Doing It Now or Later

By TED O'DONOGHUE AND MATTHEW RABIN*

We examine self-control problems—modeled as time-inconsistent, present-biased preferences—in a model where a person must do an activity exactly once. We emphasize two distinctions: Do activities involve immediate costs or immediate rewards, and are people sophisticated or naive about future self-control problems? Naive people procrastinate immediate-cost activities and preproperate—do too soon—immediate-reward activities. Sophistication mitigates procrastination, but exacerbates preproperation. Moreover, with immediate costs, a small present bias can severely harm only naive people, whereas with immediate rewards it can severely harm only sophisticated people. Lessons for savings, addiction, and elsewhere are discussed. (JEL A12, B49, C70, D11, D60, D74, D91, E21)

People are impatient—they like to experience rewards soon and to delay costs until later. Economists almost always capture impatience by assuming that people discount streams of utility over time exponentially. Such preferences are *time-consistent*: A person's relative preference for well-being at an earlier date over a later date is the same no matter when she is asked.

Casual observation, introspection, and psychological research all suggest that the assumption of time consistency is importantly wrong.[1] It ignores the human tendency to grab immediate rewards and to avoid immediate costs in a way that our "long-run selves" do not appreciate. For example, when presented a choice between doing seven hours of an unpleasant activity on April 1 versus eight hours on April 15, if asked on February 1 virtually everyone would prefer the seven hours on April 1. But come April 1, given the same choice, most of us are apt to put off the work until April 15. We call such tendencies *present-biased preferences*: When considering trade-offs between two future moments, present-biased preferences give stronger relative weight to the earlier moment as it gets closer.[2]

[1] George Loewenstein (1992) reviews how the economics profession evolved from perceiving exponential discounting as a useful, ad hoc approximation of intertemporal-choice behavior, to perceiving it as a fundamental axiom of (rational) human behavior. For some recent discussions of empirical evidence of time inconsistency, see Richard H. Thaler (1991) and Thaler and Loewenstein (1992).

[2] Many researchers have studied time-inconsistent preferences. A small set of economists have over the years proposed formal, general models of time-inconsistent preferences. See, for instance, Robert H. Strotz (1956), E. S. Phelps and Robert A. Pollak (1968), Pollak (1968), Bezalel Peleg and Menahem E. Yaari (1973), Yaari (1977), and Steven M. Goldman (1979, 1980). Other researchers have posited a specific functional form, hyperbolic discounting, to account for observed tendencies for immediate gratification [see Shin-Ho Chung and Richard J. Herrnstein (1967), George Ainslie and Herrnstein (1981), Ainslie (1991, 1992), Ainslie and Nick Haslam (1992b), and Loewenstein and Drazen Prelec (1992)]. We have contrived the term "present-biased preferences" as a more descriptive term for the underlying human characteristic that hyperbolic discounting represents.

In this paper, we explore the behavioral and welfare implications of present-biased preferences in a simple model where a person must engage in an activity exactly once during some length of time. This simple model encompasses an important class of situations, and also allows us to lay bare some basic principles that might apply more generally to formal models of time-inconsistent preferences.

Our analysis emphasizes two sets of distinctions. The first distinction is whether choices involve *immediate costs*—where the costs of an action are immediate but any rewards are delayed—or *immediate rewards*—where the benefits of an action are immediate but any costs are delayed. By exploring these two different settings under the rubric of present-biased preferences, we unify the investigation of phenomena (e.g., procrastination and overeating) that have often been explored separately, but which clearly come from the same underlying propensity for immediate gratification.[3]

The second distinction is whether people are *sophisticated*, and foresee that they will have self-control problems in the future, or are *naive*, and do not foresee these self-control problems. By explicitly comparing these competing assumptions—each of which has received attention in the economics literature—we hope to delineate which predictions come from present-biased preferences per se, and which come from these assumptions about foresight.[4]

In Section I, we further motivate and formally define a simplified form of present-biased preferences [originally proposed by Phelps and Pollak (1968) and later employed by Laibson (1994)] that we study in this paper: Relative to time-consistent preferences, a person always gives extra weight to well-being now over any future moment, but weighs all future moments equally. In Section II, we set up our model of a one-time activity. We suppose that a person must engage in an activity exactly once during some length of time. Importantly, at each moment the person can choose only whether or not to do it now, and cannot choose when later she will do it. Within this scenario, we consider a general class of reward and cost schedules for completing the activity.

Section III explores the behavioral implications of present-biased preferences in our model. We present two simple results characterizing how behavior depends on whether rewards or costs are immediate, and on whether people are sophisticated or naive. The *present-bias effect* characterizes the direct implications of present-biased preferences: You *procrastinate*—wait when you should do it—if actions involve immediate costs (writing a paper), and *preproperate*—do it when you should wait—if actions involve immediate rewards (seeing a movie). Naive people are influenced solely by the present-bias effect. The *sophistication effect* characterizes the direct implications of sophistication versus naivete: A sophisticated person does the activity sooner than does a naive person with the same preferences, irrespective of whether rewards or costs are immediate. Intuitively, a sophisticated person is correctly pessimistic about her future behavior—a naive person believes she will behave herself in the future while a sophisticated person knows she may not. As a result, waiting always seems less attractive for a sophisticated person. Although the direction is the same, the sophistication effect has very different connotations for immediate costs versus immediate rewards. When costs are immediate, sophistication mitigates the tendency to procrastinate. (And in fact, the sophistication effect can outweigh the present-bias effect so that a sophisticated person may perform an onerous activity before she would if she had

---

[3] Throughout this paper, our emphasis is impulsive choice driven by a tendency to overweight rewards and costs that are in close temporal proximity. But there are clearly other aspects of impulsive choice as well: People also tend to overweight rewards and costs that are in close spatial proximity, and more generally are attentive to rewards and costs that are salient (see Loewenstein, 1996).

[4] Strotz (1956) and Pollak (1968), two of the seminal papers on time-inconsistent preferences, carefully lay out these two assumptions, but do not much consider the implications of one versus the other. More recent papers have assumed either one or the other, without attempting to justify the choice on behavioral grounds. For instance, George A. Akerlof (1991) assumes naive beliefs, while David Laibson (1994, 1995, 1997) and Carolyn Fischer (1997) assume sophisticated beliefs. Each paper states its assumption about beliefs used [and Akerlof (1991) posits that his main welfare finding depends on his assumption of naive beliefs], but conspicuously does not argue why its assumption is correct.

no self-control problem.) When rewards are immediate, on the other hand, sophistication exacerbates the tendency to preproperate.

In Section IV, we turn to the welfare results.[5] Again, the two distinctions—immediate costs vs. immediate rewards and sophistication vs. naivete—are crucial. When costs are immediate, a person is always better off with sophisticated beliefs than with naive beliefs. Naivete can lead you to repeatedly procrastinate an unpleasant activity under the incorrect belief that you will do it tomorrow, while sophistication means you know exactly how costly delay would be. In fact, even with an arbitrarily small bias for the present, for immediate costs naive people can experience severe welfare losses, while the welfare loss from a small present bias is small if you are sophisticated. When rewards are immediate, however, a person can be better off with naive beliefs. In this case, people with present-biased preferences tend to do the activity when they should wait. Naivete helps motivate you to wait because you overestimate the benefits of waiting. Sophistication makes you (properly) skeptical of future behavior, so you are more tempted to grab today's immediate reward. This can lead to "unwinding" similar to that in the finitely repeated prisoner's dilemma: In the end, you will give in to temptation and grab a reward too soon; because you realize this, near the end you will cave in a little sooner than if you thought you would resist temptation in the end; realizing this, you will cave in a little sooner, etc. As a result, for immediate rewards it is sophisticated people who can experience severe welfare losses with an arbitrarily small present bias, while the wel-

fare loss from a small present bias is small if you are naive.

Researchers looking for empirical proof of time-inconsistent preferences often explore the use of self-limiting "commitment devices" (e.g., Christmas clubs, fat farms), because such devices represent "smoking guns" that cannot be explained by any time-consistent preferences. We show in Section V that even within our simple setting, certain behaviors induced by present-biased preferences are inconsistent with any time-consistent preferences. Hence, we illustrate that smoking guns need not involve external commitment devices. Furthermore, while previous literature has focused on smoking guns for sophisticated people, we show that smoking guns exist for naive people as well.

Although many of the specific results described above are special to our one-activity model, these results illustrate some more general intuitions. To begin the process of generalizing our model, in Section VI we present an extension where, rather than being performed exactly once, the activity must be performed more than once during some length of time. In Section VII, we discuss more broadly (and less formally) what our model suggests about general implications of self-control problems, and describe how some of these implications might play out in specific economic contexts, such as saving and addiction. We then conclude with a discussion of some lessons to take away from our analysis, both for why it is important that economists start to study self-control problems, and for how we should go about doing so.

I. Present-Biased Preferences

Let u_t be a person's *instantaneous utility* in period t. A person in period t cares not only about her present instantaneous utility, but also about her future instantaneous utilities. We let $U^t(u_t, u_{t+1}, \ldots, u_T)$ represent a person's *intertemporal preferences* from the perspective of period t, where U^t is continuous and increasing in all components.[6] The standard

[6] Note that this formalization is entirely agnostic about

simple model employed by economists is exponential discounting: For all t, $U^t(u_t, u_{t+1}, \ldots, u_T) \equiv \sum_{\tau=t}^T \delta^\tau u_\tau$, where $\delta \in (0, 1]$ is a "discount factor."

Exponential discounting parsimoniously captures the fact that people are impatient. Yet exponential discounting is more than an innocuous simplification of a more general class of preferences, since it implies that preferences are *time-consistent*: A person's relative preference for well-being at an earlier date over a later date is the same no matter when she is asked. But intertemporal preferences are not time-consistent. People tend to exhibit a specific type of time-inconsistent preferences that we call *present-biased preferences*: When considering trade-offs between two future moments, present-biased preferences give stronger relative weight to the earlier moment as it gets closer.[7]

In this paper, we adopt an elegant simplification for present-biased preferences developed by Phelps and Pollak (1968), and later employed by Laibson (1994, 1995, 1997), Fischer (1997), and O'Donoghue and Rabin (1999). They capture the most basic form of present-biased preferences — a bias for the "present" over the "future" — with a simple two-parameter model that modifies exponential discounting.

Definition 1: (β, δ)-*preferences* are preferences that can be represented by:

For all t, $U^t(u_t, u_{t+1}, \ldots, u_T)$

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

where $0 < \beta, \delta \leq 1$.

In this model, δ represents long-run, time-consistent discounting. The parameter β, on the other hand, represents a "bias for the present" — how you favor now versus later. If $\beta = 1$, then (β, δ)-preferences are simply exponential discounting. But $\beta < 1$ implies present-biased preferences: The person gives more relative weight to period τ in period τ than she did in any period prior to period τ.

Researchers have converged on a simple strategy for modeling time-inconsistent preferences: The person at each point in time is modeled as a separate "agent" who is choosing her current behavior to maximize current preferences, where her future selves will control her future behavior. In such a model, we must ask what a person believes about her future selves' preferences. Strotz (1956) and Pollak (1968) carefully lay out two extreme assumptions. A person could be *sophisticated* and know exactly what her future selves' preferences will be. Or, a person could be *naive* and believe her future selves' preferences will be identical to her current self's, not realizing that as she gets closer to executing decisions her tastes will have changed. We could, of course, also imagine more intermediate assumptions. For instance, a person might be aware that her future selves will have present-biased preferences, but underestimate the degree of the present bias. Except for a brief comment in Section VII, we focus in this paper entirely on the two extreme assumptions.

Are people sophisticated or naive?[8] The use of self-commitment devices, such as alcohol

---

what factors appear as arguments in the instantaneous utility function. For instance, while it is common to assume that a person's instantaneous utility u_t depends only on her consumption bundle in period t, our formulation also allows for instantaneous utilities to depend on past consumption (as suggested by Gary S. Becker and Kevin M. Murphy, 1988; Daniel Kahneman et al., 1991).

[7] We have contrived the term "present-biased preferences" to connote that people's preferences have a bias for the "present" over the "future" (where the "present" is constantly changing). This is merely our term for an array of older models that went under different names. In fact, the (β, δ)-preferences that we will use in this paper are identical to the preferences studied by Laibson (1994), who uses the term "hyperbolic discounting," and are essentially identical to the preferences used in Akerlof (1991), although Akerlof frames his discussion very differently. For more general definitions of present-biased preferences and related elements of our model, see O'Donoghue and Rabin (1996). For an alternative formulation of the same phenomenon, see Prelec (1990), who uses the term "decreasing impatience."

[8] Most economists modeling time-inconsistent preferences assume sophistication. Indeed, sophistication implies that people have "rational expectations" about future behavior, so it is a natural assumption for econo-

clinics, Christmas clubs, or fat farms, provides evidence of sophistication.[9] Only sophisticated people would want to commit themselves to smaller choice sets: If you were naive, you would never worry that your tomorrow self might choose an option you do not like today. Despite the existence of some sophistication, however, it does appear that people underestimate the degree to which their future behavior will not match their current preferences over future behavior. For example, people may repeatedly not have the "will power" to forgo tempting foods or to quit smoking, while predicting that tomorrow they will have this will power. We think there are elements of both sophistication and naivete in the way people anticipate their own future preferences. In any event, our goal is to clarify the logic of each, and in the process we delineate which predictions come purely from present-biased preferences, and which come from the "sophistication effects" of people being aware of their own time inconsistency.

II. Doing It Once

Suppose there is an activity that a person must perform exactly once, and there are $T < \infty$ periods in which she can do it. Let $\mathbf{v} \equiv (v_1, v_2, \ldots, v_T)$ be the *reward schedule*, and let $\mathbf{c} \equiv (c_1, c_2, \ldots, c_T)$ be the *cost schedule*, where $v_t \geq 0$ and $c_t \geq 0$ for each $t \in \{1, 2, \ldots, T\}$. In each period $t \leq T - 1$, the person must choose either to do it or to wait. If she does the activity in period t, she receives reward v_t

but incurs cost c_t, and makes no further choices. If she waits, she then will face the same choice in period $t + 1$. Importantly, if the person waits she cannot commit in period t to when later she will do it. If the person waits until period T, she must do it then.

The reward schedule $\mathbf{v}$ and the cost schedule $\mathbf{c}$ represent rewards and costs as a function of when the person does the activity. However, the person does not necessarily receive the rewards and costs immediately upon completion of the activity. Indeed, we differentiate cases precisely by when rewards and costs are experienced. Some activities, such as writing a paper or mowing the lawn, are unpleasant to perform, but create future benefits. We refer to activities where the cost is incurred immediately while the reward is delayed as activities having *immediate costs*. Other activities, such as seeing a movie or taking a vacation, are pleasurable to perform, but may create future costs. We refer to activities where the reward is received immediately while the cost is delayed as activities having *immediate rewards*.[10]

We analyze these two cases using the (β, δ)-preferences outlined in Section I. For simplicity, we assume $\delta = 1$; i.e., we assume that there is no "long-term" discounting.[11] Given $\delta = 1$, without loss of generality we can interpret delayed rewards or costs as being experienced in period $T + 1$. We can then describe a person's intertemporal utility from the perspective of period t of completing the activity in period $\tau \geq t$, which we denote by $U^t(\tau)$.[12]

---

mists. Akerlof (1991) uses a variant of the naivete assumption.

[9] The very term "self-control" implies that people are aware that it may be prudent to control their future selves. For analyses of self-control in people, see Ainslie (1974, 1975, 1987, 1992), Thomas C. Schelling (1978, 1984, 1992), Thaler (1980), Thaler and Hersh M. Shefrin (1981), David C. Funder and Jack Block (1989), Stephen J. Hoch and Loewenstein (1991), Ainslie and Haslam (1992a), Jacob Glazer and Andrew Weiss (1992), Shefrin and Thaler (1992), Klaus Wertenbroch (1993), and Laibson (1994, 1995, 1997). Ainslie (1974) explores similar issues with pigeons. As many have emphasized, especially Ainslie (1992) and Bill Watterson (1993 pp. 83–88), a sort of intrapersonal "bargaining" can arise because of the basic disagreements we have with ourselves about when we should do something.

[10] We occasionally make reference to a third case where both rewards and costs are immediate. The fourth case—neither rewards nor costs are immediate—is not of interest because it is equivalent to the case of time consistency, which we study.

[11] The results are easily generalized to $\delta < 1$. Suppose the "true" reward schedule is $\boldsymbol{\pi} \equiv (\pi_1, \pi_2, \ldots, \pi_T)$, the "true" cost schedule is $\boldsymbol{\phi} \equiv (\phi_1, \phi_2, \ldots, \phi_T)$, and $\delta < 1$. If, for instance, costs are immediate and rewards are received in period $T + 1$, then if we let $v_t \equiv \delta^{T+1} \pi_t$ and $c_t \equiv \delta^t \phi_t$ for each t, doing the analysis with $\mathbf{v}$, $\mathbf{c}$, and no discounting is identical to doing the analysis with $\boldsymbol{\pi}$, $\boldsymbol{\phi}$, and δ.

[12] This formulation normalizes the instantaneous utility from not completing the activity to be zero. For instance, when costs are immediate and rewards are received in period $T + 1$, we are assuming that if the person does the

1. *Immediate Costs.*—If a person completes the activity in period τ, then her intertemporal utility in period $t \le \tau$ is

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

2. *Immediate Rewards.*—If a person completes the activity in period τ, then her intertemporal utility in period $t \le \tau$ is

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

We will focus in this environment on three types of agents. We refer to people with standard exponential, time-consistent preferences (i.e., $\beta = 1$) as *TCs*. We then focus on two types of people with present-biased preferences (i.e., $\beta < 1$), representing the two extremes discussed in Section I. We call people with sophisticated perceptions *sophisticates,* and people with naive perceptions *naifs.* Sophisticates and naifs have identical preferences (throughout we assume they have the same β), and therefore differ only in their perceptions of future preferences.

A person's behavior can be fully described by a *strategy* $s \equiv (s_1, s_2, ... , s_T)$, where $s_t \in \{Y, N\}$ specifies for period $t \in \{1, 2, ... , T\}$ whether or not to do the activity in period t given she has not yet done it. The strategy s specifies doing it in period t if $s_t = Y$, and waiting if $s_t = N$. In addition to specifying when the person will actually complete the activity, a strategy also specifies what the person "would" do in periods after she has already done it; e.g., if $s_t = Y$, we still specify $s_{t'}$ for all $t' > t$. This feature will prove useful in our analysis. Since the person must do it in period T if she has not yet done it, without loss of generality we require $s_T = Y$.

To describe behavior given our assumptions, we define a "solution concept": A *perception-perfect strategy* is a strategy that in all periods (even those after the activity is performed) a person chooses the optimal action given her current preferences and her perceptions of future behavior. Rather than give a general formal definition, we simply define a perception-perfect strategy for each of the three types of agents that we consider. Definition 2 describes a perception-perfect strategy for TCs. Reflecting the fact that TCs do not have a self-control problem, Definition 2 says that in any period, TCs will complete the activity if and only if it is the optimal period of those remaining given her current preferences.

Definition 2: A *perception-perfect strategy for TCs* is a strategy $s^{tc} \equiv (s_1^{tc}, s_2^{tc}, ... , s_T^{tc})$ that satisfies for all $t < T$ $s_t^{tc} = Y$ if and only if $U^t(t) \ge U^t(\tau)$ for all $\tau > t$.

Naifs have present-biased preferences (since $\beta < 1$), but naifs believe that they are time-consistent. As a result, the decision process for naifs is identical to that for TCs (although naifs have different preferences). Definition 3 says that in any period, naifs will complete the activity if and only if it is the optimal period of those remaining given her current preferences.

Definition 3: A *perception-perfect strategy for naifs* is a strategy $s^n \equiv (s_1^n, s_2^n, ... , s_T^n)$ that satisfies for all $t < T$ $s_t^n = Y$ if and only if $U^t(t) \ge U^t(\tau)$ for all $\tau > t$.

Although naifs and TCs have essentially the same decision process, it is important to realize that naifs have incorrect perceptions about future behavior, and therefore may plan to behave one way but in fact behave differently. With (β, δ)-preferences, these incorrect perceptions take a convenient form: At all times, naifs believe that if they wait they will behave like TCs in the future.

Sophisticates also have present-biased preferences and a self-control problem. But unlike naifs, sophisticates know they will have self-control problems in the future, and therefore correctly predict future behavior. Definition 4 says that in period t, sophisticates calculate

---

activity in period τ, the instantaneous utilities are $u_\tau = -c_\tau$, $u_{T+1} = v_\tau$, and $u_t = 0$ for all $t \notin \{\tau, T + 1\}$. This assumption is purely for convenience. In particular, for any $\bar{u}$ we would get identical results if we normalize the utility from not doing the activity to be $\bar{u}$, or if we normalize the utility from completing the activity to be $\bar{u}$.

when their future selves will complete the activity if they wait now, and then do the activity now if and only if given their current preferences doing it now is preferred to waiting for their future selves to do it.

Definition 4: A *perception-perfect strategy for sophisticates* is a strategy $\mathbf{s}^s \equiv (s_1^s, s_2^s, \dots , s_T^s)$ that satisfies for all $t < T$ $s_t^s = Y$ if and only if $U^t(t) \geq U^t(\tau')$ where $\tau' \equiv \min_{\tau > t} \{\tau | s_\tau^s = Y\}$.

Note that in Definitions 2, 3, and 4, we have assumed that people do it when indifferent, which implies that there is a unique perception-perfect strategy for each type. In addition, this assumption implies that a perception-perfect strategy must be a pure strategy. For generic values of $\mathbf{v}$, $\mathbf{c}$, and β, nobody will ever be indifferent, so these assumptions are irrelevant. In nongeneric games, more general definitions could lead to additional equilibria. For sophisticates, a perception-perfect strategy is the identical solution concept to that used by Strotz (1956), Pollak (1968), Laibson (1994, 1995, 1997), and others. For naifs, it is essentially the same solution concept as those used by Pollak (1968) and Akerlof (1991).

It will be useful in the analysis of this model to have notation for when a person will actually complete the activity (i.e., the outcome): Given the perception-perfect strategies $\mathbf{s}^{tc}$, $\mathbf{s}^s$, and $\mathbf{s}^n$, we let τ_{tc}, τ_s, and τ_n be the periods in which each of the three types of agents do the activity. That is, given $a \in \{tc, s, n\}$, $\tau_a \equiv \min_t \{t | s_t^a = Y\}$.

III. Behavior

In this section, we compare the behavior of TCs, naifs, and sophisticates who have identical long-run preferences. Comparing naifs or sophisticates to TCs reflects how people with present-biased preferences behave relative to how they would like to behave from a long-run perspective; and comparing sophisticates to naifs reflects the implications of sophistication about self-control problems.

We begin by analyzing in some detail a pair of related examples to illustrate the intuitions behind many of the results. Consider the following scenario: Suppose you usually go to the movies on Saturdays, and the schedule at the local cinema consists of a mediocre movie this week, a good movie next week, a great movie in two weeks, and (best of all) a Johnny Depp movie in three weeks. Now suppose you must complete a report for work within four weeks, and to do so you must skip the movie on one of the next four Saturdays. When do you complete the report?

The activity you must do exactly once is writing the report. The reward from doing the report is received at work in the future. We will assume the reward is independent of when you complete the report, and denote it by $\bar{v}$. The cost of doing the report on a given Saturday—not seeing the movie shown that day—is experienced immediately. Letting valuations of the mediocre, good, great, and Depp movies be 3, 5, 8, and 13, we formalize this situation in the following example, where we present both the parameters of the example and the perception-perfect strategy for each type of agent.

Example 1: 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 $\mathbf{c} = (3, 5, 8, 13)$.

$\mathbf{s}^{tc} = (Y, Y, Y, Y)$, so TCs do the report in

period $\tau_{tc} = 1$.

$\mathbf{s}^n = (N, N, N, Y)$, so naifs do the report in

period $\tau_n = 4$.

$\mathbf{s}^s = (N, Y, N, Y)$, so sophisticates do the

report in period $\tau_s = 2$.

TCs do the report on the first Saturday, skipping the mediocre movie. TCs always do the activity in the period t that maximizes $v_t - c_t$. Since Example 1 has a stationary reward schedule, TCs do the report in the period with the minimum cost.

Naifs procrastinate until the last Saturday, forcing themselves to skip the Depp movie. On the first Saturday, naifs give in to their self-control problem and see the mediocre movie because they believe they will skip the good

movie in week 2 and still be able to see the great movie and the Depp movie. The period-1 naif prefers incurring a cost of 5 next week as opposed to a cost of 3 now. However, when the second Saturday arrives, naifs again give in to their self-control problem and see the good movie, now believing they will skip the great movie in week 3 and still get to see the Depp movie. Finally, when the third Saturday arrives, naifs have self-control problems for a third time and see the great movie, forcing themselves to miss the Depp movie. This example demonstrates a typical problem for naifs when costs are immediate: They incorrectly predict that they will not procrastinate in the future, and consequently underestimate the cost of procrastinating now.

Sophisticates procrastinate one week, but they do the report on the second Saturday, skipping the good movie and enabling themselves to see the great movie and the Depp movie. The period-1 sophisticate correctly predicts that he would have self-control problems on the third Saturday and see the great movie. However, the period-1 sophisticate also correctly predicts that knowing about period-3 self-control problems will induce him to do the report on the second Saturday. Hence, the period-1 sophisticate can safely procrastinate and see the mediocre movie. Example 1 illustrates typical behavior for sophisticates when costs are immediate. Although sophisticates have a tendency to procrastinate (they do not write the report right away, which their long-run selves prefer), perfect foresight can mitigate this problem because sophisticates will do it now when they (correctly) foresee costly procrastination in the future.

Example 1 illustrates an intuition expressed by Strotz (1956) and Akerlof (1991) that sophistication is "good" because it helps overcome self-control problems. As in Akerlof's (1991) procrastination example, naifs repeatedly put off an activity because they believe they will do it tomorrow. Akerlof intuits that sophistication could overcome this problem, and Example 1 demonstrates this intuition.

However, this intuition may not hold when rewards are immediate. Consider a similar scenario: Suppose you have a coupon to see one movie over the next four Saturdays, and your allowance is such that you cannot afford to pay

for a movie. The schedule at the local cinema is the same as for the above example—a mediocre movie this week, a good movie next week, a great movie in two weeks, and (best of all) a Johnny Depp movie in three weeks. Which movie do you see?

Now, the activity you must do exactly once is going to a movie, and the reward, seeing the movie, is experienced immediately.[13] Using the same payoffs for seeing a movie as in Example 1, we have the following formalization.

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

$\mathbf{s}^{tc} = (N, N, N, Y)$, so TCs see the movie in

period $\tau_{tc} = 4$.

$\mathbf{s}^{n} = (N, N, Y, Y)$, so naifs see the movie in

period $\tau_n = 3$.

$\mathbf{s}^{n} = (Y, Y, Y, Y)$, so sophisticates see the

movie in period $\tau_s = 1$.

TCs wait and see the Depp movie since it yields the highest reward. Naifs see merely the great movie. On the first two Saturdays, naifs skip the mediocre and good movies incorrectly believing they will wait to see the Depp movie. However, on the third Saturday, they give in to self-control problems and see the great movie. For activities with immediate rewards, the self-control problem leads naifs to do the activity too soon.

Sophisticates have even worse self-control problems in this situation. They see merely the mediocre movie because of an unwinding similar to that in the finitely repeated prisoner's dilemma. The period-2 sophisticate would choose to see the good movie because he correctly predicts that he would give in to self-control problems on the third Saturday, and see merely the great movie rather than the

---

[13] That seeing a movie is a "cost" in Example 1 and a "reward" in Example 2 reflects that the rewards and costs are defined with respect to the activity being done once.

Depp movie. The period-1 sophisticate correctly predicts this reasoning and behavior by his period-2 self. Hence, the period-1 sophisticate realizes that he will see merely the good movie if he waits, so he concludes he might as well see the mediocre movie now. This example demonstrates a typical problem for sophisticates when rewards are immediate: Knowing about future self-control problems can lead you to give in to them today, because you realize you will give in to them tomorrow.[14]

We now present some propositions that characterize present-biased behavior more generally. We refer to the most basic intuition concerning how present-biased preferences affect behavior as the *present-bias effect*:[15] When costs are immediate people with present-biased preferences tend to *procrastinate*—wait when they should do it—while when rewards are immediate they tend to *preproperate*—do it when they should wait.[16] For immediate costs, they wait in periods where they should do it because they want to avoid the immediate cost. For immediate rewards, they do it in periods where they should wait because they want the immediate reward now. Proposition 1 captures that naifs are influenced solely by the present-bias effect—for immediate costs naifs always procrastinate, and for immediate rewards naifs always preproperate.[17]

PROPOSITION 1: *(1) If costs are immediate, then* $\tau_n \geq \tau_{tc}$.
(2) If rewards are immediate, then $\tau_n \leq \tau_{tc}$.

Proposition 1 is as simple as it seems: Naifs believe they will behave like TCs in the future but are more impatient now. Hence, the qualitative behavior of naifs relative to TCs intuitively and solely reflects the present-bias effect.

The behavior of sophisticates is more complicated because there is a second effect influencing their behavior. The *sophistication effect* reflects that sophisticates are fully aware of any self-control problems they might have in the future, and this awareness can influence behavior now. The sophistication effect is captured in comparisons of sophisticates to naifs. In our one-activity model, the sophistication effect is straightforward: Because sophisticates are (correctly) pessimistic that they will behave themselves in the future, they are more inclined than naifs to do it now, irrespective of whether it is costs, rewards, or both that are immediate.

PROPOSITION 2: *For all cases,* $\tau_s \leq \tau_n$.

Even though sophisticates complete the activity before naifs for both immediate costs and immediate rewards, the sophistication effect lends itself to different interpretations in these cases. For immediate costs, that sophisticates do it before naifs reflects that sophistication helps mitigate the tendency to procrastinate, as discussed in Example 1. For immediate rewards, that sophisticates do it before naifs reflects that sophistication can exacerbate the tendency to preproperate, as discussed in Example 2. These alternative interpretations will have important welfare implications, as we discuss in Section IV.

---

[14] The example also shows why sophisticates would like ways to "commit" the behavior of their future selves, as discussed by many researchers: If the period-1 sophisticate could commit himself to seeing the Depp or great movie, he would do so—even given his taste for immediate rewards. Note that with a reasonable assumption that a person does not bind himself when indifferent, the existence of commitment devices will never affect the behavior of naifs in our model, since naifs think they will always behave in the future according to their current preferences.

[15] By the present-bias effect, we mean the effect that the present bias has on the one-shot choice between doing it now versus doing it in some fixed future period. Note that for any one-shot choice, whether a person is sophisticated or naive is irrelevant.

[16] Throughout this paper, "procrastination" means that an agent chooses to wait when her long-run self (i.e., a TC) would choose to do it, and "preproperation" means that an agent chooses to do it when her long-run self would choose to wait. We derived the word "preproperate" from the Latin root "praeproperum," which means "to do before the proper time." We later found this word in a few sufficiently unabridged dictionaries, with the definition we had intended.

[17] All propositions are stated with weak inequalities; but in each case, examples exist where the inequalities are strict. All proofs are in the Appendix.

Because sophisticates are influenced by the sophistication effect in addition to the present-bias effect, the qualitative behavior of sophisticates relative to TCs is complicated. In particular, it can be that sophisticates do not even exhibit the basic present-bias intuition. Consider the following scenario: Suppose you must write a paper this weekend, on Friday night, Saturday, or Sunday. You know the paper will be better if written on either Saturday or Sunday (when you have an entire day). However, it is a mid-November weekend with plenty of sports on TV—pro basketball on Friday night, college football on Saturday, and pro football on Sunday. You prefer watching pro football to college football, and prefer college football to pro basketball. Which sports event do you miss to write the paper? We can represent this scenario with the following example, where the activity to be done once is writing the paper and the costs correspond to the attractiveness of the sports event missed.

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

TCs write the paper on Saturday because the marginal benefit of a better paper outweighs the marginal cost of giving up college football for pro basketball. Since the example involves immediate costs, the present-bias effect suggests that sophisticates should procrastinate. However, the sophistication effect leads sophisticates to write the paper on Friday night, before TCs. On Friday, sophisticates correctly predict that they will end up writing the paper on Sunday if they do not do it now. Hence, although sophisticates would prefer to write the paper on Saturday, they do it on Friday to prevent themselves from procrastinating until Sunday.

In Example 3, sophisticates behave exactly opposite from what present-biased preferences would suggest, a result we will see again in Sections VI and VII. Of course, this is not always the case. Indeed, when rewards are immediate, sophisticates always preproperate because the sophistication effect exacerbates

the self-control problem. Even so, situations like that in Example 3 are not particularly pathological, and "preemptive overcontrol" is likely to arise in real-world environments (especially when choices are discrete). We highlight this result to emphasize the importance of sophistication effects. If you assume present-biased preferences and sophistication (as economists are prone to do), you must be careful to ask whether results are driven by present-biased preferences per se, or by present-biased preferences in conjunction with sophistication effects.

IV. Welfare

Our emphasis in the previous section on qualitative behavioral comparisons among the three types of people masks what we feel may be a more important question about present-biased preferences: When does the taste for immediate gratification severely hurt a person? In this section, we examine the welfare implications of present-biased preferences with an eye towards this question. We show that even a small bias for the present can lead a person to suffer severe welfare losses, and characterize conditions when this can happen.

Welfare comparisons for people with time-inconsistent preferences are in principle problematic; the very premise of the model is that a person's preferences at different times disagree, so that a change in behavior may make some selves better off while making other selves worse off. The savings literature (e.g., Goldman, 1979, 1980; Laibson, 1994) often addresses this issue by defining a Pareto-efficiency criterion, asking when all period selves (weakly) prefer one strategy to another. If a strategy is Pareto superior to another, then it is clearly better. However, we feel this criterion is too strong: When applied to intertemporal choice, the Pareto criterion often refuses to rank two strategies even when one is much preferred by virtually all period selves, while the other is preferred by only one period self. Since present-biased preferences are often meant to capture self-control problems, where people pursue immediate gratification on a day-to-day basis, we feel the natural perspec-

tive in most situations is the "long-run perspective". (See Schelling [1984] for a thoughtful discussion of some of these issues.)[18]

To formalize the long-run perspective, we suppose there is a (fictitious) period 0 where the person has no decision to make and weights all future periods equally. We can then denote a person's *long-run utility* from doing it in period τ by $U^0(\tau) \equiv v_\tau - c_\tau$. Our welfare analysis throughout this section will involve comparisons of long-run utilities. Even so, most of our welfare comparisons can be roughly conceived of as "Pareto comparisons," and we will note Pareto-efficiency "analogues" for our two main welfare results at the end of this section.

We begin with some brief qualitative comparisons of sophisticates and naifs. The language in Section III implied that sophistication is good when costs are immediate because it mitigates the tendency to procrastinate. Indeed, it is straightforward to show that when costs are immediate, sophisticates always do at least as well as naifs [i.e., $U^0(\tau_s) \geq U^0(\tau_n)$]. Intuitively, since sophisticates never procrastinate in a period where naifs do it, the only way their utilities can differ is when sophisticates preempt costly procrastination. When sophisticates choose to preempt costly procrastination, they do so despite their exaggerated aversion to incurring immediate costs, so this decision must also be preferred by the long-run self.

When rewards are immediate, on the other hand, the discussion in Section III implied that sophistication is bad because it exacerbates the tendency to preoperate. More severe preproperation will often lead to lower long-run utility (as in Example 2), but this is not necessarily the case. In particular, if there is a future period that is very tempting (i.e., it has a large reward) but very bad from a long-run

perspective (i.e., it also has an even larger delayed cost), then more severe preproperation by sophisticates may in fact mean that sophisticates avoid this "temptation trap" while naifs do not. Hence, for immediate rewards we cannot say in general whether sophisticates or naifs are better off.

Rather than simple comparisons between sophisticates and naifs, however, our main focus for welfare analysis is the question of when a small bias for the present (i.e., β close to 1) can cause severe welfare losses. Since sophisticates, naifs, and TCs have identical long-run utility, we can measure the welfare loss from self-control problems by the deviation from TC long-run utility [i.e., $U^0(\tau_{tc}) - U^0(\tau_s)$ and $U^0(\tau_{tc}) - U^0(\tau_n)$].

We first note that if rewards and costs can be arbitrarily large, then a person with present-biased preferences can suffer arbitrarily severe welfare losses even from one-shot decisions. Suppose rewards are immediate, for instance, in which case a person with present-biased preferences is willing to grab a reward today for a delayed cost that is larger than the reward (by factor $1/\beta$). Even if β is very close to one, this decision can create an arbitrarily large welfare loss if the reward and cost are large enough.

We feel the more interesting case is when there is an upper bound on how large rewards and costs can be. In this case, the welfare loss from any individual bad decision will become very small as the self-control problem becomes small. But even if the welfare loss from any individual decision is small, severe welfare losses can still arise when self-control problems are compounded. To demonstrate this result, we suppose the upper bound on rewards and costs is $\bar{X}$. Then the welfare loss for both sophisticates and naifs cannot be larger than $2\bar{X}$.

Consider the case of immediate costs, where the self-control problem leads you to procrastinate. As in Example 1, naifs can compound self-control problems by making repeated decisions to procrastinate, each time believing they will do it next period. With each decision to procrastinate, they incur a small welfare loss, but the total welfare loss is the sum of these increments. No matter how small the individual welfare losses, naifs can suffer severe

---

[18] Indeed, Akerlof (1991) frames his discussion of procrastination in a way that emphasizes that a person's true preferences are her long-run preferences. Procrastination occurs in his model because costs incurred today are "salient"—a person experiences a cognitive illusion where costs incurred today loom larger than they are according to her true preferences.

welfare losses if they procrastinate enough times. Sophisticates, in contrast, know exactly when they will do it if they wait, so delaying from period τ_{tc} to period τ_s is a single decision to procrastinate. Hence, for sophisticates small self-control problems cannot cause severe welfare losses. The following proposition formalizes these intuitions.

PROPOSITION 3: *Suppose costs are immediate, and consider all* **v** *and* **c** *such that* $v_t \leq \bar{X}$ *and* $c_t \leq \bar{X}$ *for all t:*

(1) $\lim_{\beta \to 1}(\sup_{(\mathbf{v},\mathbf{c})}[U^0(\tau_{tc}) - U^0(\tau_s)]) = 0$, *and*

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

When rewards are immediate, however, and the self-control problem leads you to preproperate, we get the exact opposite result. For immediate rewards, naifs always believe that if they wait they will do it when TCs do it, so doing it in period τ_n as opposed to waiting until period τ_{tc} is a single decision to preproperate for naifs. Hence, for naifs small self-control problems cannot cause severe welfare losses. But sophisticates can compound self-control problems because of an unwinding: In the end, sophisticates will preproperate; because they realize this, near the end they will preproperate; realizing this they preproperate a little sooner, etc. For each step of this unwinding, the welfare loss may be small, but the total welfare loss is the sum of multiple steps. As with naifs and immediate costs, no matter how small the individual welfare losses, sophisticates can suffer severe welfare losses if the unraveling occurs over enough periods. These intuitions are formalized in Proposition 4.

PROPOSITION 4: *Suppose rewards are immediate, and consider all* **v** *and* **c** *such that* $v_t \leq \bar{X}$ *and* $c_t \leq \bar{X}$ *for all t:*

(1) $\lim_{\beta \to 1}(\sup_{(\mathbf{v},\mathbf{c})}[U^0(\tau_{tc}) - U^0(\tau_n)]) = 0$, *and*

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

As discussed at the beginning of this section, we feel that examining welfare losses in terms of long-run utility is the appropriate cri-terion to use when examining the welfare implications of present-biased preferences. Using this criterion, Propositions 3 and 4 formalize when a small bias for the present can be very costly from a long-run perspective.[19] Even so, we note that there is also a less strong formalization using Pareto comparisons: If costs are immediate, sophisticates always choose a Pareto-optimal strategy while naifs may not; and if rewards are immediate, naifs always choose a Pareto-optimal strategy while sophisticates may not.

V. Smoking Guns

Many researchers studying time-inconsistent preferences have searched for

---

[19] We feel that these limit results qualitatively capture very real differences in when moderately impatient sophisticates and naifs can suffer severe welfare losses, but there are reasons to be cautious in interpreting them too literally. For instance, since "unwinding" drives severe preproperation for sophisticates, it seems natural to ask whether a small amount of uncertainty could reverse this tendency, much as David M. Kreps et al. (1982) showed that a small amount of uncertainty can lead to extensive cooperation in the finitely repeated prisoner's dilemma. We suspect that there is something to this story, but the analogy is problematic on two fronts. First, although players may cooperate for most of a very long horizon, there is still a long duration at the end of the repeated prisoner's dilemma where players are unlikely to cooperate. Such an "endgame" could still create significant welfare losses. Second, in the Kreps et al. result a player's current behavior will signal something about her future behavior to other players. Since each "player" in our game plays only once, the comparable signal is that a person in period t infers something about the propensity of her period-$(t + 1)$ self to wait from the fact that her period-$(t - 1)$ self waited, which requires that the period-t self does not know β. While we believe that such self-inference and self-signaling go on, there are many issues to be worked out to understand the strategic logic and psychological reality of such phenomena.

A comparable worry about our extreme results for naifs is that they will eventually learn that they have a tendency to procrastinate. Again, we think there is something to this intuition, but we suspect the issue is complicated. The issue of self-inference again arises. Further, people seem to have a powerful ability not to apply general lessons they understand well to specific situations. For instance, we are all familiar with the sensation of being simultaneously aware that we tend to be overoptimistic in completing projects, but *still* being overoptimistic regarding our current project. (See Kahneman and Dan Lovallo [1993] for evidence on related issues.)

empirical proof that people have such preferences. Efforts to indirectly prove time inconsistency have focused on the use of external "commitment devices" that limit future choice sets, because the use of such devices provides smoking guns that prove time consistency wrong. In this section, we show that smoking guns exist in our simple one-activity model, where no external commitment devices are available.

There are two properties that a person with time-consistent preferences will never violate. The first is "dominance": For intertemporal choice, one strategy dominates another if it yields in every period an instantaneous utility at least as large as the instantaneous utility from the other strategy, and strictly larger for some periods. In our model, one strategy is dominated by another if and only if the first strategy implies doing it at a cost with no reward while the second strategy implies doing it for a reward with no cost.[20]

Definition 5: A person obeys *dominance* if whenever there exists some period τ with $v_\tau > 0$ and $c_\tau = 0$ the person does not do it in any period τ' with $c_{\tau'} > 0$ and $v_{\tau'} = 0$.

The second property that a person with time-consistent behavior will never violate is independence of irrelevant alternatives — eliminating an option from the choice set that is not chosen should not change the person's choice from the remaining options.

Definition 6: For any $\mathbf{v} \equiv (v_1, v_2, \ldots, v_T)$ and $\mathbf{c} \equiv (c_1, c_2, \ldots, c_T)$, define

$$\mathbf{v}^{-t} \equiv (v_1, \ldots, v_{t-1}, v_{t+1}, \ldots, v_T) \quad and$$

$$\mathbf{c}^{-t} \equiv (c_1, \ldots, c_{t-1}, c_{t+1}, \ldots, c_T).$$

A person's behavior is *independent of irrelevant alternatives* if whenever she chooses period $\tau' \neq t$ when facing $\mathbf{v}$ and $\mathbf{c}$ she also chooses τ' when facing $\mathbf{v}^{-t}$ and $\mathbf{c}^{-t}$.

A time-consistent person will never violate dominance nor independence of irrelevant alternatives. These results hold for any time-consistent preferences, including time-consistent preferences that discount differently from period to period, and even time-consistent preferences that are not additively separable. Proposition 5 establishes that these results do not hold for people with present-biased preferences.

PROPOSITION 5: *For any β and δ such that $0 < \delta \leq 1$ and $0 < \beta < 1$, and for both sophistication and naivete:*
(1) There exists $(\mathbf{v}, \mathbf{c})$ and assumptions about immediacy such that a person with (β, δ)-preferences will violate dominance, and
(2) There exists $(\mathbf{v}, \mathbf{c})$ and assumptions about immediacy such that a person with (β, δ)-preferences will violate independence of irrelevant alternatives.

To give some intuition for these results, we describe examples where each type violates dominance. The intuition for why each type violates independence of irrelevant alternatives is related. Sophisticates violate dominance when they choose a dominated early time to do an activity because they (correctly) worry that their future selves will not choose the dominating later time. For example, suppose rewards are immediate, $T = 3$, $\mathbf{v} = (0, 5, 1)$ and $\mathbf{c} = (1, 8, 0)$. Doing it in period 1 is clearly dominated by doing it in period 3. Even so, a sophisticate with $\beta = \frac{1}{2}$ will complete the activity in period 1. She does so not because it is her most preferred period, but rather to avoid doing it in period 2. In period 1, the person prefers period 3 to period 1. Unfortunately, the period-2 self gets to choose between periods 2 and 3, and she will choose period 2.

Naifs can violate dominance because of incorrect perceptions about future behavior. For example, suppose costs are immediate, $T = 3$, $\mathbf{v} = (1, 8, 0)$ and $\mathbf{c} = (0, 5, 1)$. Doing it in

---

[20] E.g., consider a three-period example where $\mathbf{v} = (1, x, 0)$ and $\mathbf{c} = (0, y, 1)$. Then if costs are immediate, doing it in period 1 yields the stream of instantaneous utilities $(0, 0, 0, 1)$ while doing it in period 3 yields the stream of instantaneous utilities $(-1, 0, 0, 0)$. Clearly the former dominates the latter.

period 3 is dominated by doing it in period 1, and yet a naif with $\beta = \frac{1}{2}$ will choose period 3. Even though in period 1 she prefers period 1 to period 3, she waits in period 1 incorrectly believing she will do it in period 2. Unfortunately, in period 2 she prefers waiting until period 3.[21]

Proposition 5 has important implications for the literature on smoking guns. First, Proposition 5 implies that smoking guns need not involve the use of external commitment devices. Even simple behaviors can sometimes represent smoking guns. Furthermore, the literature on external commitment devices provides smoking guns for sophisticates but not for naifs, since naifs would not pay to limit future choice sets. Proposition 5 implies that smoking guns exist for naifs as well. Finally, the intuitions above (and in the proof) suggest ways to design experiments attempting to find smoking guns, as well as the types of real world situations without external commitment devices where smoking guns might be found.

VI. Multi-Tasking

We now begin to explore how our results might carry over to more general settings. Consider a simple extension of our model where the activity must be performed more than once. The basic structure of the model is exactly as in Section II, but now the person must do the activity exactly $M \geq 1$ times, and she can do it at most once in any given period. We let $\tau^i(M)$ denote the period in which a person completes the activity for the ith time, and define $\Theta(M) \equiv \{\tau^1(M), \tau^2(M), \dots, \tau^M(M)\}$. For each period τ in which the person does it, she receives reward v_τ and incurs cost c_τ, and these can be experienced immediately or with some delay. Using the interpretations of immediate costs and immediate rewards from Section II, preferences take the following form.

1. *Immediate Costs.* —Given $\Theta(M)$, the set of periods in which she does it, a person's intertemporal utility in period t is given by equation (1) below.

2. *Immediate Rewards.* —Given $\Theta(M)$, the set of periods in which she does it, a person's intertemporal utility in period t is given by equation (2) below.

Given these preferences, we can define perception-perfect strategies analogously to Definitions 2, 3, and 4. We omit the formal definitions here. Let $\Theta_a(M) \equiv \{\tau_a^1(M), \dots, \tau_a^M(M)\}$ be the set of periods that an agent of type $a \in \{tc, s, n\}$ completes the activity according to her perception-perfect strategy. We begin by showing that the behavior of TCs and naifs in the multiactivity model is "normal" and intuitive.

$$
(1) \qquad U^t(\Theta(M)) \equiv \begin{cases} -(1-\beta)c_t + \beta\left(\displaystyle\sum_{\tau \in \Theta(M)} v_\tau - \sum_{\tau \in \Theta(M)} c_\tau\right) & \text{if } t \in \Theta(M) \\[2em] \beta\left(\displaystyle\sum_{\tau \in \Theta(M)} v_\tau - \sum_{\tau \in \Theta(M)} c_\tau\right) & \text{if } t \notin \Theta(M). \end{cases}
$$

$$
(2) \qquad U^t(\Theta(M)) \equiv \begin{cases} (1-\beta)v_t + \beta\left(\displaystyle\sum_{\tau \in \Theta(M)} v_\tau - \sum_{\tau \in \Theta(M)} c_\tau\right) & \text{if } t \in \Theta(M) \\[2em] \beta\left(\displaystyle\sum_{\tau \in \Theta(M)} v_\tau - \sum_{\tau \in \Theta(M)} c_\tau\right) & \text{if } t \notin \Theta(M). \end{cases}
$$

---

[21] The proof of Proposition 5 essentially involves generalizing these examples for all values of β and δ.

PROPOSITION 6: *(1) For all cases and for any* **v** *and* **c**, *for each* $M \in \{1, 2, ..., T - 1\}$: $\Theta_{tc}(M) \subset \Theta_{tc}(M + 1)$ *and* $\Theta_n(M) \subset \Theta_n(M + 1)$; *and*

(2) If costs are immediate, then for all $i \in \{1, 2, ..., M\}$, $\tau_n^i(M) \geq \tau_{tc}^i(M)$, *and if rewards are immediate, then for all* $i \in \{1, 2, ..., M\}$, $\tau_n^i(M) \leq \tau_{tc}^i(M)$.

Part 1 of Proposition 6 addresses how behavior depends on M: If TCs or naifs must do the activity an extra time, they do it in all periods they used to do it, and some additional period. If in any period they have k activities remaining, both TCs and naifs do it now if and only if the current period is one of the k best remaining periods given their current preferences. Having more activities remaining, therefore, makes it more likely that they perform an activity now. Part 2 of Proposition 6 states that the qualitative behavior of naifs relative to TCs in the multiactivity model is exactly analogous to that in the one-activity model. If costs are immediate, naifs procrastinate: They are always behind TCs in terms of activities completed so far. If rewards are immediate, naifs preproperate: They are always ahead of TCs in terms of activities completed so far. Hence, the present-bias effect extends directly to the multiactivity setting; and again naifs exhibit the pure effects of present-biased preferences.

While the behavior of naifs in the multiactivity model is a straightforward and intuitive analogue of their behavior in the one-activity model, the effects of sophistication are significantly complicated. Consider the following example.

Example 4: Suppose rewards are immediate, $T = 3$, and $\beta = \frac{1}{2}$ for naifs and sophisticates. Let **v** $= (6, 11, 21)$ and **c** $= (0, 0, 0)$.

If $M = 1$, then $\tau_s = 1$, $\tau_n = 2$, and $\tau_{tc} = 3$.

If $M = 2$, then $\Theta_s(2) = \{2, 3\}$, $\Theta_n(2) = \{1, 2\}$, and $\Theta_{tc}(2) = \{2, 3\}$.

There are a couple of aspects of Example 4 worth emphasizing. First, changing M dramatically changes the behavior of sophisticates: While sophisticates always preproperate when there is one activity, they do not preproperate here with two activities. Hence, the an-

alogue to Part 1 of Proposition 6 does not hold for sophisticates. Sophisticates are always looking for ways to influence their future behavior, and for $M > 1$ waiting can be a sort of "commitment device" to influence future behavior. If there is only one activity, there is no way to commit future selves not to preproperate. In Example 4, when $M = 1$ the period-1 sophisticate does the activity because he (correctly) predicts that he will just do it in period 2 if he waits. If there is a second activity, however, a commitment device becomes available: Waiting now prevents you from doing the activity for the second time tomorrow; you can only do it for the first time tomorrow. Thus, forgoing the reward today makes you delay until period 3. When $M = 2$, the period-1 sophisticate knows he will do the second activity in period 2 if he does the first now, but he can force himself to do it in periods 2 and 3 if he waits now.

Example 4 also illustrates that the simple comparison of Proposition 2—that for $M = 1$ sophisticates always do it before naifs—does not extend to the multiactivity case. In Example 4 with $M = 2$, sophisticates do it after naifs. The intuition behind Proposition 2 was that sophisticates are correctly pessimistic about their utility from completing the activity in the future, and are therefore less willing to wait than naifs. But for $M > 1$ the relevant question is how pessimism affects the marginal utility of delaying one activity. As a result, there is no general result for the implications of sophistication versus naivete. Example 4 shows for immediate rewards that sophistication can sometimes mitigate rather than exacerbate preproperation. Likewise, for immediate costs one can also find cases where sophistication exacerbates procrastination (and where sophisticates are worse off than naifs). These examples illustrate that, in general environments, identifying when sophistication mitigates self-control problems and when it exacerbates them is more complicated than in the one-activity model. It is still true that sophisticates are more pessimistic than naifs about future behavior. But in more general environments, comparisons of sophisticates to naifs depends on whether pessimism increases or decreases the marginal cost of current indulgence. As we discuss in Section

VII, in many contexts there are identifiable patterns as to how pessimism will affect incentives to behave oneself—but these patterns will not always correspond to the simple case of Proposition 2.

We conclude this section by returning to a point made in Section III–that sophistication can lead a person to behave in ways that are seemingly contrary to having present-biased preferences. In Section III, we showed that sophisticates may do it before TCs even though costs are immediate. In the following example, sophisticates do things after TCs even though rewards are immediate.

Example 5: Suppose rewards are immediate, and $\beta = \frac{1}{2}$ for naifs and sophisticates. Let $\mathbf{v} = (12, 6, 11, 21)$ and $\mathbf{c} = (0, 0, 0, 0)$.

If $M = 2$, then $\Theta_{tc}(2) = \{1, 4\}$, $\Theta_n(2) = \{1, 3\}$, and $\Theta_s(2) = \{3, 4\}$.

In Example 5, the situation beginning in period 2 is identical to Example 4, and the intuition for why sophisticates do it later than TCs is related to the intuition of Example 4. The period-1 sophisticate knows that if he has one activity left in period 2, he will do it in period 2, while if he has two activities left in period 2, he will wait until periods 3 and 4. Hence, even though the period-1 sophisticate's most preferred periods for doing it are periods 1 and 4, he realizes he will not do it in period 4 if he does it in period 1. The choice for the period-1 sophisticate is between doing it in periods 1 and 2 versus doing it in periods 3 and 4. Of course situations like Example 5 are somewhat special; but we do not feel they are so pathological that they will never occur in real-world environments (particularly for discrete choices).

VII. Discussion and Conclusion

Many economic applications where present-biased preferences are clearly important cannot readily be put into the framework of this paper. Nonetheless, we feel our analysis provides some insight into such realms. In this section, we discuss some general lessons to take away from our analysis, and illustrate how these general lessons might play out in particular economic applications, such as savings and addiction.[22]

In our model, the behavior of naifs intuitively and directly reflects their bias for the present. We suspect this simplicity in predicting the effects of naive self-control problems will hold in a broad array of economic models. Since consuming now yields immediate payoffs whereas the increased future payoffs that saving allows is delayed, naifs will undersave in essentially any savings model; and since addictive activities involve yielding to some immediate desire today that has future costs, naifs will overindulge in essentially any addiction model.

In contrast to naifs, sophisticates in our model can behave in ways that seemingly contradict having present-biased preferences. We saw in Section III that sophisticates may complete an unpleasant task before they would if they had no self-control problem, and in Section VI that they may consume tempting goods later than they would if they had no self-control problem. We suspect this complexity in predicting the effects of sophisticated self-control problems will also hold more generally. Sophistication effects that operate in addition to, and often in contradiction to, the present-bias effect can be quite significant. In the realm of saving, sophisticates can have a negative marginal propensity to consume over some ranges of income; and sophisticates can sometimes save more than TCs (i.e., they can behave exactly opposite from what a present bias would suggest).[23] In the realm of

---

[22] There has been a lot of previous research on time inconsistency in savings models; see, for instance, Strotz (1956), Phelps and Pollak (1968), Pollak (1968), Thaler and Shefrin (1981), Shefrin and Thaler (1988, 1992), Laibson (1994, 1995, 1997), and Thaler (1994). Recently, economists have proposed models of "rational addiction" (Becker and Murphy, 1988; Becker et al., 1991, 1994). These models insightfully formalize the essence of (bad) addictive goods: Consuming more of the good today decreases overall utility but increases marginal utility for consumption of the same good tomorrow. However, these models a priori rule out the time-inconsistency and self-control issues modeled in this paper, and which many observers consider important in addiction.

[23] For simple examples of such behaviors, consider the following savings interpretation of a multiactivity model with $\mathbf{c} = (0, 0, \ldots, 0)$: People have time-variant

addiction, when it is optimal to consume an addictive product in moderation, sophisticates may not consume at all as a means of self-control—they know they will lose control if they try to consume in moderation. It is even possible to construct models where addictive goods are Giffen goods for sophisticates—non-addicts may buy more of a good in response to a permanent price increase, because high prices act as a sort of commitment device not to become addicted in the future.

People clearly have some degree of sophistication, and many sophistication effects—particularly attempts at self-control—seem very real. Other examples of sophistication effects seem perverse, however, and the corresponding behavior is likely to be somewhat rare. Hence, economists should be cautious when exploring present-biased preferences solely with the assumption of sophistication (which economists are prone to do since sophistication is closer to the standard economic assumptions). Because our analysis shows that sophistication effects can have large behavioral implications, and since people are clearly not completely sophisticated, researchers should be careful to clarify which results are driven by present-biased preferences per se, and which results arise from present-biased preferences in conjunction with sophistication effects.

We suspect one reason economists are so prone to assume sophistication in their models is the rule of thumb that less extreme depar-

tures from classical economic assumptions will lead to less extreme departures from classical predictions; hence, it is presumed that whatever novel predictions arise assuming sophistication will hold a fortiori assuming naivete. This rule of thumb does not apply here, of course, because many commitment strategies and other behaviors arise only because of sophistication. Moreover, our analysis also shows that even when sophistication does not affect the qualitative predictions, it does not always yield "milder" departures from conventional predictions: In many situations, being aware of self-control problems can exacerbate self-control problems.[24]

Indeed, another major theme of our analysis is to characterize the types of situations where sophistication mitigates versus exacerbates self-control problems. Extrapolating from our results, sophistication helps you when knowing about future misbehavior increases your perceived cost of current misbehavior, thereby encouraging you to behave yourself now. Sophistication hurts you when knowing about future misbehavior decreases the perceived cost of current misbehavior. In our one-activity model, this manifests itself in a simple fashion: When costs are immediate, you tend to procrastinate; if you are aware you will procrastinate in the future, that makes you perceive it as more costly to procrastinate now. Hence, sophistication helps when costs are immediate. When rewards are immediate, you tend to preproperate; if you are aware you will preproperate in the future, that makes you perceive

instantaneous utility functions, where in any period t the marginal utility of consuming the first dollar is v_t, and the marginal utility for any consumption beyond the first dollar is negligible. Then given wealth $M \in \{ \$1, \$2, ... , \$T \}$, you must decide in which periods to consume. With this savings interpretation, sophisticates have a negative marginal propensity to consume in Example 4: With wealth $1, sophisticates consume $1 in period 1, while with wealth $2, sophisticates consume $0 in period 1. And sophisticates save more than TCs in Example 5: With wealth $2, TCs consume $1 in year 1 and save $1 (which is consumed in year 4), while sophisticates consume $0 in year 1 and save $2 (which is consumed in years 3 and 4). Although Examples 4 and 5 use rather special utility functions, it is relatively straightforward to find similar examples where utility functions are concave, increasing, and differentiable. We suspect, but have not proven, that sophisticates will never save more than TCs if utility functions are constant over time.

[24] We have seen little discussion in the literature of how sophistication might affect the implications of self-control problems. Strotz (1956) and Akerlof (1991) discuss how sophistication might help improve behavior. We suspect their discussion reflects the prevalent intuition that sophistication can only help, and in fact have found no explicit discussion anywhere of how awareness of self-control problems might hurt. That sophistication can hurt you is, however, implicit in Pollak (1968). In the process of demonstrating a mathematical result, Pollak shows that sophisticates and naifs behave the same for logarithmic utility. From this, it is straightforward to show that for utility functions more concave than the log utility function, sophisticates save more than naifs (i.e., sophistication mitigates self-control problems), whereas for less concave utility functions, sophisticates save less than naifs (i.e., sophistication exacerbates self-control problems).

it as less costly to preproperate now. Hence, sophistication hurts when rewards are immediate.

In richer economic environments, whether sophistication helps or hurts will be more complicated. Nonetheless, our analysis suggests some simple conjectures. Consider, for example, the realm of addiction. Our analysis suggests sophistication might help when one wants to quit an addiction. A naive person may repeatedly delay quitting smoking believing he will quit tomorrow; and Proposition 4 suggests that this problem could lead to significant welfare losses. Sophistication should prevent this problem. In contrast, sophistication may hurt when a person is sure she will eventually get addicted, because this might lead to an unwinding logic along the lines of our Example 2, by which she decides that since she will eventually succumb to temptation she might as well get addicted now.[25]

We conclude by reviewing two motivations for incorporating present-biased preferences into economic analysis. First, present-biased preferences may be useful in predicting behavior. There seem to be numerous applications where present-biased preferences can explain a prevalent behavior in a simple and plausible way, whereas post hoc and contrived explanations are required if one insists on interpreting phenomena through the prism of time-consistent preferences. For instance, Fischer (1997) observes that episodes of procrastination might be consistent with time consistency—but only if one assumes an absurd discount factor or implausibly low costs of delay. In contrast, present-biased preferences can explain the same episode of procrastination with a reasonable discount factor and a small bias for the present.[26]

But in many situations, present-biased preferences and time-consistent preferences both provide perfectly plausible explanations for behavior. Even so, a second motivation for incorporating present-biased preferences into economic analysis is that these two explanations can have vastly different welfare implications. For example, suppose a person becomes fat from eating large quantities of potato chips. She may do so because of a harmful self-control problem, or merely because the pleasure from eating potato chips outweighs the costs of being fat. Both hypotheses are reasonable explanations for the observed behavior; however, the two hypotheses have very different normative implications. The former says people buy too many potato chips at the prevailing price; the latter says they buy the right amount. Because welfare analyses are often the main contribution economists can make, distinguishing between these two hypotheses is crucial. To further emphasize this point, consider the more policy-relevant example of an economic analysis of cigarette taxation that a priori assumes away self-control problems. This analysis may (or may not) yield a very accurate prediction of how cigarette taxes will affect consumption. But by ignoring self-control and related problems, it is likely to be either useless or very misleading as a guide to optimal cigarette-tax policy.

There are clearly many reasons to be cautious about welfare analyses that abandon rational-choice assumptions, and research ought to employ the most sophisticated methods available to carefully discern whether behaviors truly reflect harmful self-control problems. But the existence of present-biased preferences is overwhelmingly supported by psychological evidence, and strongly accords to common sense and conventional wisdom. And recall that our analysis in Section IV suggests that even relatively mild self-control problems can lead to significant welfare losses. Hence, even if the psychological evidence, common sense, and conventional wisdom are just a little right, and economists' habitual assumption of time consistency is just a little wrong, welfare economics ought be attentive to the role of self-control problems.

---

[25] We believe it is likely that in most contexts—including addiction—sophistication will mitigate self-control problems rather than exacerbate them; but our analysis makes clear that there is no general principle guaranteeing this.

[26] O'Donoghue and Rabin (1999) show, in turn, that efforts to combat procrastination arising from present-biased preferences may help explain why incentive schemes involve deadlines that punish delays in completing a task much more harshly after some date than before that date—even when the true costs of delay are stationary. (Of course, it is likely there are plausible "time-consistent" explanations for the use of deadlines as well.)

By analyzing the implications of present-biased preferences in a simple model, and positing some general lessons that will likely carry over to other contexts, we hope that our paper will add to other research in developing a tractable means for economists to investigate both the behavioral and welfare implications of present-biased preferences.

APPENDIX

PROOF OF PROPOSITION 1:

(1) We show that when costs are immediate, for any period if naifs do it then TCs do it. Consider period t, and let $t' \equiv \max_{\tau > t}(v_\tau - c_\tau)$. Naifs do it in period t only if $\beta v_t - c_t \geq \beta(v_{t'} - c_{t'})$, or $v_t - (1/\beta)c_t \geq v_{t'} - c_{t'}$; TCs do it in period t if $v_t - c_t \geq v_{t'} - c_{t'}$; and $v_t - c_t \geq v_t - (1/\beta)c_t$ for any $\beta \leq 1$. The result follows.

(2) We show that when rewards are immediate, for any period if TCs do it then naifs do it. Consider period t, and let $t' \equiv \max_{\tau > t}(v_\tau - c_\tau)$. TCs do it in period t only if $v_t - c_t \geq v_{t'} - c_{t'}$; naifs do it in period t if $v_t - \beta c_t \geq \beta(v_{t'} - c_{t'})$, or $(1/\beta)v_t - c_t \geq v_{t'} - c_{t'}$; and $(1/\beta)v_t - c_t \geq v_t - c_t$ for any $\beta \leq 1$. The result follows.

PROOF OF PROPOSITION 2:

We show that for any period, if naifs do it then sophisticates do it. Recall naifs and sophisticates have identical preferences. The result follows directly because naifs do it in period t only if $U^t(t) \geq U^t(\tau)$ for all $\tau > t$, while sophisticates do it in period t if $U^t(t) \geq U^t(\tau')$ for $\tau' \equiv \min_{\tau > t}\{\tau | s_\tau^s = Y\}$.

PROOF OF PROPOSITION 3:

(1) We first argue that when costs are immediate, for any $t < t'$ such that $s_t^s = s_{t'}^s = Y$, $U^0(t) \geq U^0(t')$. This follows because for any t and $t' \equiv \min_{\tau > t}\{\tau | s_\tau^s = Y\}$, $s_t^s = Y$ only if $\beta v_t - c_t \geq \beta(v_{t'} - c_{t'})$, which implies $v_t - c_t \geq v_{t'} - c_{t'}$.

Now let $\bar{\tau} \equiv \min_{\tau > \tau_{tc}}\{\tau | s_\tau^s = Y\}$, so $\bar{\tau}$ is when sophisticates would do it if they waited in all $t \leq \tau_{tc}$. If $U^0(\tau_s) < U^0(\tau_{tc})$ then $s_{\tau_{tc}}^s = N$, so either $\tau_s = \bar{\tau}$ or $\tau_s < \tau_{tc}$. But using the

result above, in either case $U^0(\tau_s) \geq U^0(\bar{\tau})$, which implies $U^0(\tau_{tc}) - U^0(\tau_s) \leq U^0(\tau_{tc}) - U^0(\bar{\tau})$. Given the definition of $\bar{\tau}$, $s_{\bar{\tau}}^s = N$ only if $\beta v_{\tau_{tc}} - c_{\tau_{tc}} < \beta U^0(\bar{\tau})$ or $-((1 - \beta)/\beta)c_{\tau_{tc}} + U^0(\tau_{tc}) < U^0(\bar{\tau})$. Given the upper bound on costs $\bar{X}$, we must have $U^0(\tau_{tc}) - U^0(\tau_s) < ((1 - \beta)/\beta)\bar{X}$. It is straightforward to show we can get arbitrarily close to this bound, so $\sup_{(\mathbf{v},\mathbf{c})}[U^0(\tau_{tc}) - U^0(\tau_s)] = ((1 - \beta)/\beta)\bar{X}$. Hence, $\lim_{\beta \to 1}(\sup_{(\mathbf{v},\mathbf{c})}[U^0(\tau_{tc}) - U^0(\tau_s)]) = 0$.

(2) Fix $\beta < 1$. We will show that for any $\varepsilon \in (0, \bar{X})$ there exist reward/cost schedule combinations such that $U^0(\tau_{tc}) - U^0(\tau_n) = 2\bar{X} - \varepsilon$, from which the result follows. Choose $\gamma > 0$ such that $\beta + \gamma < 1$. Let i be the integer satisfying $(\varepsilon)/(\beta + \gamma)^i < \bar{X} \leq (\varepsilon)/(\beta + \gamma)^{i+1}$, and let j be the integer satisfying $\bar{X} - j((1 - \beta)/(\beta + \gamma))\bar{X} > 0 \geq \bar{X} - (j + 1)((1 - \beta)/(\beta + \gamma))\bar{X}$. Consider the following reward and cost schedules where $T = i + j + 3$ is finite:

$$\mathbf{v} = (\bar{X}, \bar{X}, \ldots, \bar{X}, \bar{X} - ((1 - \beta)/(\beta$$
$$+ \gamma))\bar{X}, \bar{X} - 2((1 - \beta)/(\beta + \gamma))\bar{X},$$
$$\ldots, \bar{X} - j((1 - \beta)/(\beta + \gamma))\bar{X}, 0)$$

$$\mathbf{c} = (\varepsilon, \varepsilon/(\beta + \gamma), \varepsilon/(\beta + \gamma)^2, \ldots,$$
$$\varepsilon/(\beta + \gamma)^i, \bar{X}, \bar{X}, \ldots, \bar{X}).$$

Under $\mathbf{v}$ and $\mathbf{c}$, $\tau_{tc} = 1$ so $U^0(\tau_{tc}) = \bar{X} - \varepsilon$, and $\tau_n = T$ so $U^0(\tau_n) = -\bar{c}$. Hence, we have $U^0(\tau_{tc}) - U^0(\tau_n) = 2\bar{X} - \varepsilon$.

PROOF OF PROPOSITION 4:

(1) When rewards are immediate, by Proposition 1 $\tau_n \leq \tau_{tc}$. For any $t < \tau_{tc}$, naifs believe they will do it in period τ_{tc} if they wait. Hence, $v_{\tau_n} - \beta c_{\tau_n} \geq \beta U^0(\tau_{tc})$, which we can rewrite as $((1 - \beta)/\beta)v_{\tau_n} + U^0(\tau_n) \geq U^0(\tau_{tc})$. Given the upper bound on rewards $\bar{X}$, we have $U^0(\tau_{tc}) - U^0(\tau_n) \leq ((1 - \beta)/\beta)\bar{X}$. Since the bound is easily achieved, $\sup_{(\mathbf{v},\mathbf{c})}[U^0(\tau_{tc}) - U^0(\tau_n)] = ((1 - \beta)/\beta)\bar{X}$, and $\lim_{\beta \to 1}(\sup_{(\mathbf{v},\mathbf{c})}[U^0(\tau_{tc}) - U^0(\tau_n)]) = 0$.

(2) Fix $\beta < 1$. We will show that for any $\varepsilon \in (0, \bar{X})$ there exist reward/cost schedule combinations such that $U^0(\tau_{tc}) - U^0(\tau_s) = 2\bar{X} - \varepsilon$, from which the result follows. Let i

be the integer satisfying $(\varepsilon)/(\beta^i) < \bar{X} \le (\varepsilon)/(\beta^{i+1})$, and let j be the integer satisfying $\bar{X} - j((1 - \beta)/\beta)\bar{X} > 0 \ge \bar{X} - (j + 1)((1 - \beta)/\beta)\bar{X}$. Consider the following reward and cost schedules where $T = i + j + 3$ is finite:

$$\mathbf{v} = (\varepsilon, \varepsilon/(\beta), \varepsilon/(\beta^2), \ldots, \varepsilon/(\beta^i), \bar{X},$$

$$\bar{X}, \ldots, \bar{X})$$

$$\mathbf{c} = (\bar{X}, \bar{X}, \ldots, \bar{X}, \bar{X} - ((1 - \beta)/\beta)\bar{X},$$

$$\bar{X} - 2((1 - \beta)/\beta)\bar{X}, \ldots,$$

$$\bar{X} - j((1 - \beta)/\beta)\bar{X}, 0).$$

Under $\mathbf{v}$ and $\mathbf{c}$, $\tau_{tc} = T$ so $U^0(\tau_{tc}) = \bar{X}$, and $\tau_s = 1$ so $U^0(\tau_s) = \varepsilon - \bar{X}$. Hence, we have $U^0(\tau_{tc}) - U^0(\tau_s) = 2\bar{X} - \varepsilon$.

PROOF OF PROPOSITION 5:

We prove each part by constructing examples.

(1) Suppose rewards are immediate, $T = 3$, $\mathbf{v} = (0, x, 1)$ and $\mathbf{c} = (1, y, 0)$. Sophisticates choose dominated strategy (Y, Y, Y) if $(x) - \beta\delta^2(y) \ge \beta\delta(1) - \beta\delta^2(0)$ and $0 - \beta\delta^3(1) \ge \beta\delta(x) - \beta\delta^3(y)$. We can rewrite these conditions as $\delta^2 y - \delta^2 \ge x \ge \beta\delta + \beta\delta^2 y$. If $y > (\beta + \delta)/(\delta(1 - \beta))$ then $\delta^2 y - \delta^2 > \beta\delta + \beta\delta^2 y$. Hence, for any β and δ there exists $y > (\beta + \delta)/(\delta(1 - \beta))$ and $x \in (\beta\delta + \beta\delta^2 y, \delta^2 y - \delta^2)$, in which case $\mathbf{s}^s = (Y, Y, Y)$.

Suppose costs are immediate, $T = 3$, $\mathbf{v} = (1, x', 0)$ and $\mathbf{c} = (0, y', 1)$. Naifs choose dominated strategy (N, N, Y) if $\beta\delta^3(1) - (0) < \beta\delta^3(x') - \beta\delta(y')$ and $\beta\delta^2(x') - (y') < \beta\delta^2(0) - \beta\delta(1)$. We can rewrite these conditions as $\delta^2 x' - \delta^2 > y' > \beta\delta + \beta\delta^2 x'$. If $x' > (\beta + \delta)/(\delta(1 - \beta))$ then $\delta^2 x' - \delta^2 > \beta\delta + \beta\delta^2 x'$. Hence, for any β and δ there exists $x' > (\beta + \delta)/(\delta(1 - \beta))$ and $y' \in (\beta\delta + \beta\delta^2 x', \delta^2 x' - \delta^2)$, in which case $\mathbf{s}^n = (N, N, Y)$.

(2) For any β and δ, choose $\phi \in (\sqrt{\beta}, 1)$, let $\mathbf{v} = (0, 0, 0)$ and $\mathbf{c} = (1, \phi/(\beta\delta), \phi^2/(\beta^2\delta^2))$, and suppose costs are immediate. Then sophisticates choose $\tau_s = 1$ when facing $\mathbf{v}$ and $\mathbf{c}$, but $\tau_s = 2$ when facing $\mathbf{v}^{-T}$ and $\mathbf{c}^{-T}$, and this violates independence of irrelevant alternatives.

For any β and δ, choose $\phi \in (\sqrt{\beta}, 1)$, let $\mathbf{v} = (1, \phi/(\beta\delta), \phi^2/(\beta^2\delta^2))$ and $\mathbf{c} = (0, 0, 0)$, and suppose rewards are immediate. Then naifs choose $\tau_n = 2$ when facing $\mathbf{v}$ and $\mathbf{c}$, but $\tau_n = 1$ when facing $\mathbf{v}^{-T}$ and $\mathbf{c}^{-T}$, and this violates independence of irrelevant alternatives.

PROOF OF PROPOSITION 6:

(1) For both TCs and naifs, if they have k activities remaining in period t, then they do it in period t if and only if period t is one of the k best remaining periods given period-t preferences. Hence, for any $k' > k$, if TCs or naifs do it in period t with k activities remaining, then they do it in period t with k' activities remaining. Given this, the result is straightforward.

(2) We first show that for any t and k, when TCs and naifs each have k activities remaining in period t, then (i) for immediate costs if naifs do it in period t then TCs do it in period t; and (ii) for salient rewards if TCs do it in period t then naifs do it in period t. Let t' be such that $v_{t'} - c_{t'}$ is the kth best $v_\tau - c_\tau$ for $\tau \in \{t + 1, t + 2, \ldots, T\}$. (i) follows because for immediate costs, naifs do it in period t only if $\beta v_t - c_t \ge \beta(v_{t'} - c_{t'})$, or $v_t - (1/\beta)c_t \ge v_{t'} - c_{t'}$; TCs do it in period t if $v_t - c_t \ge v_{t'} - c_{t'}$; and $v_t - c_t \ge v_t - (1/\beta)c_t$ for any $\beta \le 1$. (ii) follows because for immediate rewards, TCs do it in period t only if $v_t - c_t \ge v_{t'} - c_{t'}$; naifs do it in period t if $v_t - \beta c_t \ge \beta(v_{t'} - c_{t'})$, or $(1/\beta)v_t - c_t \ge v_{t'} - c_{t'}$; and $(1/\beta)v_t - c_t \ge v_t - c_t$ for any $\beta \le 1$. The result then follows because (i) implies that for immediate costs naifs can never get ahead of TCs, and (ii) implies that for immediate rewards TCs can never get ahead of naifs.

REFERENCES

Ainslie, George. "Impulse Control in Pigeons." *Journal of the Experimental Analysis of Behavior*, May 1974, *21*(3), pp. 485–89.

_____. "Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control." *Psychological Bulletin*, July 1975, *82*(4), pp. 463–96.

_____. "Self-Reported Tactics of Impulse Control." *International Journal of the Addictions*, February 1987, *22*(2), pp. 167–79.

_____. "Derivation of 'Rational' Economic Behavior from Hyperbolic Discount Curves." *American Economic Review*, May 1991 (*Papers and Proceedings*), *81*(2), pp. 334–40.

_____. *Picoeconomics: The strategic interaction of successive motivational states within the person*. New York: Cambridge University Press, 1992.

Ainslie, George and Haslam, Nick. "Self-Control," in George Loewenstein and Jon Elster, eds., *Choice over time*. New York: Russell Sage Foundation, 1992a, pp. 177–209.

_____. "Hyperbolic Discounting," in George Loewenstein and Jon Elster, eds., *Choice over time*. New York: Russell Sage Foundation, 1992b, pp. 57–92.

Ainslie, George and Herrnstein, Richard J. "Preference Reversal and Delayed Reinforcement." *Animal Learning and Behavior*, November 1981, *9*(4), pp. 476–82.

Akerlof, George A. "Procrastination and Obedience." *American Economic Review*, May 1991 (*Papers and Proceedings*), *81*(2), pp. 1–19.

Becker, Gary S.; Grossman, Michael and Murphy, Kevin M. "Rational Addiction and the Effect of Price on Consumption." *American Economic Review*, May 1991 (*Papers and Proceedings*), *81*(2), pp. 237–41.

_____. "An Empirical Analysis of Cigarette Addiction." *American Economic Review*, June 1994, *84*(3), pp. 396–418.

Becker, Gary S. and Murphy, Kevin M. "A Theory of Rational Addiction." *Journal of Political Economy*, August 1988, *96*(4), pp. 675–700.

Chung, Shin-Ho and Herrnstein, Richard J. "Choice and Delay of Reinforcement." *Journal of the Experimental Analysis of Behavior*, 1967, *10*(1), pp. 67–74.

Fischer, Carolyn. "Read This Paper Even Later: Procrastination with Time-Inconsistent Preferences." Mimeo, University of Michigan, 1997.

Funder, David C. and Block, Jack. "The Role of Ego-Control, Ego-Resiliency, and IQ in Delay of Gratification in Adolescence." *Journal of Personality and Social Psychology*, December 1989, *57*(6), pp. 1041–50.

Glazer, Jacob and Weiss, Andrew. "Conflicting Preferences and Voluntary Restrictions on Choices." Mimeo, Tel Aviv University and Boston University, 1992.

Goldman, Steven M. "Intertemporally Inconsistent Preferences and the Rate of Consumption." *Econometrica*, May 1979, *47*(3), pp. 621–26.

_____. "Consistent Plans." *Review of Economic Studies*, July 1980, *47*(3), pp. 533–37.

Hoch, Stephen J. and Loewenstein, George. "Time-Inconsistent Preferences and Consumer Self-Control." *Journal of Consumer Research*, March 1991, *17*(4), pp. 492–507.

Kahneman, Daniel; Knetsch, Jack L. and Thaler, Richard H. "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives*, Winter 1991, *5*(1), pp. 193–206.

Kahneman, Daniel and Lovallo, Dan. "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking." *Management Science*, January 1993, *39*(1), pp. 17–31.

Kreps, David M.; Milgrom, Paul; Roberts, John and Wilson, Robert. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." *Journal of Economic Theory*, August 1982, *27*(2), pp. 245–52.

Laibson, David. "Essays in Hyperbolic Discounting." Ph.D. dissertation, MIT, 1994.

_____. "Hyperbolic Discount Functions, Undersaving, and Savings Policy." Mimeo, Harvard University, 1995.

_____. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, May 1997, *112*(2), pp. 443–77.

Loewenstein, George. "The Fall and Rise of Psychological Explanations in the Economics of Intertemporal Choice," in George Loewenstein and Jon Elster, eds., *Choice over time*. New York: Russell Sage Foundation, 1992, pp. 3–34.

_____. "Out of Control: Visceral Influences on Behavior." *Organizational Behavior and Human Decision Processes*, March 1996, *65*(2), pp. 272–92.

Loewenstein, George and Prelec, Drazen. "Anomalies in Intertemporal Choice: Evidence and an Interpretation." *Quarterly Journal of Economics*, May 1992, *107*(2), pp. 573–97.

O'Donoghue, Ted and Rabin, Matthew. "Doing It Now or Later." Math Center Discussion Paper No. 1172, Northwestern University, 1996.

_____. "Incentives for Procrastinators." Math Center Discussion Paper No. 1181, Northwestern University, 1997; *Quarterly Journal of Economics*, 1999 (forthcoming).

Peleg, Bezalel and Yaari, Menahem E. "On the Existence of a Consistent Course of Action When Tastes Are Changing." *Review of Economic Studies*, July 1973, *40*(3), pp. 391–401.

Phelps, E. S. and Pollak, Robert A. "On Second-Best National Saving and Game-Equilibrium Growth." *Review of Economic Studies*, April 1968, *35*(2), pp. 185–99.

Pollak, Robert A. "Consistent Planning." *Review of Economic Studies*, April 1968, *35*(2), pp. 201–08.

Prelec, Drazen. "Decreasing Impatience: Definition and Consequences." Working Paper No. 90-015, Harvard University, 1990.

Schelling, Thomas C. "Egonomics, or the Art of Self-Management." *American Economic Review*, May 1978 (*Papers and Proceedings*), *68*(2), pp. 290–94.

_____. "Self-Command in Practice, in Policy, and in a Theory of Rational Choice." *American Economic Review*, May 1984 (*Papers and Proceedings*), *74*(2), pp. 1–11.

_____. "Self-Command: A New Discipline," in George Loewenstein and Jon Elster, eds., *Choice over time*. New York: Russell Sage Foundation, 1992, pp. 167–76.

Shefrin, Hersh M. and Thaler, Richard H. "The Behavioral Life-Cycle Hypothesis." *Economic Inquiry*, October 1988, *26*(4), pp. 609–43.

_____. "Mental Accounting, Saving, and Self-Control," in George Loewenstein and Jon Elster, eds., *Choice over time*. New York: Russell Sage Foundation, 1992, pp. 287–329.

Strotz, Robert H. "Myopia and Inconsistency in Dynamic Utility Maximization." *Review of Economic Studies*, 1956, *23*(3), pp. 165–80.

Thaler, Richard H. "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization*, March 1980, *1*(1), pp. 39–60.

_____. "Some Empirical Evidence on Dynamic Inconsistency," in Richard H. Thaler, ed., *Quasi rational economics*. New York: Russell Sage Foundation, 1991, pp. 127–33.

_____. "Psychology and Savings Policies." *American Economic Review*, May 1994 (*Papers and Proceedings*), *84*(2), pp. 186–92.

Thaler, Richard H. and Loewenstein, George. "Intertemporal Choice," in Richard H. Thaler, ed., *The winners curse: Paradoxes and anomalies of economic life*. New York: Free Press, 1992, pp. 92–106.

Thaler, Richard H. and Shefrin, Hersh M. "An Economic Theory of Self-Control." *Journal of Political Economy*, April 1981, *89*(2), pp. 392–406.

Watterson, Bill. *The days are just packed: A Calvin and Hobbes collection*. Kansas City, MO: Andrews and McMeel, 1993.

Wertenbroch, Klaus. "Consumer Self-Control: How Much Vice and Virtue Do We Buy?" Mimeo, University of Chicago, 1993.

Yaari, Menahem E. "Consistent Utilization of an Exhaustible Resource, or, How to Eat an Appetite-Arousing Cake." Center for Research in Mathematical Economics and Game Theory Research Memorandum No. 26, 1977.