Naive Quasi-Hyperbolic Discounting (O'Donoghue and Rabin)

Todd Sarver

Econ 885 – Duke University

Outline

Quasi-Hyperbolic Discounting

A simple extension of the exponential discounting model that permits a present bias is the quasi-hyperbolic discounting model of Phelps and Pollak (1968) and Laibson (1997):

$$\begin{aligned} U_t &= u_t + \beta \sum_{\tau=t+1}^{\infty} \delta^{\tau-t} u_\tau \\ &= u_t + \beta\delta u_{t+1} + \beta\delta^2 u_{t+2} + \beta\delta^3 u_{t+3} + \dots \end{aligned} \quad (1)$$

where $\beta, \delta \in (0, 1)$. The preferences represented by this utility function are also referred to as the (β, δ) preferences.

Quasi-Hyperbolic Discounting: Special Case

- O'Donoghue and Rabin (1999) assumed a finite time horizon, so adopt the finite-horizon version of the (β, δ) model in Equation (1) with T periods. Then, for every $1 \leq t \leq T$,

$$U_t = u_t + \beta \sum_{\tau=t+1}^{T} \delta^{\tau-t} u_\tau.$$

- They also assumed that $\delta = 1$. This assumption greatly simplifies the analysis, and we will see that it will be without loss of generality for the particular choice problems that they consider. Thus,

$$\begin{aligned} U_t &= u_t + \beta \sum_{\tau=t+1}^{T} u_\tau \\ &= u_t + \beta u_{t+1} + \cdots + \beta u_T. \end{aligned}$$

Decision Problem: "Doing It Once"

O'Donoghue and Rabin (1999) considered a single activity that needs to be completed once. The costs and benefits may change with the dates completed.

- Let $\mathbf{v} = (v_1, v_2, \ldots, v_T)$ be the reward schedule.
- Let $\mathbf{c} = (c_1, c_2, \ldots, c_T)$ be the cost schedule.
- Thus completing the activity in period t yields reward v_t and cost c_t, where $v_t \geq 0$ and $c_t \geq 0$ for all t.
- Note that the subscripts correspond to the date of task completion, but are not necessarily the date at which the rewards and costs are experienced: Some activities may incur delayed costs or benefits.

Decision Problem: "Doing It Once"

O'Donoghue and Rabin (1999) considered two possibilities for the timing of costs and rewards:

1. Immediate Costs: The cost c_t is incurred in the period t of task completion and the reward v_t is experienced in a later period. (Note that the exact future period after t in which the reward is experienced is irrelevant since $\delta = 1$.) The individual's utility from the perspective of period t of completing the activity in period $\tau \geq t$ is therefore

$$U_t(\tau; \beta) = \begin{cases} \beta v_\tau - c_\tau & \text{if } \tau = t \\ \beta v_\tau - \beta c_\tau & \text{if } \tau > t. \end{cases}$$

2. Immediate Rewards: The reward v_t is experienced in the period t of task completion and the cost c_t is incurred in a later period. Utility in this case is

$$U_t(\tau; \beta) = \begin{cases} v_\tau - \beta c_\tau & \text{if } \tau = t \\ \beta v_\tau - \beta c_\tau & \text{if } \tau > t. \end{cases}$$

Outline

Preferences and Choice Setting

Sophistication and Naivete

- Perception-Perfect Strategies
- Procrastination and Preproperation
- Welfare

Partial Naivete

- Perception-Perfect Strategies
- Continuity of Behavior in Beliefs

Outline

Preferences and Choice Setting

Sophistication and Naivete

- Perception-Perfect Strategies
- Procrastination and Preproperation
- Welfare

Partial Naivete

- Perception-Perfect Strategies
- Continuity of Behavior in Beliefs

Sophistication and Naivete

O'Donoghue and Rabin (1999) examine the behavior of three types of individuals:

- A time-consistent individual has $\beta = 1$.
- A sophisticated time-inconsistent individual has $\beta < 1$ and recognizes that she will make choices to maximize her (β, δ) preferences in every period. Thus, she realizes that she will be dynamically inconsistent.
- A naive time-inconsistent individual has $\beta < 1$ but incorrectly believes she will be dynamically consistent going forward. She has a (false) belief that in the future she will use exponential discounting.
    - Has $\beta < 1$ in the current period, but incorrectly believes she will have $\beta = 1$ for preferences in future periods.

Game Between Multiple Selves

For time-inconsistent preferences, dynamic choices can be modeled as a game between multiple selves:

- Think of the individual as a sequence of different selves acting at each of the time periods.
- Then, describe the strategies and beliefs of each of these selves, and look for the equilibrium play between them.

What does "equilibrium play" look like?

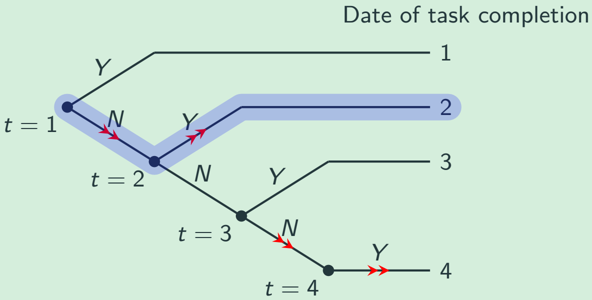
- In the case of a sophisticated time-inconsistent individual, the solution concept will be precisely subgame perfect equilibrium.
- For a naive individual, we require a modified solution concept that takes into account her incorrect beliefs about future play.

Game Between Multiple Selves: Strategies

A strategy is a vector $\mathbf{s} = (s_1, s_2, \dots, s_T)$, where $s_t \in \{Y, N\}$ specifies whether or not to do the activity in period t given she has not yet done it. Since the activity must be completed in one of the periods, require that $s_T = Y$.

Example

$T = 4$ and $\mathbf{s} = (N, Y, N, Y)$.



Perception-Perfect Strategies

O'Donoghue and Rabin (1999) define a perception-perfect strategy as a strategy such that in all periods the person chooses the optimal action given her current preferences and her perceptions of future behavior. This solution concept can be applied to each of the three types of individuals—time-consistent (TC), naifs, and sophisticates:

Definition

A perception-perfect strategy (PPS) is defined as follows:

1. A PPS for TCs is a strategy $\mathbf{s}^{tc} = (s_1^{tc}, s_2^{tc}, \ldots, s_T^{tc})$ such that for all $t < T$, $s_t^{tc} = Y$ if and only if $U_t(t; 1) \geq U_t(\tau; 1)$ for all $\tau > t$.
2. A PPS for naifs is a strategy $\mathbf{s}^n = (s_1^n, s_2^n, \ldots, s_T^n)$ such that for all $t < T$, $s_t^n = Y$ if and only if $U_t(t; \beta) \geq U_t(\tau; \beta)$ for all $\tau > t$.
3. A PPS for sophisticates is a strategy $\mathbf{s}^s = (s_1^s, s_2^s, \ldots, s_T^s)$ such that for all $t < T$, $s_t^n = Y$ if and only if $U_t(t; \beta) \geq U_t(\tau'; \beta)$ where $\tau' = \min\{\tau > t : s_\tau^s = Y\}$.

Perception-Perfect Strategies

- Note that the definition of a perception-perfect strategy is the same for TCs and naifs because both believe that in the future they will follow the optimal plan made today. The strategies for these two types of preferences will be different, however, since $U_t(\tau; \beta)$ is different for $\beta = 1$ and $\beta < 1$.

- The PPS for a sophisticate is a subgame perfect equilibrium where each self optimizes given the strategy of future selves. (Note that a particular SPE in the game between selves is selected—the one where each self completes the activity if indifferent.) This strategy can be derived through backward induction using preferences in future periods.

The equilibrium completion date (period when the activity is completed) is the main issue in the analysis. These periods are denoted τ_{tc}, τ_n, and τ_s, respectively, and are defined by

$$\tau_a = \min\{t : s_t^a = Y\} \quad \text{for} \quad a \in \{tc, n, s\}.$$

Outline

Procrastination and Preproperation

- O'Donoghue and Rabin (1999) use the term procrastination to mean that an agent chooses to wait when a time-consistent individual with the same δ (which they argue is the "long-run self") would choose to do it.
- They use preproperation to mean that an agent chooses to do it when her long-run self would choose to wait. In this section, we show how procrastination and preproperation can arise in this model.

Example: Procrastination

The following example could represent the costs and benefits of completing a term paper when the opportunity cost of time is increasing as the deadline approaches.

Example (Procrastination)

Suppose costs are immediate, $T = 4$, and $\beta = \frac{1}{2}$ for naifs and sophisticates. Let $\mathbf{v} = (\bar{v}, \bar{v}, \bar{v}, \bar{v})$ and $\mathbf{c} = (3, 5, 8, 13)$.

- $\mathbf{s}^{tc} = (Y, Y, Y, Y)$ and $\tau_{tc} = 1$.
- $\mathbf{s}^{n} = (N, N, N, Y)$ and $\tau_{n} = 4$.
- $\mathbf{s}^{s} = (N, Y, N, Y)$ and $\tau_{s} = 2$.

The following example illustrates how the main argument behind the previous example can be generalized to any finite number of periods.

Example (Procrastination)

Suppose costs are immediate and $\beta = \frac{1}{2}$ for naifs and sophisticates. Let $v_t = \bar{v}$ and $c_t = (3/2)^t$ for $t \in \{1, \ldots, T\}$.

- $\mathbf{s}^{tc} = (Y, Y, \ldots, Y, Y)$ and $\tau_{tc} = 1$.
- $\mathbf{s}^{n} = (N, N, \ldots, N, Y)$ and $\tau_{n} = T$.
- If T is even, then $\mathbf{s}^{s} = (N, Y, N, \ldots, N, Y)$ and $\tau_{s} = 2$.
- If T is odd, then $\mathbf{s}^{s} = (Y, N, Y, \ldots, N, Y)$ and $\tau_{s} = 1$.

Unrealistic Behavior in this Example?

The alternating pattern in the PPS for sophisticates seems a bit strange and potentially unrealistic.

- Part of the unintuitive conclusion might be attributed to the solution concept. Recall there are a number of examples of implausible implications of subgame perfect equilibrium in standard game theory (e.g., the centipede game).
- Another factor in this example may be the starkness of the (deterministic, fixed-horizon) decision problem. A later paper by Carroll et al. (2009) introduces uncertainty and takes the infinite horizon version of this type of decision problem, and it obtains more intuitive conclusions (stationary threshold strategies).

Example: Preproperation

Consider now an example of immediate rewards, such as choosing a weekend to go to the movies.

Example (Preproperation)

Suppose rewards are immediate, $T = 4$, and $\beta = \frac{1}{2}$ for naifs and sophisticates. Let $\mathbf{v} = (3, 5, 8, 13)$ and $\mathbf{c} = (0, 0, 0, 0)$.

- $\mathbf{s}^{tc} = (N, N, N, Y)$ and $\tau_{tc} = 4$.
- $\mathbf{s}^{n} = (N, N, Y, Y)$ and $\tau_{n} = 3$.
- $\mathbf{s}^{s} = (Y, Y, Y, Y)$ and $\tau_{s} = 1$.

Results

These examples are suggestive of some general patterns, which are formalized in the following result.

Theorem

1. *If costs are immediate, then* $\tau_{tc} \leq \tau_n$.
2. *If rewards are immediate, then* $\tau_n \leq \tau_{tc}$.
3. *In both cases,* $\tau_s \leq \tau_n$.

Proof.

(1): When costs are immediate, for any period, if a naif would complete the activity then so would a TC. This follows from simple algebra:

$$\beta v_t - c_t \geq \beta v_\tau - \beta c_\tau \implies v_t - c_t \geq v_\tau - c_\tau .$$

The intuition is that the naif assigns the same relative weights all costs and benefits the same as the TC, with the exception of the immediate cost, which the naif overweights.

(2): When rewards are immediate, for any period, if a TC would complete the activity then so would a naif. The intuition is the same as part (1), except in this case the naif assigns higher relative weight to present benefits.

(3): A naif will only complete the activity in period t if $U_t(t;\beta) \geq U_t(\tau;\beta)$ for all $\tau > t$. The sophisticate will do it in period t if the same inequality holds just for $\tau' = \min\{\tau > t : s^s_\tau = Y\}$. Intuitively, the naif is too optimistic about the period in which she will complete the activity in the future and therefore overweights delay. □

Example: Preproperation to Avoid Procrastination

Note that in the case of immediate costs, the theorem does not exclude the possibility that $\tau_s < \tau_{tc} < \tau_n$. The following example shows how this ordering is possible. When a sophisticate correctly predicts that in the next period she will delay the activity, she may complete it early to prevent excessive procrastination.

Example (Preproperation to avoid procrastination)

Suppose costs are immediate, $T = 3$, and $\beta = \frac{1}{2}$ for naifs and sophisticates. Let $\mathbf{v} = (12, 18, 18)$ and $\mathbf{c} = (3, 8, 13)$.

- $\mathbf{s}^{tc} = (N, Y, Y)$ and $\tau_{tc} = 2$.
- $\mathbf{s}^{n} = (N, N, Y)$ and $\tau_{n} = 3$.
- $\mathbf{s}^{s} = (Y, N, Y)$ and $\tau_{s} = 1$.

Outline

Example: Welfare Loss from Naivete

A sufficient condition for a naive individual to procrastinate is

$$-c_t + \beta v_t < -\beta c_{t+1} + \beta v_{t+1}.$$

- If $v_t = v_{t+1} = \bar{v}$, this simplifies to

$$c_{t+1} < \left(\frac{1}{\beta}\right) c_t.$$

- If $c_t = c_{t+1} = \bar{c}$, this simplifies to

$$v_{t+1} > v_t - \left(\frac{1-\beta}{\beta}\right) \bar{c}.$$

Using these conditions, we can construct an example that shows how large the negative effects of procrastination can be.

Example: Welfare Loss from Naivete

Example

Fix any $c_1, v_1 > 0$. Also fix some (large) $\bar{X}$ to be the target cost we would like to show the agent could ultimately be made to pay. Take any growth rate $1 < g < \frac{1}{\beta}$ and let

$$c_t = \begin{cases} g^{t-1}c_1 & \text{if } t < T' \\ \bar{X} & \text{if } t \geq T', \end{cases}$$

where T' is the smallest integer that satisfies $g^{T'-1}c_1 \geq \bar{X}$. Similarly, take any $0 < L < \frac{1-\beta}{\beta}\bar{X}$ and let

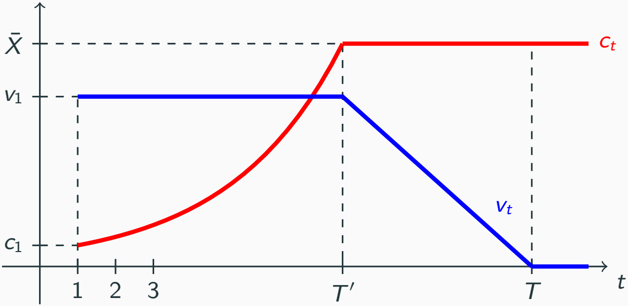
$$v_t = \begin{cases} v_1 & \text{if } t \leq T' \\ v_1 - L(t - T') & \text{if } T' < t < T \\ 0 & \text{if } t = T, \end{cases}$$

where T is the smallest integer such that $v_1 - L(T - T') \leq 0$.

Illustration

$$1 < g < \frac{1}{\beta} \qquad c_t = \begin{cases} g^{t-1} c_1 & \text{if } t < T' \\ \bar{X} & \text{if } t \geq T', \end{cases}$$

$$0 < L < \frac{1-\beta}{\beta}\bar{X} \qquad v_t = \begin{cases} v_1 & \text{if } t \leq T' \\ v_1 - L(t - T') & \text{if } T' < t < T \\ 0 & \text{if } t = T, \end{cases}$$



Welfare Impacts of Time-Inconsistency and Naivete

O'Donoghue and Rabin (1999) define the individual's long-run utility from completing the activity in period τ by $U_0(\tau) = v_\tau - c_\tau$.

Theorem (Welfare: immediate costs case)

Suppose costs are immediate, and consider all $\mathbf{v}$ *and* $\mathbf{c}$ *such that* $v_t \leq \bar{X}$ *and* $c_t \leq \bar{X}$ *for all* t*:*

1. $\lim_{\beta \to 1} \sup_{(\mathbf{v},\mathbf{c})} [U_0(\tau_{tc}) - U_0(\tau_s)] = 0$.
2. *For any* $\beta < 1$, $\sup_{(\mathbf{v},\mathbf{c})} [U_0(\tau_{tc}) - U_0(\tau_n)] = 2\bar{X}$.

- Part 2 follows from the previous example, taking $c_1 = \varepsilon$ and $v_1 = \bar{X}$.
- Note that this example can also be interpreted using other welfare criteria, such as Pareto efficiency between selves (all selves agree that completing at date $t = 1$ is better than $t = T$).

Welfare Impacts of Time-Inconsistency and Naivete

Theorem (Welfare: immediate rewards case)

Suppose rewards are immediate, and consider all $\mathbf{v}$ *and* $\mathbf{c}$ *such that* $v_t \leq \bar{X}$ *and* $c_t \leq \bar{X}$ *for all* t*:*

1. $\lim_{\beta \to 1} \sup_{(\mathbf{v},\mathbf{c})}[U_0(\tau_{tc}) - U_0(\tau_n)] = 0$.
2. *For any* $\beta < 1$, $\sup_{(\mathbf{v},\mathbf{c})}[U_0(\tau_{tc}) - U_0(\tau_s)] = 2\bar{X}$.

Outline

Preferences and Choice Setting

Sophistication and Naivete

- Perception-Perfect Strategies
- Procrastination and Preproperation
- Welfare

Partial Naivete

- Perception-Perfect Strategies
- Continuity of Behavior in Beliefs

Outline

Preferences and Choice Setting

Sophistication and Naivete

- Perception-Perfect Strategies
- Procrastination and Preproperation
- Welfare

Partial Naivete

- Perception-Perfect Strategies
- Continuity of Behavior in Beliefs

Partial Naivete

- In the previous section, an individual either had correct beliefs about her future present-bias parameter β (sophistication) or incorrectly believed that she would be time consistent in all future periods (naivete).
- In this section we consider the possibility that the individual understands that she will be time inconsistent in her future choices, but she potentially underestimates the degree of her future present bias (partial naivete).

Definition

An individual with present-bias factor $\beta < 1$ is partially naive if she believes that her future present-bias factor will be $\hat{\beta} \in (\beta, 1)$. She is sophisticated if $\hat{\beta} = \beta$ and is (fully) naive if $\hat{\beta} = 1$. We refer to quasi-hyperbolic discounting with these partially-naive beliefs as $(\beta, \hat{\beta}, \delta)$ preferences.

Choice Environment

- O'Donoghue and Rabin (2001) studied partially-naive preferences within a stationary infinite-horizon environment where the agent decides in each period whether to complete one of a fixed set of possible tasks. Moreover, the agent completes at most one task and can complete that task at most once.

- In this section, we instead study partially naive preferences within the finite-horizon environment from O'Donoghue and Rabin (1999).

As in the previous section, we assume there is a single task that can completed, and we allow the costs and benefits to change with the dates completed. As before:

- Let $\mathbf{v} = (v_1, v_2, \ldots, v_T)$ be the reward schedule.

- Let $\mathbf{c} = (c_1, c_2, \ldots, c_T)$ be the cost schedule.

- Completing the activity in period t yields reward v_t and cost c_t, where $v_t \geq 0$ and $c_t \geq 0$ for all t.

Choice Environment

- We will focus on immediate cost case and again take $\delta = 1$. Thus the cost c_t is incurred in the period t of task completion, but the reward v_t arrives in a later period.

- The individual's utility from the perspective of period t of completing the activity in period $\tau \geq t$ is therefore

$$U_t(\tau; \beta) = \begin{cases} \beta v_\tau - c_\tau & \text{if } \tau = t \\ \beta v_\tau - \beta c_\tau & \text{if } \tau > t. \end{cases}$$

- Given her partially-naive beliefs, the agent believes that preferences in future periods will be given by $U_t(\tau; \hat{\beta})$, where $\beta \leq \hat{\beta} \leq 1$.

Perception-Perfect Strategies

We now define perception-perfect strategies for partially-naive agents. We will again assume that the task must be completed in the final period, so $s_T = Y$ and $\hat{s}_T = Y$.[1]

Definition

Suppose $\beta \leq \hat{\beta} \leq 1$. A strategy $\mathbf{s} = (s_1, s_2, \ldots, s_T)$ is a perception-perfect strategy if there exist beliefs $\hat{\mathbf{s}} = (\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_T)$ that satisfy:

1. The belief $\hat{\mathbf{s}}$ is a PPS for sophisticated $(\hat{\beta}, \delta)$ preferences. That is, for all $t < T$, $\hat{s}_t = Y$ if and only if $U_t(t; \hat{\beta}) \geq U_t(\tau'; \hat{\beta})$ where $\tau' = \min\{\tau > t : \hat{s}_\tau = Y\}$.
2. For all $t < T$, $s_t = Y$ if and only if $U_t(t; \beta) \geq U_t(\tau'; \beta)$ where $\tau' = \min\{\tau > t : \hat{s}_\tau = Y\}$.

[1] However, this definition can also be applied to the case where never completing the task is permitted by setting the utility from never completing equal to zero and by taking $s_T = Y$ if and only if $U_T(T; \beta) \geq 0$, and $\hat{s}_T = Y$ if and only if $U_T(T; \hat{\beta}) \geq 0$.

Example (Procrastination, revisited)

Suppose $\beta = \frac{1}{2}$. Let $\mathbf{v} = (\bar{v}, \bar{v}, \bar{v}, \bar{v})$ and $\mathbf{c} = (3, 5, 8, 13)$.

- $\frac{5}{8} \leq \hat{\beta} \leq 1$:[2]
  - $\hat{\mathbf{s}} = (Y, Y, Y, Y)$
  - $\mathbf{s} = (N, N, N, Y)$
- $\frac{8}{13} \leq \hat{\beta} < \frac{5}{8}$:
  - $\hat{\mathbf{s}} = (Y, N, Y, Y)$
  - $\mathbf{s} = (Y, N, N, Y)$
- $\frac{3}{5} \leq \hat{\beta} < \frac{8}{13}$:
  - $\hat{\mathbf{s}} = (Y, Y, N, Y)$
  - $\mathbf{s} = (N, Y, N, Y)$
- $\frac{1}{2} \leq \hat{\beta} < \frac{3}{5}$:
  - $\hat{\mathbf{s}} = (N, Y, N, Y)$
  - $\mathbf{s} = (N, Y, N, Y)$

[2] Note that: $\frac{3}{5} = 0.6 < \frac{8}{13} \approx 0.615 < \frac{5}{8} = 0.625$

Dynamically Consistent Beliefs

- Note that the beliefs about future play are forced to be consistent across time periods in this definition. That is, for $\tau > t' > t$, the belief about the strategy in period τ is $\hat{s}_\tau$ in both periods t and t'.

- O'Donoghue and Rabin (2001) begin with a more general definition that permits a potentially different set of beliefs in each period, so the beliefs $\hat{\mathbf{s}}^t$ in period t may disagree with the beliefs $\hat{\mathbf{s}}^{t'}$ in period t' about the play in some future period.

- However, they then rule out this possibility in their definition of the perception-perfect strategy: They require that beliefs in each period must be "dynamically consistent," meaning that $\hat{s}^t_\tau = \hat{s}^{t'}_\tau$ for any $\tau > t' > t$, just as we have assumed here.

Importance of Dynamically Consistent Beliefs

- In the infinite-horizon environment of O'Donoghue and Rabin (2001), it is easy to create examples where the individual prefers to delay one period, but would rather complete the task immediately than wait two periods.
- Now, suppose that the individual is sophisticated, so $\beta = \hat{\beta}$. We should expect the PPS to be an alternating sequences of Y and N.
- However, we can nonetheless end up with endless procrastination if we permit beliefs that are not dynamically consistent: Let the belief be $\hat{\mathbf{s}}^t = (Y, N, Y, N, \ldots)$ in even periods t, and $\hat{\mathbf{s}}^t = (N, Y, N, Y, \ldots)$ in odd periods t. Both of these are PPS for sophisticated $(\hat{\beta}, \delta)$ preferences under the assumptions provided.
- Since the beliefs alternate across periods (the agent thinks she is playing a different equilibrium in even periods than in odd periods), in every period the agent believes that she will complete the task in the next period. Since she is prefers one period of delay, her optimal strategy is then $\mathbf{s} = (N, N, N, \ldots)$.
- The assumption of dynamically-consistent beliefs rules out this possibility.

Outline

Continuity of Behavior as Naivete Vanishes

The next two results illustrate continuity of behavior in the limit as $\hat{\beta}$ gets close to β.

Lemma

Fix any $\mathbf{v}$, $\mathbf{c}$, *and* β. *There exists* $\varepsilon > 0$ *such that for any* $\hat{\beta} \in [\beta, \beta + \varepsilon)$ *and any* t *and* $\tau > t$*:*

1. $U_t(t; \beta) \geq U_t(\tau; \beta) \implies U_t(t; \hat{\beta}) \geq U_t(\tau; \hat{\beta})$.
   (But equality on the left may correspond to strict inequality on the right.)
2. $U_t(t; \beta) < U_t(\tau; \beta) \implies U_t(t; \hat{\beta}) < U_t(\tau; \hat{\beta})$.

Proof.

1. The first part is true for any $\hat{\beta} \geq \beta$:

$$\begin{aligned} U_t(t;\beta) \geq U_t(\tau;\beta) &\iff \beta v_t - c_t \geq \beta v_\tau - \beta c_\tau \\ &\iff v_t - \frac{1}{\beta} c_t \geq v_\tau - c_\tau \\ &\implies v_t - \frac{1}{\hat{\beta}} c_t \geq v_\tau - c_\tau \\ &\iff U_t(t;\hat{\beta}) \geq U_t(\tau;\hat{\beta}). \end{aligned}$$

2. The second part is continuity:

$$U_t(t;\beta) < U_t(\tau;\beta) \iff v_t - \frac{1}{\beta} c_t < v_\tau - c_\tau \tag{2}$$

Therefore, there exists $\varepsilon > 0$ such that $\hat{\beta} < \beta + \varepsilon$ implies that for all pairs $t < \tau$ where Equation (2) holds,

$$v_t - \frac{1}{\hat{\beta}} c_t < v_\tau - c_\tau \iff U_t(t;\hat{\beta}) < U_t(\tau;\hat{\beta}).$$

□

Continuity of Behavior as Naivete Vanishes

Theorem

Fix any $\mathbf{v}$, $\mathbf{c}$, *and* β. *Let* $\mathbf{s}^s$ *denote the perception-perfect strategy for a sophisticate with* (β, δ) *preferences (we are taking* $\delta = 1$*, but the result is true in general). There exists* ε *such that for any* $\hat{\beta} \in [\beta, \beta + \varepsilon)$*:*

1. $\mathbf{s}^s$ *is also the perception-perfect strategy for sophisticated* $(\hat{\beta}, \delta)$ *preferences.*
2. *The perception-perfect strategy for* $(\beta, \hat{\beta}, \delta)$ *preferences is* $\mathbf{s} = \mathbf{s}^s$ *and the corresponding beliefs are also* $\hat{\mathbf{s}} = \mathbf{s}^s$*.*

Proof.

Part 1 follows directly from the lemma.

Part 2 follows directly from Part 1. □

Example: Welfare Loss from Naivete

A set of sufficient conditions on **c** and **v** for a $(\beta, \hat{\beta}, \delta)$ preference to procrastinate until the last period is (assume still that $\delta = 1$):

1. The individual always prefers to wait one period:

$$-c_t + \beta v_t < -\beta c_{t+1} + \beta v_{t+1} \quad \forall t.$$

2. A sophisticated $(\hat{\beta}, \delta)$ preference would do it as soon as possible:

$$-c_t + \hat{\beta} v_t \geq -\hat{\beta} c_{t+1} + \hat{\beta} v_{t+1} \quad \forall t.$$

   This implies the individual with $(\beta, \hat{\beta}, \delta)$ preferences thinks she will do it one period later if she doesn't do it now.

If $v_t = v_{t+1} = \bar{v}$, these conditions simplify to

$$\left(\frac{1}{\hat{\beta}}\right) c_t \leq c_{t+1} < \left(\frac{1}{\beta}\right) c_t.$$

If $c_t = c_{t+1} = \bar{c}$, these conditions simplify to

$$v_t - \left(\frac{1-\hat{\beta}}{\hat{\beta}}\right) \bar{c} \geq v_{t+1} > v_t - \left(\frac{1-\beta}{\beta}\right) \bar{c}.$$

Example (Procrastination for Partial Naif)

Fix any $(\beta, \hat{\beta}, \delta)$ preferences with $\beta < \hat{\beta}$. We can construct an example that shows how large the negative effects of procrastination can be, even for the slightest degree of naivete. Fix any $c_1, v_1 > 0$. Take any growth rate $\frac{1}{\hat{\beta}} < g < \frac{1}{\beta}$ and let

$$c_t = \begin{cases} g^{t-1} c_1 & \text{if } t < T' \\ \bar{X} & \text{if } t \geq T', \end{cases}$$

where $g^{T'-1} c_1 = \bar{X}$ for some large T'. The idea is that $\bar{X}$ is some large cost that the agent can ultimately be made to pay. Similarly, take any $\frac{1-\hat{\beta}}{\hat{\beta}} \bar{X} < L < \frac{1-\beta}{\beta} \bar{X}$ and let

$$v_t = \begin{cases} v_1 & \text{if } t \leq T' \\ v_1 - L(t - T') & \text{if } T' < t \leq T, \end{cases}$$

where T is the largest integer such that $v_1 - L(T - T') \geq 0$. If the individual has the option to never complete the project, then she will not complete it in this example.

How Are the Theorem and the Example Connected?

- The theorem shows that for any fixed decision problem $\mathbf{v}, \mathbf{c}$ and any β, when $\hat{\beta}$ is close enough to β, the behavior of the sophisticate and the partial naif will be the same.

- The example shows that for any given $\hat{\beta} > \beta$ (no matter how close), there exists a decision problem $\mathbf{v}, \mathbf{c}$ such that the behavior of the sophisticate and the partial naif are quite different:
  - The naif suffers a large welfare loss from procrastinating rather than completing it in the first period.
  - The sophisticate will complete the task in a period τ_s that gives
  $$v_{\tau_s} - c_{\tau_s} \geq v_1 - \frac{1}{\beta} c_1.$$
  (This condition must be satisfied for a sophisticate to delay in period 1, and is trivially satisfied if $\tau_s = 1$.)

- Combining these observations: There is "pointwise" convergence of behavior as $\hat{\beta} \to \beta$ (for each decision problem), but not "uniform" convergence (across all decision problems).

Carroll, G. D. et al. (2009): "Optimal Defaults and Active Decisions", *Quarterly Journal of Economics*, 124, 1639–1674 (page 18).

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting", *Quarterly Journal of Economics*, 112, 443–478 (page 3).

O'Donoghue, T. and M. Rabin (1999): "Doing it Now or Later", *American Economic Review*, 89, 103–124 (pages 4–6, 9, 12, 15, 27, 32).

O'Donoghue, T. and M. Rabin (2001): "Choice and Procrastination", *Quarterly Journal of Economics*, 116, 121–160 (pages 32, 36, 37).

Phelps, E. S. and R. A. Pollak (1968): "On Second-Best National Saving and Game-Equilibrium Growth", *Review of Economic Studies*, 35, 185–199 (page 3).