_* \{x\}$. From NC it follows that $\{x, x', \hat{x}\} \sim_* \{x, x'\}$ or $\{x, x', \hat{x}\} \sim_* \{x', \hat{x}\}$. It also follows that $\{x, x', \hat{x}\} \sim_* \{x, x'\}$ or $\{x, x', \hat{x}\} \sim_* \{\hat{x}\}$. Therefore, $\{x, x', \hat{x}\} \sim_* \{x\}$. Applying NC again, we observe that $\{x, x', \hat{x}\} \sim_* \{x, \hat{x}\}$ or $\{x, x', \hat{x}\} \sim_* \{x'\}$. Since $\{x, x', \hat{x}\} \sim_* \{x\}$ we may rule out the latter case and conclude that $\{x, \hat{x}\} \sim_* \{x\}$. Since $\{x\} \nsim_* \{\hat{x}\}$ this implies $xR\hat{x}$ as desired.

Next, assume that $\{x\} \sim_* \{x'\} \sim_* \{\hat{x}\}$. Then, $\{\bar{x}, \hat{x}\} \sim_* \hat{x}$ implies $\{\bar{x}, x'\} \sim_* \{x'\}$ which in turn implies $\{\bar{x}, x\} \sim_* \{x\}$. By transitivity of $\succeq$ we have that $\{x\} \sim_* \{\hat{x}\}$ and hence $xR\hat{x}$.

Next, assume $\{x\} \sim_* \{x'\} \nsim_* \{\hat{x}\}$. Since $\{x', \hat{x}\} \sim_* \{x'\}$ it follows that $\{x, \hat{x}\} \sim_* \{x\}$. This shows $xR\hat{x}$ since by transitivity $\{x\} \nsim_* \{\hat{x}\}$.

Next, assume $\{x\} \nsim_* \{x'\} \sim_* \{\hat{x}\}$. Then $\{x\} \nsim_* \{\hat{x}\}$ and hence it is sufficient to show that $\{x, \hat{x}\} \sim_* \{x\}$. But $\{x, x'\} \nsim_* \{x'\}$ implies $\{x, \hat{x}\} \nsim_* \{\hat{x}\}$ and hence $\{x, \hat{x}\} \sim_* \{x\}$.

Finally, assume $\{x\} \sim_* \{\hat{x}\} \nsim_* \{x'\}$ then $\{x, x'\} \sim_* \{x\}$ and $\{x', \hat{x}\} \nsim_* \{\hat{x}\}$ and hence not $\hat{x}Rx$ and by completeness $xR\hat{x}$. ∥

Lemma 2. *Let $\succeq_*$ satisfy NC. If W represents $\succeq_*$ and v represents $R_{\succeq_*}$, then $W(A) = \max_{x \in A} W(\{x\})$ subject to $v(x) \geq v(x')$ for all $x' \in A$.*

Proof. Let x^* be a solution to $\max_{x \in A} W(\{x\})$ subject to $v(x) \geq v(x')$ for all $x' \in A$. Note that $z = \bigcup_{x' \in A} \{x^*, x'\}$ and since W represents NC preference $\succeq_*$, we have $W(A) = W(\{x^*, x'\})$ for some $x' \in A$. Since v represents $R_{\succeq_*}$ we have $W(\{x^*, x'\}) = W(\{x^*\})$ and hence $W(A) = W\{(x^*\}) = \max_{x \in A} W(\{x\})$ subject to $v(x) \geq v(x')$ for all $x' \in A$ as desired. ∥

Proof of Theorem 1. To prove that P and NC imply the existence of a weak Strotz representation, note that since C and hence Z_1 are finite, there exists a function W that represents $\succeq$. By Lemmas 1 and 2, there exists $V_T(h_{T-1}, \cdot)$ such that

$$\begin{aligned} &W(\{h_{T-1}, z_T\}) := \max\nolimits_{c \in C} W(\{h_{T-1}, c\}) \\ &\text{subject to } V_T(h_{T-1}, c) \geq V_T(h_{T-1}, c') \text{ for all } c' \in z_T. \end{aligned}$$

Similarly, for $t = 1, \ldots, T-1$, there exist $V_t(h_{t-1}, \cdot)$ such that

$$\begin{aligned} &W(\{h_{t-1}, z_t\}) := \max\nolimits_{(c, z_{t+1}) \in z_t} W(\{h_{t-1}, c, z_{t+1}\}) \\ &\text{subject to } V_t(h_{t-1}, c, z_{t+1}) \geq V_t(h_{t-1}, c^*, z^*_{t+1}) \text{ for all } (c^*, z^*_{t+1}) \in z_t. \end{aligned}$$

Set $\bar{V} = (V_1, \ldots, V_T)$ and $U(c_1, \ldots, c_T) = W(\{c_1, \ldots, c_T\})$. Applying a standard dynamic programming argument establishes

$$W(z_1) = \max\nolimits_{\phi \in P^*(\bar{V})} U(\phi^T(z_1))$$

as desired. The converse is straightforward and hence omitted. ∥

Proof of Theorem 2. Let $(U', \bar{V}')$ be a Strotz representation and let $(c, z_{t+1}) \in z_t$ be a choice from z_t after history $h \in H_{t-1}$. We must show that (c, z_{t+1}) is optimal for the canonical representation $(U, \bar{V})$.

Axiom NC implies that $\{h, c, z_{t+1}\} \sim \{h, \{(c, z_{t+1}), (c', z'_{t+1})\}\}$ for all $(c', z'_{t+1}) \in z_t$ since $\{h, c, z_{t+1}\} \sim \{h, z_t\}$. If $\{h, c, z_{t+1}\} \nsim \{h, c', z'_{t+1}\}$ for all $(c', z'_{t+1}) \in z_t$ then $V_t(h, c, z_{t+1}) > V_t(h, c', z'_{t+1})$ for all $(c', z'_{t+1}) \in z_t$ and the theorem follows.

If $\{h, c, z_{t+1}\} \sim \{h, c', z'_{t+1}\}$ for $(c', z'_{t+1}) \in z_t$ and $V_t(h, c', z'_{t+1}) > V_t(h, c, z_{t+1})$ then for some $(\hat{c}, \hat{z}_{t+1})$ we have $\{h, \hat{c}, \hat{z}_{t+1}\} \sim \{h, \{(c, z_{t+1}), (\hat{c}, \hat{z}_{t+1})\}, \{h, c', z'_{t+1}\}\} \sim \{h, \{(c', z'_{t+1}), (\hat{c}, \hat{z}_{t+1})\}\}$ and $\{h, c, z_{t+1}\} \nsim \{h, \hat{c}, \hat{z}_{t+1}\}$. If $\{h, c, z_{t+1}\} \succ \{h, \hat{c}, \hat{z}_{t+1}\}$ then it follows that $V'_t(h, c', z'_{t+1}) \geq V'_t(h, \hat{c}, \hat{z}_{t+1}) > V'_t(h, c, z_{t+1})$ contradicting the fact that (c, z_{t+1}) is chosen from z_t. If $\{h, \hat{c}, \hat{z}_{t+1}\} \succ \{h, c, z_{t+1}\}$ then it follows that $V'_t(h, c', z'_{t+1}) > V'_t(h, \hat{c}, \hat{z}_{t+1}) \geq V'_t(h, c, z_{t+1})$ again contradicting the fact that (c, z_{t+1}) is chosen from z_t. Hence $V_t(h, c', z'_{t+1}) \leq V_t(h, c, z_{t+1})$ for all (c', z'_{t+1}) with $\{h, c, z_{t+1}\} \sim \{h, c', z'_{t+1}\}$. But this implies that (c, z_{t+1}) is an optimal choice from z_t for the canonical representation. ||

Proof of Theorem 3. Suppose $\mathcal{C}_1, \dots, \mathcal{C}_T$ all satisfy Axiom H. Then, there exists V_t such that $V_t(h, c, z_{t+1}) \geq V_t(h, c', z'_{t+1})$ if and only if $(c, z_{t+1}) \in \mathcal{C}_t(h, \{(c, z_{t+1}), (c', z'_{t+1})\})$. Define $U : C^T \to \mathbb{R}$ so that $U(\cdot)$ represents the restriction of $\succeq$ to consumption paths. That is, $U(\bar{c}) \geq U(\bar{c}')$ iff $\{\bar{c}\} \succeq \{\bar{c}'\}$. Then, define W inductively by setting $W(\{h, z_{T+1}\}) = U(h)$ and

$$\begin{aligned}&W(\{h_{t-1}, z_t\}) := \max\nolimits_{(c, z_{t+1}) \in z_t} W(\{h_{t-1}, c, z_{t+1}\})\\ &\text{subject to } V_t(h_{t-1}, c, z_{t+1}) \geq V_t(h_{t-1}, c^*, z^*_{t+1}) \text{ for all } (c^*, z^*_{t+1}) \in z_t\end{aligned}$$

for all $t < T$. Then, IRA implies that W and hence $(U, \bar{V})$, where $\bar{V} = (V_1, \dots, V_T)$ represents $\succeq$. ||

Proof of Theorem 4. That $\succeq$ satisfies NC and TCC if it is a PP preference is obvious. To prove the converse, note that by Theorem 2, there exists a weak Strotz representation $(U, V_1, \dots, V_T)$ of $\succeq$. Let $U_t(\bar{c}) = V_t(\bar{c})$ for all $\bar{c} \in C^T$ and $\bar{U} = (U_1, \dots, U_T)$, where we identify $\bar{c}$ with the decision problem in which the agent is committed to the consumption path $\bar{c}$. Recall the definitions of W and V_t's from Theorem 2 and Lemma 1. In order to avoid having to make separate statements for the cases of $t = T$ and $t < T$, let (c_T, z_{T+1}) denote (c_T).

First, we prove that for all $t \leq T - 1$,

$$\begin{aligned}&W(h_{t-1}, c, z_{t+1}) \geq W(h_{t-1}, c, z'_{t+1}) \text{ iff}\\ &V_t(h_{t-1}, c, z_{t+1}) \geq V_t(h_{t-1}, c, z'_{t+1}).\end{aligned} \tag{3}$$

If $W(h_{t-1}, c, z_{t+1}) > W(h_{t-1}, c, z'_{t+1})$ then TCC implies $\{(h_{t-1}, c, z_{t+1}), (h_{t-1}, c, z'_{t+1})\} \sim \{(h_{t-1}, c, z_{t+1})\}$ establishing that $V_t(h_{t-1}, c, z_{t+1}) > V_t(h_{t-1}, c, z'_{t+1})$. If $W(h_{t-1}, c, z_{t+1}) = W(h_{t-1}, c, z'_{t+1})$ let (h_{t-1}, c, z''_t) satisfy $\{(h_{t-1}, c, z_{t+1}), (h_{t-1}, c, z''_t)\} \sim \{(h_{t-1}, c, z_{t+1})\}$. If $\{(h_{t-1}, c, z''_t)\} \sim \{(h_{t-1}, c, z_{t+1})\}$ then $\{(h_{t-1}, c, z'_{t+1}), (h_{t-1}, c, z''_t)\} \sim \{(h_{t-1}, c, z'_{t+1})\}$ by NC. On the other hand, if $\{(h_{t-1}, c, z''_t)\} \nsim \{(h_{t-1}, c, z_{t+1})\}$ then we have by TCC that $\{(h_{t-1}, c, z_{t+1}), (h_{t-1}, c, z'_{t+1}), (h_{t-1}, c, z''_t)\} \sim \{(h_{t-1}, c, z''_t), (h_{t-1}, c, z'_{t+1})\}$ and by NC $\{(h_{t-1}, c, z_{t+1}), (h_{t-1}, c, z'_{t+1}), (h_{t-1}, c, z''_t)\} \sim \{(h_{t-1}, c, z'_{t+1})\}$ follows, establishing that

$$\{(h_{t-1}, c, z'_{t+1}), (h_{t-1}, c, z''_t)\} \sim \{(h_{t-1}, c, z'_{t+1})\}.$$

That is, $V_t(h_{t-1}, c, z'_{t+1}) \geq V_t(h_{t-1}, c, z_{t+1})$. Then, $V_t(h_{t-1}, c, z'_{t+1}) = V_t(h_{t-1}, c, z_{t+1})$ follows from a symmetric argument. We have therefore established (3).

Next, we observe that, for all $t \leq T - 1$,

$$\begin{aligned} &V_t(h_t, z_{t+1}) = \max_{(c_{t+1}, z_{t+2}) \in z_{t+1}} V_t(h_t, c_{t+1}, z_{t+2}) \\ &\text{subject to } V_{t+1}(h_t, c_{t+1}, z_{t+2}) \geq V_{t+1}(h_t, c'_{t+1}, z'_{t+2}) \end{aligned} \tag{4}$$

for all $(c'_{t+1}, z'_{t+2}) \in z_{t+1}$. To see this, note that

$$\begin{aligned} &W(h_t, z_{t+1}) = \max_{(c_{t+1}, z_{t+2}) \in z_{t+1}} W(h_t, c_{t+1}, z_{t+2}) \\ &\text{subject to } V_{t+1}(h_t, c_{t+1}, z_{t+2}) \geq V_{t+1}(h_t, c'_{t+1}, z'_{t+2}) \end{aligned} \tag{5}$$

for all $(c'_{t+1}, z'_{t+2}) \in z_{t+1}$ follows from the definition of W and V_{t+1}. Then, (3) and (5) yield (4).

It follows from (4) (and induction) that

$$V_t(\eta_{t+1}) = \max_{\phi \in P_{t+1}(\bar{U})} V_t(\phi^T(\eta_{t+1})). \tag{6}$$

To conclude the proof we show by induction that

$$W(\eta_t) = \max_{\phi \in P_t(\bar{U})} U_t(\phi^T(\eta_t)) \tag{7}$$

for all $t \leq T$ and $\eta_t \in N_t$. For $t = T$, (7) follows from (4) and the fact that $U_T = V_T$. Suppose that the result is true for $t + 1$. Then, (7) follows from (4)–(6) and our induction hypothesis. ||

Proof of Theorem 5. Let 1_t denote the t-period history $h = (1, \ldots, 1)$. Let $a, b \in (0, 1)^2$ denote consumption vectors for periods $T - 1$ and T. By local preference for commitment there exist $a, b \in (0, 1)^2$ such that

$$U_{T-2}(1_{T-2}, a) < U_{T-2}(1_{T-2}, b) \qquad \text{and} \qquad U_{T-1}(1_{T-2}, a) > U_{T-1}(1_{T-2}, b).$$

Also by local preference for commitment, we can choose $c_1, c_2 \in (0, 1]$ and $b^1, b^2 \in (0, 1)^2$ such that

$$\begin{aligned} &U_{T-3}(1_{T-3}, c_1, b^1) > U_{T-3}(1_{T-3}, c_2, b^2) \qquad \text{and} \\ &U_{T-2}(1_{T-3}, c_1, b^1) < U_{T-2}(1_{T-3}, c_2, b^2). \end{aligned}$$

Moreover, we can choose b^1 and b^2 arbitrarily close to b and c_1, c_2 arbitrarily close to 1. Hence, by the continuity of U_{T-2} we can choose c_1, c_2, b^1, b^2 such that

$$\begin{aligned} &U_{T-2}(1_{T-3}, c_2, b^2) > U_{T-2}(1_{T-3}, c_1, b^1) > U_{T-2}(1_{T-3}, c_2, a) \qquad \text{and} \\ &U_{T-1}(1_{T-3}, c_2, b^2) < U_{T-1}(1_{T-3}, c_2, a). \end{aligned}$$

By monotonicity, there exists $\lambda \in (0, 1)$ such that

$$U_{T-1}(1_{T-3}, c_2, b^2) = U_{T-1}(1_{T-3}, c_2, \lambda a).$$

Of course, we still have

$$U_{T-2}(1_{T-3}, c_2, b^2) > U_{T-2}(1_{T-3}, c_1, b^1) > U_{T-2}(1_{T-3}, c_2, \lambda a).$$

By continuity and monotonicity, for $\varepsilon > 0$ sufficiently small, there exists a unique $\gamma(\varepsilon) \in (0, 1)$ such that

$$\begin{aligned} &U_{T-1}(1_{T-4}, 1 - \varepsilon, c_2, b^2) = U_{T-1}(1_{T-4}, 1 - \varepsilon, c_2, \gamma(\varepsilon)a) - \varepsilon \\ &U_{T-2}(1_{T-4}, 1 - \varepsilon, c_2, b^2) > U_{T-2}(1_{T-4}, 1 - \varepsilon, c_1, b^1) > U_{T-2}(1_{T-4}, 1 - \varepsilon, c_2, \gamma(\varepsilon)a) \\ &\quad U_{T-3}(1_{T-4}, 1, c_2, b^2) < U_{T-3}(1_{T-4}, 1 - \varepsilon, c_1, b^1). \end{aligned}$$

Choose $\alpha > 0$ small enough such that for all $\varepsilon \in (0, \alpha]$ there exists $\gamma(\varepsilon)$ satisfying all of the above (in)equalities. Note that $\gamma(\cdot)$ is a continuous function and $\lim_{\varepsilon \to 0} \gamma(\varepsilon) = \lambda$.

For $\varepsilon \le \alpha$, let $z_{T-1}(\varepsilon)$ denote the decision problem in which the decision-maker chooses between $b^2 \in (0,1)^2$ and $\gamma(\varepsilon)a \in (0,1)^2$ in period $T-1$. Let z^*_{T-1} denote the decision problem that commits the agent to the consumption $b^1 \in (0,1)^2$ for the last two periods. Let $z_{T-2}(\varepsilon) = \{(c_2, z_{T-1}(\varepsilon))\}$ and $z^*_{T-2} = \{c_1, z^*_{T-1}\}$. Finally define

$$z_{T-3}(\varepsilon) = \{(1-\varepsilon, z_{T-2}(\varepsilon) \cup z^*_{T-2} \mid \varepsilon' \in [0,\varepsilon]\}.$$

Since γ is continuous and $[0,\varepsilon]$ is compact, $z_{T-3}(\varepsilon) \in \tilde{Z}_{T-3}$. It is easy to verify that by choosing any $\varepsilon' \in (0,\varepsilon]$ in period $T-3$ the decision-maker ends up with the consumption path $(1_{T-4}, 1-\varepsilon', c_1, b^1)$ while by choosing $\varepsilon' = 0$ he/she ends up with $(1_{T-3}, c_2, b^2)$. Hence, no optimal choice exists for the decision-maker in period $T-3$. ||

Proof of Theorem 6. Define

$$\tilde{U}_{ht} = \max_{C\times Z_{t+1}} U(h,c,z_{t+1}) - \min_{C\times Z_{t+1}} U(h,c,z_{t+1})$$

and define $\tilde{U} = \max_t \max_{H_t} \tilde{U}_{ht}$. Let

$$\begin{aligned}\underline{V_t} = \min_{(C\times Z_{t+1})\times(C\times Z_{t+1})}\{V_t(h,c,z_{t+1}) - V_t(h,c',z'_{t+1})\}\\ \text{subject to } V_t(h,c,z_{t+1}) > V_t(h,c',z'_{t+1})\end{aligned}$$

and let $\underline{V} = \min_t \min_{H_t} \tilde{V}_{ht}$. Choose α so that $\alpha\underline{V} > \tilde{U}$. Then, a simple inductive argument ensures that for all $z_1 \in Z_1$, the set of solutions to

$$\max_{\phi\in\Phi}\left\{U(\phi^T(z_1)) + \sum\nolimits_{t=1}^{T} \alpha V_t(\phi^t(z_t)) - \sum\nolimits_{t=1}^{T} \max_{\psi\in\Phi}\{\alpha V_t(\psi(\phi^{t-1}(z_1)))\}\right\}$$

coincides with the set of solutions to

$$\max_{\phi\in P^*(\bar{V})} U(\phi^T(z_1)).$$

Therefore, the self-control preference $(U, \alpha V_1, \dots, \alpha V_T)$ represents the weak Strotz preference $(U, V_1, \dots, V_T)$ on Z_1.

It remains to show that we can extend the self-control preference $(U, \alpha V_1, \dots, \alpha V_T)$ to $\tilde{Z}_1$. Since Z_1 is finite, continuous extensions of the functions $U, \alpha V_1, \dots, \alpha V_T$ to $\tilde{Z}_1$ are possible. ||

Acknowledgements. This research was supported by grants SES9911177, SES0236882, SES9905178 and SES0214050 from the National Science Foundation.

REFERENCES

GUL, F. and PESENDORFER, W. (2001), “Temptation and Self-Control”, *Econometrica*, **69**, 1403–1435.

GUL, F. and PESENDORFER, W. (2004), “Self-Control and the Theory of Consumption”, *Econometrica*, **72**, 119–158.

HARRIS, C. and LAIBSON, D. (2001), “Dynamic Choices of Hyperbolic Consumers”, *Econometrica*, **69**, 935–958.

KREPS, D. M. (1979), “A Representation Theorem for ‘Preference for Flexibility”’, *Econometrica*, **47**, 565–576.

KRUSELL, P., KURUSCU, B. and SMITH, T. Jr. (2002), “Temptation and Taxation” (Mimeo).

LAIBSON, D. (1997), “Golden Eggs and Hyperbolic Discounting”, *Quarterly Journal of Economics*, **112**, 443–477.

O’DONOGHUE, T. and RABIN, M. (1999), “Doing it Now or Later”, *American Economic Review*, **89**, 103–124.

PELEG, M. and YAARI, M. E. (1973), “On the Existence of a Consistent Course of Action when Tastes are Changing”, *Review of Economic Studies*, **40**, 391–401.

PHELPS, E. S. and POLLAK, R. A. (1968), “On Second-Best National Saving and Game-Equilibrium Growth”, *Review of Economic Studies*, **35**, 185–199.

POLLAK, R. A. (1968), “Consistent Planning”, *Review of Economic Studies*, **35**, 201–208.

STROTZ, R. H. (1955), “Myopia and Inconsistency in Dynamic Utility Maximization”, *Review of Economic Studies*, **23**, 165–180.