To appear as a chapter in *Now or Later: Economic and Psychological Perspectives on Intertemporal Choice*, edited by Roy Baumeister, George Loewenstein, and Daniel Read, published by Russell Sage Foundation Press.

Self Awareness and Self Control

Ted O'Donoghue
Department of Economics
Cornell University

and

Matthew Rabin
Department of Economics
University of California, Berkeley

December 30, 2001

Abstract

People have self-control problems: From a prior perspective, they want to behave relatively patiently, but as the moment of action approaches, they want to behave relatively impatiently. Recently economists have studied the implications of self-control problems for a variety of economic behaviors, such as consumption-savings decisions, procrastination, addiction, information acquisition, and job search. But while there is very little behavioral evidence on degree to which people are aware of their future self-control problems, much of this economic research has assumed full awareness. In this chapter, we discuss alternative assumptions to full awareness, ranging from full unawareness to partial awareness. We argue with some simple illustrations that the degree to which a person is aware of self-control problems is a crucial determinant of the implications of those self-control problems, and hence analyses that assume complete awareness can sometimes be misleading. Because it seems clear that people are, at least to some degree, naive, we conclude that to fully understand the implications of self-control problems, researchers must seriously address the possibility of naivete.

Acknowledgments: For financial support, we thank the National Science Foundation (Awards SBR-9709485 and SES-0078796), and Rabin thanks the Russell Sage and MacArthur Foundations.

Mail: Ted O'Donoghue / Department of Economics / Cornell University / 414 Uris Hall / Ithaca, NY 14853-7601. Matthew Rabin / Department of Economics / 549 Evans Hall #3880 / University of California / Berkeley, CA 94720-3880. Email: edo1@cornell.edu and rabin@econ.berkeley.edu. CB handles: "golf boy" and "game boy". Webpages (with related papers): www.people.cornell.edu/pages/edo1/ and elsa.berkeley.edu/~rabin/index.html.

1. Introduction

People have self-control problems: From a prior perspective, they want to behave relatively patiently, but as the moment of action approaches, they want to behave relatively impatiently.[1] While the existence of self-control problems is well established and much discussed in psychological research, a standard assumption used in economic models of intertemporal choice is that a person's preferences cannot change over time. Recently, however, a small set of economists have studied the implications of self-control problems for a variety of economic behaviors, including consumption-saving decisions, procrastination, addiction, information acquisition, and job search.

When a person has self-control problems and her preferences change over time, the question arises to what extent is she aware of her own future self-control problems. While there is very little behavioral evidence, much of the economic research on self-control problems has assumed full awareness. In this chapter, we discuss alternative assumptions to full awareness, ranging from full unawareness to partial awareness. We argue with some simple illustrations that the degree to which a person is aware of self-control problems is a crucial determinant of the implications of those self-control problems, and hence analyses that assume complete awareness can sometimes be misleading.

In Section 2, we briefly describe the approach used by economists to study self-control problems. We also outline the different assumptions one might make about a person's awareness of her future self-control problems. Our analysis focuses on three possible assumptions. People could be *sophisticated*, fully aware of their future self-control problems and therefore prone to correctly predict how they will behave in the future; people could be *naive*, fully *un*aware of their future self-control problems and therefore prone to (wrongly) predict that they will behave themselves in the future; or people could be *partially naive*, aware that they will have self-control problems, but underestimating their magnitude.

In Section 3, we describe the role of awareness in some simple environments in order to outline some basic principles. Self-control problems can lead to misbehavior — behaving differently from what one would have preferred if asked from a prior perspective. In some situations, awareness has no effect on this misbehavior. More often, awareness affects behavior, but we demonstrate that awareness can sometimes mitigate and sometimes exacerbate misbehavior. We also discuss the welfare implications of self-control problems, focusing on the

question of whether self-control problems generate severe harm vs. minor suboptimalities. We show that awareness can play an important role in answering this question. In Section 4, we discuss specific economic applications, and use the basic principles described in Section 3 to frame our discussion.

In Section 5, we discuss an alternative, boundedly rational approach to incomplete awareness. Even if a person is fully aware of her future self-control problems, deciding what to do can involve some very complicated reasoning. We discuss one particular way in which a person might go about simplifying this reasoning process. Finally, we conclude in Section 6 with a few thoughts for how researchers ought to proceed in exploring the implications of self-control problems. Because it seems clear that people are, at least to some degree, naive, we conclude that to fully understand the implications of self-control problems we must seriously address the possibility of naivete.

2. Self-Control Problems and Self Awareness

Self-Control Problems

A standard assumption used in economic models of intertemporal choice is that people have *time-consistent preferences*: 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. If, for example, from a prior perspective a person prefers a larger-later reward to a smaller-sooner reward, the passage of time cannot change this preference. More concretely, if on Monday a person chooses to work on Saturday rather than Sunday, the person cannot change her mind when Saturday arrives; and if she decides to save next year, she cannot change her mind when next year arrives.

A mass of evidence, however, suggests that people have self-control problems: From a prior perspective, they want to behave relatively patiently, but as the moment of action approaches, they want to behave relatively impatiently. The existence of self-control problems is well established and much discussed in psychological research. There is a long tradition in psychology that seeks to identify the discount function used for intertemporal choice. Perhaps the most robust conclusion from this literature is that people have declining discount rates — that is, a person's relative preference for date τ over date $\tau + \Delta$ is larger the closer is τ to the present moment ("now"). This conclusion is not always explicitly framed in terms of declining discount

rates. For instance, a large strand of research discusses how hyperbolic discount functions — which impose declining discount rates — better fit the data than exponential discount functions — which impose constant discount rates. Another strand of research discusses how people exhibit "preference reversals", wherein a larger-later reward is preferred to a smaller-sooner reward when both dates are far in the future, but the sooner-smaller reward becomes preferred if both dates are moved close enough to now.[2]

In recent years, a number of economists have incorporated self-control problems into their analyses. The goal of this research has been to understand the implications of self-control problems in specific economic environments; we discuss specific applications in Section 4.[3] These researchers have modeled self-control problems in a particularly simple way, using a model originally developed by Phelps and Pollak (1968) in the context of intergenerational altruism, and later used by Laibson (1994, 1997) to model self-control problems within an individual. Let u_t be the instantaneous utility a person gets in period t, by which we mean her well-being in period t. Her intertemporal preferences at time t, U^t, can be represented by the following intertemporal utility function:

$$U^t(u_t, u_{t+1}, ..., u_T) \equiv u_t + \beta \sum_{\tau=t+1}^{T} \delta^{\tau-t} u_\tau .$$

This model is a simple modification of the standard discounted-utility model. The parameter δ is the standard discount rate, and represents "time-consistent" impatience. The parameter β introduces a time-inconsistent preference for immediate gratification, and represents the person's self-control problem. In particular, for any $\beta < 1$, at any given moment the person has an extra bias for the present over the future.

To better understand how these preferences incorporate self-control problems, consider the following example:

Example 1: Suppose a person can choose to see either *Sleepy Hollow* in period *2* or *Ed Wood* in period *3*, and these options yield the following instantaneous utilities:

Sleepy Hollow in period 2: $u_1 = 0$, $u_2 = 4$, and $u_3 = 0$.

Ed Wood in period 3 $u_1 = 0$, $u_2 = 0$, and $u_3 = 6$.

Consider the person's preferences when $\delta = 1$ and $\beta = 1/2$. From a period-*1* perspective, the person's preferences are $U^1(u_1, u_2, u_3) \equiv u_1 + (1/2)u_2 + (1/2)u_3$. Hence, the person prefers to see *Ed Wood*, because doing so yields intertemporal utility of *(1/2)6 = 3* whereas seeing *Sleepy Hollow* would only yield intertemporal utility of *(1/2)4 = 2*. When period *2* arrives, the person's preferences change to $U^2(u_2, u_3) \equiv u_2 + (1/2)u_3$. As a result, she now prefers to see *Sleepy Hollow*, because doing so yields intertemporal utility of *(1)4 = 4* whereas seeing *Ed Wood* would only yield intertemporal utility of *(1/2)6 = 3*. This example illustrates how these (β, δ) preferences give rise to a self-control problem: Whereas from a prior perspective the person wants to behave relatively patiently and attend the better movie, at the moment of action she wants to behave relatively impatiently and see the inferior movie now.

Self Awareness

When a person has self-control problems and her preferences change over time, the question arises to what extent is she aware of her own future self-control problems. Two extreme assumptions about such awareness have appeared in the economics literature on self-control problems. Most researchers assume that people are *sophisticated*, fully aware of their future self-control problems and therefore prone to correctly predict how they will behave in the future. Fewer researchers have assumed people are *naive*, fully *un*aware of their future self-control problems and therefore prone to (wrongly) predict that they will behave themselves in the future.[4]

While casual observation and introspection suggest that people lie somewhere in between these two extremes — that people are aware that they will have self-control problems, but underestimate their magnitude — the behavioral evidence is quite limited. One study worth mentioning is by Ariely and Wertenbroch (2001), discussed in Wertenbroch (this volume). They offer one group of subjects the ability to impose costly deadlines on themselves (e.g., binding deadlines for course papers), while for a second group evenly spaced deadlines are exogenously imposed. Subjects in the first group chose to impose deadlines on themselves, suggesting that they are not completely naive. But the deadlines they chose allowed more delay than evenly spaced deadlines, and by some performance measures — e.g., their grade for the course — they faired worse than people with exogenously imposed, evenly spaced deadlines. These results are

consistent with people being to some degree aware, but not completely aware, of future self-control problems.

O'Donoghue and Rabin (2001) formulate an approach to the more realistic assumption *partial naivete*. We suppose that a person has true self-control problem β, but perceives that in the future she will have self-control problem $\hat{\beta}$. Formally, we assume the person believes that in the future she will behave like a sophisticated person with self-control problem $\hat{\beta}$. Given these beliefs, the person chooses her current behavior to maximize her current preferences, which are of course determined by her true self-control problem β. With this formulation, people with standard time-consistent preferences — whom we refer to as TCs — have $\hat{\beta} = \beta = 1$, sophisticates have $\hat{\beta} = \beta < 1$, naifs have $\beta < \hat{\beta} = 1$, and partial naifs have $\beta < \hat{\beta} < 1$.

To illustrate this approach, consider what it implies in Example 1. As argued above, in period *1* the person prefers to see *Ed Wood*, while in period *2* she prefers to see *Sleepy Hollow*. What does the person believe in period 1 about her period-2 preferences? Given beliefs $\hat{\beta}$, the person perceives her period-*2* preferences to be $\hat{U}^2(u_2, u_3) \equiv u_2 + \hat{\beta}u_3$. If $\hat{\beta} < 2/3$, she believes that in period-*2* she'll prefer to see *Sleepy Hollow* — that is, she correctly predicts that her preferences will change. If, in contrast, $\hat{\beta} > 2/3$, she believes that in period-*2* she'll prefer to see *Ed Wood* — that is, she incorrectly predicts that her preferences will not change.

Whether the different beliefs about future preferences influences choice behavior depends on the specific choice environment. That is the subject of the next section.

3. Some Basic Principles

In this section, we explore the role of awareness in some simple environments. Our goal is to outline some basic principles, which we will then use to frame our discussion of specific applications in Section 4. Throughout this section, we apply the (β, δ) preferences described in Section 2, where we assume for simplicity that $\delta = 1$.

"One-Shot Decisions"

In some situations, a person's awareness of future self-control problems does not affect her behavior. The most obvious such case is a simple one-shot decision. Consider, for instance, a person who is choosing whether to have dessert. Suppose the dessert would yield an immediate benefit of, say, *5*, but would create future costs of *10*. If these were truly the only payoff consequences of having dessert, then the person's decision would be simple: She would eat the dessert if $5 - \beta(10) > 0$, or $\beta < 1/2$. In other words, she would simply implement what she currently feels to be the best decision, and her awareness of future self-control problems would be irrelevant.

Several comments about this conclusion help illustrate the ways in which awareness will matter. First, one-shot decisions need not involve only short-term behavior. If, for instance, a person must commit in January to a sequence of desserts for the next three months, she will merely choose what she currently feels to be the best sequence, and again her awareness of future self-control problems would be irrelevant. More generally, for any decisions involving long-term commitments, awareness will not play an important role.[5]

Second, if a person faces a series of completely disconnected one-shot decisions, her awareness of future self-control problems still does not matter. Suppose, for instance, that on seven consecutive nights the person must choose whether to have dessert that night. These decisions are disconnected if eating dessert on any given night does not affect the payoffs from eating dessert on any other night. If the payoffs above apply for all nights, then the person will eat dessert on all seven nights if $\beta < 1/2$, and no desserts if $\beta > 1/2$. Notice that, because the benefit is smaller than the cost, on each night the person would like to skip dessert on every future night. Also notice that her beliefs about whether she will eat desserts on future nights depends on her awareness — if $\hat{\beta} < 1/2$ she predicts that she will eat dessert on all future nights, whereas if $\hat{\beta} > 1/2$ she predicts that she will not eat dessert on any future night. But when the payoffs for the different decisions are disconnected, neither of these concerns influences her decision whether to have dessert tonight.

What makes two decisions disconnected? Formally, two decisions are disconnected if the choice for each decision does not affect the payoffs for the other decision. Whether this condition is satisfied of course depends on what the two decisions are. Should a person's beliefs about whether she will rent *Ed Wood* vs. *Sleepy Hollow* next weekend affect her decision whether to

have dessert tonight? Probably not. Should her beliefs about whether she will have hamburger vs. salad for lunch tomorrow matter? Perhaps. Should her beliefs about whether she will have a banana vs. cake late tonight matter? Probably yes.

This discussion suggests that true one-shot decisions are rare. In the dessert example, if eating dessert tonight affects the payoffs from eating dessert tomorrow night — e.g., because the cost of desserts is not linear in the number of desserts, or because the person will develop a taste for desserts, or because the person has a limited budget for desserts — then it is no longer a one-shot decision, and awareness of future self-control problems matters.[6]

When to Do an Activity

When decisions are connected, awareness can sometimes mitigate and sometimes exacerbate misbehavior due to self-control problems, depending on the environment. To demonstrate these possibilities, we present a modified version of the one-activity environment introduced by O'Donoghue and Rabin (1999*a*, 2000*a*). Suppose a person must do some activity exactly once in some finite number of periods. Each period, the person chooses only whether to do the activity then, and there are no external commitment devices available to commit future behavior. In such an environment, the implications of self-control problems, and the role of awareness, depend on whether the activity is onerous or pleasurable. We first consider the case of an onerous activity, such as writing a paper:

> **Example 2:** Suppose there is an onerous task that a person must carry out in one of the next T periods, where T might be large. The task is onerous in the sense that carrying out the task requires that the person incur an immediate cost of 10. Completing the task generates a future reward, but delay in completing the task reduces this future reward. Specifically, if the person completes the task in period 1 then she gets reward V, but each period of delay reduces the reward by $1/2$. Hence, if the person completes the task in period $k+1$ (i.e., she delays for k periods), then the reward is $V - (1/2)k$.

We examine behavior in this environment when $\delta = 1$ and $\beta = .9$. Given $\delta = 1$, standard time-consistent agents would merely complete the task in the period that maximizes the reward minus the cost, which is period 1. The behavior of time-consistent agents is a useful benchmark,

because it represents how people with self-control problems would like to behave if asked from some prior perspective. Each period, people with self-control problems trade off their desire to put off incurring the onerous task cost against the lost reward from a short delay.

Naifs, who are fully unaware of future self-control problems, procrastinate completing the task until the last possible moment — that is, until period T. Because naifs always believe they will behave themselves and choose optimally in the future, they always believe that if they delay this period then they will complete the task next period (because any further delay would be suboptimal). Hence, in each period they compare the benefit of incurring the onerous task cost in the future rather than now, which is $(1-\beta)10 = 1$, to the (discounted) lost reward from a one-period delay, which is $\beta(1/2) = .45$. Because the benefit of delay is larger than the lost reward, naifs always prefer to complete the task next period rather than this period, and so they wait in all periods and end up completing the task in period T.

Sophisticates, in contrast, who are fully aware of future self-control problems, complete the task early on — specifically, in one of the first three periods. Like naifs, sophisticates would like to delay incurring the onerous task cost. But unlike naifs, sophisticates correctly predict when they would complete the task in the future if they were to delay now. Hence, if in period *1* sophisticates choose to delay, then they must be at peace with the actual realized delay. In other words, if sophisticates delay until period τ, then in period *1* they must have preferred completion in period τ to completion in period *1*. In this example, sophisticates are willing to tolerate a two-period delay (until period *3*) but not a three-period delay (until period *4*). It follows that sophisticates must complete the task in one of the first three periods.[7]

Partial naifs with beliefs $\hat{\beta} < .909$ behave exactly like sophisticates and complete the task early on, and partial naifs with $\hat{\beta} > .909$ behave exactly like naifs and complete the task at the last possible moment. Partial naifs would like to delay, but are willing to tolerate at most a two-period delay. If their beliefs are sufficiently well calibrated — $\hat{\beta}$ close enough to β — then partial naifs predict the same future behavior as sophisticates, and hence behave exactly as sophisticates do. Suppose instead that their beliefs are sufficiently over-optimistic that they view their future tolerance for delay to be at least one day shorter than their current tolerance for delay — that is, their perceived future tolerance for delay is less than two periods. Then in all periods partial naifs believe that, if they wait now, they will complete the task within two periods. As a

result, they wait in all periods and end up completing the task in period T. In this example, the perceived future tolerance for delay is less than two if $(1-\hat{\beta})10 < \hat{\beta}(2*1/2)$ or $\hat{\beta} > .909$.[8]

Example 2 illustrates how, for an onerous activity, a preference for immediate gratification implies a tendency to procrastinate — delay completion relative to TCs. But more important for our discussion of the role of awareness, this example illustrates how naivete exacerbates misbehavior in this environment. An over-optimistic belief that you won't procrastinate in the future makes it more likely that you procrastinate now. Example 2 also illustrates how it only takes a small amount of naivete to generate severe procrastination of the form we see for complete naifs. In particular, the person severely procrastinates whenever she perceives her future self-control problem to be $\hat{\beta} > .909$, and the critical level is not much larger than her true future self-control problem of $\beta = .9$

Our conclusion that only a little naivete is required to generate severe procrastination of an onerous task is quite robust. The basic intuition is exactly as in Example 2: Whenever a person is willing to tolerate, say, a k-period delay, she only needs to be sufficiently over-optimistic to make her perceived future tolerance for delay less than or equal to $k-1$ to generate severe procrastination. This conclusion illustrates how the assumption of complete sophistication can sometimes be non-robust to a little naivete. Indeed, one can show that, while sophisticates are immune to severe procrastination, a person with any degree of naivete is subject to the possibility of severe procrastination.

We next turn our attention to the case of a pleasurable activity, such as seeing a movie.

> **Example 3:** Suppose there is a pleasurable activity that a person gets to carry out exactly once in the next T periods, where T may be large. The activity is pleasurable in the sense that carrying out the activity yields some immediate reward, and this reward is growing over time. Specifically, if the person completes the activity in period T, then she gets reward V, but if the person completes the activity prior to period T, then the reward is reduced by factor $.99$ per period. Hence, if the person completes the activity in period t, then the reward is $(.99)^{T-t}V$. Finally, there are no costs associated with the activity.

We again examine behavior when $\delta = 1$ and $\beta = .9$. Time-consistent agents complete the activity in the period with the largest reward, which is period T. Hence, if asked from a prior

perspective, people with self-control problems would most like to wait until period T. Each period, however, they trade off their desire to grab the reward now against the larger reward that comes from waiting.

Naifs are able to delay until period $T - 10$.[9] As before, naifs always believe they will behave themselves and choose optimally in the future, which here means that they always believe that if they wait, they will end up completing the activity in period T. In period t, therefore, they compare grabbing the reward now, which has value $(.99)^{T-t}V$, to waiting until period T, which has value $\beta(V)$. The larger is t, the larger is the value of grabbing the reward now. It follows that naifs do the activity the first time they prefer receiving the reward now as opposed to waiting until period T, which is period $T - 10$.

Sophisticates, sadly, are unable to delay at all, and grab the reward in period 1. Each period, sophisticates are willing to delay only if they believe they will wait more than 10 periods — because $(.99)^{T-t}V < \beta(.99)^{T-t-d}V$ only if $d > 10$. Unfortunately, an unraveling similar to that of the finitely repeated prisoner's dilemma leads them to always believe that if they delay now they'll just grab the reward next period. Because they correctly predict that they would complete the activity on date $T-1$, on date $T-2$ they realize that waiting merely means waiting one additional period, which is not worth it. Hence, they would complete the activity in period $T-2$. But this means that in period $T-3$ they correctly realize that waiting merely means waiting one additional period, and so they would complete the activity in period $T-3$. This logic iterates until sophisticates decide in period 1 that they might as well complete the activity now.

Partial naifs with beliefs $\hat{\beta} > .99$ behave exactly like naifs and delay until period $T - 10$, and partial naifs with $\hat{\beta} < .99$ behave exactly like sophisticates and don't delay at all. Intuitively, when $\hat{\beta} > .99$, partial naifs, like naifs, believe that if they wait they will end up completing the activity in period T. When $\hat{\beta} < .99$, in contrast, partial naifs experience the same unraveling logic as sophisticates, and hence, like sophisticates, always believe that if they delay now they'll just grab the reward next period.

Example 3 illustrates how, for a pleasurable activity, a preference for immediate gratification implies a tendency to preproperate — accelerate completion relative to TCs. In this environment it is *sophistication* that exacerbates misbehavior. An accurate belief that you'll just preproperate in the future reduces the value of waiting, and therefore makes you more prone to

preproperate now. Moreover, even a little awareness can induce severe preproperation. This example illustrates that sophistication effects can arise even for small degrees of awareness. Hence, merely observing that people make commitments, or seem to be worried about their own future behavior, does not mean that people are completely sophisticated.

A Richer Environment

In the simple, discrete, one-activity model described above, the effects of awareness are straightforward: For onerous activities, awareness mitigates misbehavior, while for pleasurable activities, awareness exacerbates misbehavior. In richer environments, the effects of awareness are more complicated. To illustrate, consider a simple three-period consumption-saving example. Suppose a person has income \$100,000 to allocate over three years of consumption, and assume for simplicity that the interest rate is 0%. In year 1, the person chooses how to divide her income between year-1 consumption (c_1) and saving for years 2 and 3 ($s_1 = \$100{,}000 - c_1$). Then in year 2, she chooses how to divide these savings between year-2 consumption (c_2) and year-3 consumption (c_3). We assume the person's instantaneous utility (well-being) in year t is given by $u(c_t)$.

As a benchmark, we note that if u has diminishing marginal utility — if u is concave — then TCs would merely divide their income evenly across the three years, and so they would choose $c_1 = c_2 = c_3 = \$33{,}333$. People with self-control problems are prone to consume more than \$33,333 in year 1 (and less than \$33,333 in years 2 and 3). But how much more depends (among other things) on their awareness of future self-control problems. Naifs believe that whatever they save they will divide equally between years 2 and 3, and so they believe they will choose $c_2 = c_3 = s_1/2$. People who are aware of their future self-control problems recognize that in year 2 they'll consume more than half of their savings. But whether this awareness leads them to save more or less depends on their utility function.

Table 1 describes year-1 consumption as a function of awareness. For simplicity, we assume no time-consistent impatience ($\delta=1$). We use $\beta=.7$, and we consider three different utility functions. Table 1 presents year-1 consumption for four levels of awareness: complete unawareness ($\hat{\beta} = 1$), two levels of partial awareness ($\hat{\beta} = .9$ and $\hat{\beta} = .8$), and complete awareness ($\hat{\beta} = \beta = .7$).

Insert Table 1 here.

For all three utility functions, all types over-consume in year 1 relative to the benchmark of \$33,333. The role of awareness differs across the three utility functions. For the first utility function, increased awareness exacerbates over-consumption; for the second utility function, awareness has no effect on consumption; and for the third utility function, awareness mitigates over-consumption. This ambiguity reflects the more complicated role of awareness in more general environments.[10]

The implications of awareness in this example are not nearly as dramatic as in Examples 2 and 3, particularly relative to the main effect of over-consumption due to self-control problems. For utility function $u(c) = c^{1/2}$, people with self-control problems over-consume by more than \$17,000, while the difference between sophisticates and naifs is only \$766. Similarly, for utility function $u(c) = -c^{-1/2}$, people with self-control problems over-consume by more than \$5,000, while the difference between sophisticates and naifs is only \$84. These results reflect that in some circumstances, while the degree of awareness qualitatively affects behavior, the magnitudes of these effects are small relative to the main effect of having self-control problems. We'll return to this theme in the next section.

Welfare Implications

Many researchers are interested in studying self-control problems because of their welfare implications. People with self-control problems may not behave in their own best interests — that is, self-control problems may cause a person to harm herself. By better understanding the ways in which self-control problems cause harm, we can (eventually) analyze policy interventions that might help. We now discuss some simple welfare lessons in the context of the environments discussed above.

For our welfare criterion, we ask what would the person prefer if asked at some prior ("long run") perspective. To formalize this approach, we define a person's long-run utility to be

$$U^0(u_1, u_2, ..., u_T) \equiv \sum_{\tau=1}^{T} \delta^{\tau-t} u_\tau ,$$

and then conduct welfare analysis in terms of these preferences.[11] Specifically, we measure the harm generated from self-control problems by comparing long-run utility from the person's actual behavior to long-run utility from the best thing she could have done (from a long-run perspective). Our main welfare concern is understanding when self-control problems generate severe harm vs. minor suboptimalities.

In one-shot decisions, self-control problems can generate harm, but this harm cannot be significant unless the person's self-control problem is large. To illustrate, consider an extended version of our dessert example above wherein the dessert yields an immediate benefit of V but creates future costs of C. Self-control problems generate harm whenever $V - \beta(C) \geq 0 > V - C$, because the person chooses to have dessert when from a long-run perspective she would prefer not to. But one can show that the harm generated can be at most $V(1-\beta)/\beta$. Hence, unless the person's self-control problem is large — β significantly less than one — the harm generated by this one-shot decision is small. The more general point is that, for modest self-control problems, the harm generated by any single decision to indulge must be small.

Although the harm generated by individual decisions to indulge must be small, the harm generated by many decisions to indulge can be quite large. If life consists of 100,000 unrelated opportunities to indulge, even small self-control problems can generate severe net harm. This suggests one way in which a person's awareness of future self-control problems can play an important role in determining the magnitude of harm generated by those self-control problems. If a person is aware of her future self-control problems, and in particular recognizes the situations in which she is likely to indulge, then she may make "commitments" that help prevent this indulgence. She might, for instance, alter a situation in a way that will reduce the likelihood of indulging — e.g., she makes sure to have only healthy desserts in the house — or she might choose to avoid the situation altogether — e.g., she chooses to have no desserts in the house. Naivete about future self-control problems can generate harm because the person fails to engage in such "self-management".[12]

When decisions are connected, as in the one-activity model discussed above, a person's awareness of future self-control problems has a more direct effect on the harm generated by those self-control problems. In Example 2, self-control problems generate harm when they cause the person to delay, and the harm suffered is the total delay cost incurred. Because sophisticates

delay at most two periods, they suffer harm of at most $2*(1/2) = 1$. Naifs, in contrast, delay until period T, and therefore suffer harm $(T–1)*(1/2)$, which can be much larger.

The welfare conclusions in Example 2 reflect more general intuitions. For onerous activities, one can show that sophisticates suffer at most small harm. Intuitively, when sophisticates choose to delay, they know exactly when in the future they will complete the task, and (at the moment of action) they approve of the delay. Hence, this decision to delay is essentially a single decision to indulge, and therefore the harm generated is small. Naifs, on the other hand, can suffer severe harm because they might *repeatedly* decide to complete the task in the near future. While each decision to delay generates only small harm, the net effects of many decisions to delay can be quite large. Because partial naifs can also severely procrastinate, they too can suffer severe harm. Indeed, one can show that for any degree of naivete — no matter how small — there exist environments in which the person suffers severe harm. Hence, for completion of an onerous activity, a person can suffer at most minor harm if she is completely sophisticated, but she may sometimes suffer severe harm if she has any degree of naivete.

For the case of a pleasurable activity, just as our behavioral results are reversed — awareness exacerbates misbehavior — so are our welfare conclusions. In Example 3, the harm suffered by naifs is $V – (.99)^{10}V = .096V$, while the harm suffered by sophisticates can be much larger — for $T = 200$, for instance, the harm for sophisticates is $V – (.99)^{199}V = .865V$. More generally, one can show for pleasurable activities that naifs suffer at most small harm. When naifs grab the reward, they directly compare grabbing it now to grabbing it in period T, and so they approve of their preproperation. Hence, naifs are essentially making a single decision to indulge, and the harm generated is small. Sophisticates can suffer large harm because the unraveling logic leads them to always compare grabbing the reward now vs. grabbing the reward next period. Although the harm caused by each step in this unraveling is small, the net harm can be large. Hence, for completion of a pleasurable activity, a person can suffer at most minor harm if she is completely naive, but she may sometimes suffer severe harm if she has any degree of sophistication.

While we believe long-run preferences are the appropriate welfare criterion, other researchers have been troubled by the question of how to conduct welfare analysis for a person with time-inconsistent preferences, since the person herself evaluates her well-being differently at different times. Some researchers (e.g., Goldman 1979; Laibson 1994, 1996, 1997) have

avoided this problem by using a "Pareto criterion", which says one outcome is better than another outcome when the person prefers the first outcome from all perspectives. In Example 2, for instance (and assuming T is large), the Pareto criterion says completion in period *3* is better than completion in period T, because from all perspectives the person prefers completion in period *3* to completion in period T. At the same time, the Pareto criterion does not rank completion in period *3* vs. completion in period *1*, because completion in period *3* is preferred from a period-*1* perspective while completion in period *1* is preferred from the period-*2* and period-*3* perspectives.

There are several reasons why we prefer using long-run preferences for welfare analysis. First, there is our personal belief that a preference for immediate gratification (the β) is an "error" that does not reflect true well-being. On every other day of her life, a person disagrees with her March 1, 2005 preference for immediate gratification, and so it seems wrong to give this preference for immediate gratification any weight in the person's welfare function. Second, the Pareto criterion is too weak, because it refuses to rank outcomes when one perspective barely prefers one outcome and all other perspectives vastly prefer a second outcome. For example, suppose one outcome yields utilities $u_1 = 0$ and $u_2 = 1$, while another outcome yields utilities $u_1' = 1{,}000{,}000$ and $u_2' = 0.99$. The second outcome is clearly better, and yet the Pareto criterion refuses to rank these two outcomes.[13] This usual critique of the Pareto criterion is particularly problematic when applied to intertemporal choice, because we're talking about the same person. Finally, even when the Pareto criterion does make a prediction, it does not permit an analysis of the *magnitude* of harm. Indeed, it is even easy to show that Pareto-inefficient outcomes need not generate significant harm. This limitation is problematic if we are interested in analyzing which situations we should be particularly concerned about, or which types of policy interventions are most useful (and most worth the costs of implementing the policy).

4. Some Specific Applications

In this section, we discuss the role of awareness in specific applications. In some situations, awareness is relatively unimportant; but in other situations, awareness plays an

important role in the implications of self-control problems, and analyses that restrict attention to complete sophistication can be misleading.

Procrastination

Prelec (1989), Akerlof (1991), O'Donoghue and Rabin (1999*b*, 1999*c*, 2001), and Fischer (1999) consider the role of self-control problems for procrastination.[14] Procrastination seems a natural application for self-control problems, because most people delay doing unpleasant tasks that they wish they would do sooner. Moreover, procrastination may be a realm in which the extreme assumption of complete sophistication is likely to be highly misleading. Our onerous-activity example in Section 3 illustrates this point in a very simple way, by showing that severe procrastination and severe harm can arise if and only if a person has some naivete. In this section, we further discuss this point by describing how sophistication rules out realistic-sounding behaviors that can arise if a person is naive.

In O'Donoghue and Rabin (2001), we examine the implications of choice for procrastination. In many complete-a-task situations, people have some choice over what exactly they will do. In particular, people often can put in a small amount of effort to receive small benefits, or a larger amount of effort to receive larger benefits. If, for instance, the task represents choosing how to invest some money, you might thoughtlessly follow the advice of a friend, or thoroughly investigate investment strategies.

In such environments, the two aspects of the person's decision — which task to do and when to do it — are determined by two different criteria. A person plans to do the task that yields the highest net benefits; when the person completes this task depends on her desire to put off incurring the immediate task cost. This basic insight has important implications. For instance, a naive person who would not otherwise procrastinate can be induced to severely procrastinate if she is offered a new option. This outcome occurs when the new option is better than her existing options but more onerous. If the new option is better than her existing options, the person plans to carry it out; but if at the same time the new option is sufficiently onerous, the person procrastinates.

A second implication is that a naive person may be more prone to procrastinate in pursuit of important goals than unimportant ones. The more important are the person's goals, the more ambitious she is — that is, she plans to carry out a more onerous task. The more ambitious she is,

however, the more likely she is to severely procrastinate. While increasing importance *can* make procrastination more likely, this phenomenon is clearly not universal. Indeed, increasing importance makes a person less likely to procrastinate any fixed amount of effort. But, as emphasized in the intuition above, increasing importance can exacerbate procrastination through inducing the person to plan to exert more effort.

Naivete about future self-control problems is crucial for these results. For a completely sophisticated person, additional options or increasing importance always makes performance more likely. But in O'Donoghue and Rabin (2001) we show that additional options or increasing importance can lead to procrastination if the person has *any degree of naivete*. Moreover, we show that there exist environments in which for sufficiently important goals a person surely procrastinates *unless she is completely sophisticated*.

We believe the ideas discussed above may be quite relevant for important economic contexts. In O'Donoghue and Rabin (1999*c*), for instance, we apply these ideas to the question of whether and how a person invests her savings for retirement. We argue that people may be significantly delaying transferring their savings from low-interest bank accounts into more profitable investments. We show with some simple calibrations that for complete sophistication, such delays can only be short and not very costly (in terms of lost retirement savings), while even a little naivete can generate significant and costly delays. We also demonstrate how people may be more prone to procrastinate when they have the option to exert extra effort in order to find a better investment, and how people may be more prone to procrastinate the larger is their principal.

Consumption-Saving Decisions

Laibson (1994, 1996, 1997, 1998), Laibson *et al* (1998), and Angeletos *et al* (2001, this volume) consider the role of self-control problems for consumption-saving decisions. Because consumption-saving decisions are all about trading off current well-being for future well-being, it seems a natural realm in which to consider self-control problems.

As our simple consumption-saving example from Section 3 suggests, this application is a realm where awareness sometimes does not play an important role. The main effect of having self-control problems is over-consumption, or equivalently under-saving. Being aware of future

self-control problems might mitigate or exacerbate this over-consumption, but such effects are often small relative to the main effect.

Angeletos *et al* (2001, this volume) demonstrate how self-control problems may help explain a variety of empirical observations in the consumption-saving literature that seem anomalous when viewed through the lens of the standard economic model. They calibrate models of consumption-saving decisions using both the standard assumption that households are time-consistent and the assumption that households have sophisticated self-control problems. By comparing their simulated data for the two cases to real-world data, they demonstrate that sophisticated self-control problems are better able to explain things such as people's liquid asset holdings (relative to monthly income and relative to illiquid asset holdings), consumption-income comovement, drops in consumption at retirement, and high credit-card debt. Moreover, they claim that their conclusions would be the same if they were to consider naive self-control problems instead of sophisticated self-control problems, again reflecting that in some ways awareness may not matter in this realm.

Even so, there are some areas in consumption-savings decisions where awareness is clearly more important. Laibson (1997), for instance, focuses on the use of illiquid assets such as housing in creating commitment opportunities, and the detrimental effects of credit cards in creating liquidity. The intentional use of illiquid assets to influence future behavior clearly relies on people being aware of their self-control problems. Because it seems clear that people do in fact engage in such behavior — e.g., Christmas clubs — it seems clear that people are to some degree aware. Even so, it is unclear whether people in the real world are using as many commitment devices as one might expect from completely sophisticated people.

A closely related point is that if people are naive, then they won't fully recognize the commitment value of certain savings instruments, and there might be value in creating *incentives* for people to use these savings instruments. Consider the use of 401k retirement plans. Because such plans typically impose significant penalties for early withdrawals, they represent a commitment device to help keep oneself from consuming one's savings. If people are sophisticated, this commitment feature makes these instruments quite valuable independent of their tax-exempt features (see Laibson *et al* (1998) for a more complete discussion of the sophisticated case). If people are not completely aware, in contrast, the tax-exempt features may be crucial, because people may not see the commitment value of 401k plans. In other words,

naive people may use 401k plans because of the tax benefits, and end up reaping unexpected commitment benefits as well.

Addiction

O'Donoghue and Rabin (1999*d*, 2000*b*), Gruber and Koszegi (2001), and Carrillo (1999) consider the role of self-control problems for addiction. One key feature of harmful addictive products such as cigarettes is that current consumption creates future costs, and hence the decision whether to consume boils down to whether the current desire to consume outweighs the future cost of this consumption. Because self-control problems make a person more prone to conclude that current consumption is worthwhile, the main effect of self-control problems is a tendency to over-consume addictive products.

But since a second key feature of harmful addictive products is that current consumption increases the future desire to consume, awareness of future self-control problems may be particularly important in this realm. O'Donoghue and Rabin (1999*d*, 2000*b*) and Gruber and Koszegi (2001) both show that in stationary environments — where the desire to consume depends on past consumption but is otherwise constant over one's lifetime — sophistication exacerbates over-consumption for people who are relatively unaddicted, but sophistication mitigates over-consumption for people who are relatively addicted. The intuition behind this dichotomy is that the implications of sophistication depend on two effects. First, there is a pessimism effect: Sophisticates expect more future consumption than do naifs, and this pessimism tends to exacerbate over-consumption — if you're going to consume in the future, you might as well start now. Second, there is an incentive effect: Sophisticates may restrict current consumption so as to reduce the desire to smoke in the future — by consuming less now, you're less likely to consume in the future. The latter effect becomes more powerful as a person gets more addicted.

O'Donoghue and Rabin (2000*b*) go further, attempting to identify the situations in which self-control problems can generate *severe* harm in the realm of addiction. Under the assumption of complete sophistication, addiction is problematic only to the extent that the person feels that future consumption is inevitable. Sophisticates suffer severe harm when they feel that they will consume in the future no matter what they do now, and because of this inevitable future consumption they decide they might as well start consuming now. This inevitability might take

the extreme form of expecting to consume throughout one's life, or the less extreme form of expecting to consume throughout one's years at college. But since in real-world environments extreme inevitability seems unlikely, sophisticated self-control problems may not represent a realistic source of harmful addictions.

Naifs, in contrast, suffer severe harm when they repeatedly plan to consume only in the short run but end up consuming in the long run. As a result, naive self-control problems can give rise to more realistic-sounding, and plausibly quite harmful, behaviors. For instance, an addicted person might suffer severe harm because she procrastinates quitting — she wants to quit, and always plans to quit in the near future, but never gets around to it. Moreover, naifs might develop severely harmful addictions in the first place because they naively give in to temporarily high temptations believing they'll just quit after the temptation subsides, when in fact they end up with long-term addictions.

Acquiring Information

Carrillo and Mariotti (2000), Benabou and Tirole (2000*a*, 2000*b*), and Brocas and Carrillo (1999, 2000) apply self-control problems to information-acquisition decisions. In many situations, people have the option to acquire information about the costs and benefits of their actions. If, for instance, a person is deciding whether to embark on a specific research agenda, she has the option to get feedback from colleagues about the likely fruitfulness of that research agenda. The standard economic model implies that, if such information is free, people should always choose to acquire it. The intuition is simple: Having more information allows people to make better decisions.

Carrillo and Mariotti (2000) point out, however, that sophisticated self-control problems might change this intuition, because better decisions as viewed from the future — i.e., after receiving the information — may not correspond to better decisions as viewed from the present moment. Carrillo and Mariotti emphasize the possibility of "strategic ignorance" — not acquiring free information because doing so increases the likelihood of future misbehavior. Benabou and Tirole (2000*a*, 2000*b*) explore strategic ignorance in more detail, and explore the role of memory manipulation in this regard.

These results about self-control problems leading to the active manipulation of information (relative to optimal information acquisition) obviously require some degree of

sophistication, because they are driven by a desire to reduce future misbehavior. While the literature has focused on strategic ignorance, there may in fact be a variety of other embodiments of sophisticated information manipulation. For instance, whereas strategic ignorance involves people not gathering information that they should because the information *increases* the likelihood of future misbehavior, people might also gather (costly) information that they shouldn't when the information *decreases* the likelihood of future misbehavior (Carrillo and Mariotti briefly mention this possibility in their conclusion). More generally, sophisticated information manipulation might involve selective information gathering: When multiple sources of information are available, people will be biased against information that increases the likelihood of future misbehavior, and biased in favor of information that decreases the likelihood of future misbehavior.

But given its focus on complete sophistication, this literature ignores what we suspect may be an important source of distortions in information acquisition: procrastination in information gathering. Because information gathering itself is likely to be onerous — and create immediate costs — people with self-control problems may be prone to put off incurring these costs. As a result, our procrastination results have implications for information acquisition. For instance, people might fail to acquire information not because they are worried about future behavior, but rather because they never get around to it. Or if people are choosing from among multiple sources of information, they may be biased towards information that is less onerous to acquire. Moreover, while mild versions of such distortions may arise even for people who are sophisticated, these distortions are likely to be particularly pronounced for people who are not completely aware of their self-control problems.

5. Boundedly Rational Incomplete Awareness

In this section, we briefly speculate on an alternative, boundedly rational approach to incomplete awareness. To motivate this approach, consider the problem faced by naifs. It is quite simple — in order to choose their current behavior, naifs merely need to know their current preferences, and then they simply start following their most preferred behavior path.

Sophisticates of course also need to know their current preferences; but in order to carry out the backwards-induction logic required to derive the sophisticated path, they must know

much more. They must know what their preferences will be in the last period, so that they can derive the behavioral rule that they will follow in that period. They also must know what their preferences will be in the second-to-last period, which they must then combine with their behavioral rule for the last period to determine the behavioral rule that they will follow in the second-to-last period. They also must know what their preferences will be in the third-to-last period, which they must then combine with their behavioral rules for the last and second-to-last periods to determine the behavioral rule that they will follow in the third-to-last period. And so forth.

Clearly, for sophisticated people deciding what to do can involve a rather complicated reasoning process. It seems plausible that sophisticated people, even though they are fully aware of their future self-control problems, might follow some simplified procedure in deciding how to behave. We consider one particular simplification: They don't do all the rounds of backwards induction. In other words, instead of starting the backwards-induction logic in the last period, they might start the process, say, three periods hence. What would this mean? It would merely mean that in deciding what to do, sophisticates first ask what their preferences will be three periods hence, and use those preferences to derive the behavioral rule that they will follow from that period onward. They then continue the backwards induction from there. Sophisticates might think in this way as a useful heuristic to simplify their decision-making process. Such boundedly rational sophisticates would have incorrect predictions about future behavior, and might be aware of this fact, but they would view it as too costly to think through their decisions more carefully.

To illustrate this approach, we reconsider the simple consumption-saving example from Section 3, except that now we assume the person must allocate her \$100,000 income over four years of consumption. As before, we assume no interest ($r = 0\%$) and no time-consistent impatience ($\delta = 1$), and we use $\beta = .7$. Table 2 describes year-1 consumption for three types: naifs, complete sophisticates who do the entire backwards induction starting from the end (year 3), and boundedly rational sophisticates who do the backwards induction starting from year 2.[15] We note as a benchmark that time-consistent people would consume \$25,000 in each year.

Insert Table 2 here.

The simplified solution procedure fares quite well in this environment. For all three utility functions, boundedly rational sophisticates consume roughly the same amount as both sophisticates and naifs. This conclusion is largely driven by our earlier conclusion that in this environment the degree of awareness is relatively unimportant. Our point, however, is that even if people are fully aware of their future self-control problems, it might not make sense to assume that they carry out the complete, complicated, backwards-induction logic when making their decisions.

What does this approach rule out? In this example, boundedly rational sophisticates exhibit what might be called first-order sophistication effects: They recognize that they will over-consume in year 2, and their year-1 consumption partially reflects a reaction to this future over-consumption. But boundedly rational sophisticates do not exhibit what might be called second-order sophistication effects: Complete sophisticates also recognize that they will over-consume in year 3, and that their year-2 behavior will partially be driven by reactions to year-3 over-consumption, and hence their year-1 consumption also reflects reactions to future first-order sophistication effects.

We find it appealing that this approach rules out higher order sophistication effects, because we feel that results driven by higher order sophistication effects rely a little too heavily on the game-theoretic approach. The standard game-theoretic approach to sophistication that is used in the literature represents a nice formal way to understand the behavior of a person who is aware of future self-control problems. But it seems to push the framework a little too far to assume that the person literally carries out the complicated backwards-induction procedure. While results that rely on first-, second-, and perhaps even third-order sophistication effects seem quite reasonable, results that rely on higher-order sophistication effects seem less reasonable.

An important question for this approach is how often a person revises her plans. In the consumption-saving example above, when year 2 arrives, does the person stick to her year-1 plan, or does she re-evaluate her plan now doing the backward induction starting from year 3? In some situations — e.g., the one-activity model discussed in Section 3 — the answer to this question is critical. If she sticks to her plan and only occasionally re-evaluates, then her behavior will be closer to sophisticated behavior. If she constantly re-evaluates, then her behavior will be

closer to naive behavior. This approach suggests that another interpretation of naivete is as an approximation of what a boundedly rational sophisticate might do.

6. Discussion

We have attempted to illustrate throughout this chapter how the degree to which a person is aware of her own future self-control problems can be a crucial determinant of the implications of those self-control problems. Perhaps our main message has been directed at economists prone to assume complete awareness: Because awareness can play such a crucial role, and because it seems clear that people are, at least to some degree, naive, for economists to fully understand the implications of self-control problems they must seriously address the possibility of naivete. Indeed, in this chapter we have discussed how even a small degree of unawareness can give rise to realistic behaviors that could not arise under complete awareness, and how even a small degree of unawareness can give rise to different welfare conclusions.

We conclude with a few thoughts for how researchers ought to proceed in exploring the implications of self-control problems. First, one reason to study naivete is tractability. We have illustrated how in some circumstances the implications of sophistication vs. naivete are small relative to the direct effect of people having self-control problems. Our analysis in Section 5 also makes clear that the sophisticated path can be quite complicated to derive, whereas the naive path — or something similar — is relatively easy to derive. Hence, from a modeling perspective — and perhaps also from the agent's own perspective given our discussion in Section 5 — it may make sense to consider the more simple case of naivete. Indeed, tractability was exactly the motivation behind Pollak (1968) first introducing a formalization of the naive path. Pollak was interested in the sophisticated path, but since it was difficult to solve for, he identified a situation for which he could prove that the sophisticated path coincides with the easy-to-solve-for naive path.

Second, we believe the literature on self-control problems should focus more explicitly on when small self-control problems can cause severe harm. Much of the recent literature has focused on the ways in which models with self-control problems can better explain observed real-world behavior. We think this agenda is very important. But at the same time, at least implicit in these analyses is the policy concern of what types of policies might help people make

better decisions. In order to seriously address this policy concern, and evaluate the costs and benefits of proposed policies, analyses must formally analyze the magnitude of harm caused by self-control problems. As we have emphasized in this chapter, in some situations, self-control problems may lead people to misbehave, and yet this misbehavior may not be very important.

A related comment is that welfare concerns may be the main reason for being concerned with people's awareness of self-control problems. We have argued in the past that welfare concerns are perhaps the main reason for considering models of self-control problems (as well as other behavioral errors). Even if we can explain some behavior with a rational-choice model, an interest in welfare analysis implies that we should still be concerned with whether this explanation is correct. This same argument applies to the question of sophistication vs. naivete. In some situations, we may be able to explain some behavior with a model of self-control problems and complete sophistication; but an interest in welfare analysis implies that we should still ask whether the behavior is driven by sophistication vs. naivete, because there might be dramatically different welfare implications. As a simple example suggested by our analysis in Section 3, a model of self-control problems and sophistication can explain time-inconsistent procrastination in completing an onerous task, but would force us to conclude that the harm caused by such procrastination is not significant. If in fact the person is naive, however, then in fact the harm caused may be quite large.

Indeed, we believe that in real-world situations sophisticated self-control problems may be even less problematic than suggested by simple economic models. There likely exist situations in the real world in which sophisticates would suffer severe harm, but real-world sophisticates would constantly be on the lookout to avoid such situations. This type of meta-behavior would not be carried out by naifs.

It's not yet established whether we're right or wrong about people being naive; but by demonstrating the specialness of the extreme assumption of complete awareness, we hope we have convinced economists and other researchers that we ought to be interested in finding out.

References

Ainslie, George. 1975. "Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control." *Psychological Bulletin* 82(4): 463-496.

------. 1991. "Derivation of 'Rational' Economic Behavior from Hyperbolic Discount Curves." *American Economic Review* 81(2): 334-340.

------. 1992. *Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person*. New York: Cambridge University Press.

Ainslie, George, and Nick Haslam. 1992*a*. "Self-Control." In *Choice Over Time*, edited by George Loewenstein and Jon Elster. New York: Russell Sage Foundation.

------. 1992*b*. "Hyperbolic Discounting." In *Choice Over Time*, edited by George Loewenstein and Jon Elster. New York: Russell Sage Foundation.

Akerlof, George. 1991. "Procrastination and Obedience." *American Economic Review* 81(2): 1-19.

Angeletos, George-Marios, David Laibson, Andrea Repetto, Jeremy Tobacman, and Stephen Weinberg. 2001. "The Hyperbolic Buffer Stock Model: Calibration, Simulation, and Empirical Evaluation." *Journal of Economic Perspectives* 15(3), 47-68 (reprinted in this volume).

Ariely, Daniel, and Klaus Wertenbroch. 2001. "Procrastination, Deadlines, and Performance: Using Public Commitments to Regulate One's Behavior." *Psychological Science*.

Baumeister, Roy, and Kathleen Vohs. This volume.

Benabou, Roland, and Jean Tirole. 2000*a*. "Self-Confidence: Intrapersonal Strategies." Princeton University Discussion Paper #209.

------. 2000*b*. "Willpower and Personal Rules." Unpublished paper. Princeton University and MIT.

Brocas, Isabelle, and Juan Carrillo. 1999. "A Theory of Haste with Applications to Impulse Buying and Destruction of the Environment." CEPR Discussion Paper No. 2027.

------. 2000. "The Value of Information When Preferences are Dynamically Inconsistent." *European Economic Review* 44: 1104-1115.

Carrillo, Juan. 1999. "Self-Control, Moderate Consumption, and Craving." CEPR Discussion Paper 2017.

Carrillo, Juan, and Thomas Mariotti. 2000. "Strategic Ignorance as a Self-Disciplining Device." *Review of Economic Studies* 67: 529-544.

Fischer, Carolyn. 1999. "Read This Paper Even Later: Procrastination with Time-Inconsistent Preferences." Resources for the Future Discussion Paper 99-20.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue. 2002. "Time Discounting: A Critical Review." *Journal of Economic Literature* (reprinted in this volume).

Goldman, Steven. 1979. "Intertemporally Inconsistent Preferences and the Rate of Consumption." *Econometrica* 47(3): 621-626.

Gruber, Jonathan, and Botond Koszegi. 2001. "Is Addiction 'Rational'? Theory and Evidence." *Quarterly Journal of Economics* 116(4): 1261-1303.

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

------. 1996. "Hyperbolic Discount Functions, Undersaving, and Savings Policy." NBER Working Paper 5635.

------. 1997. "Hyperbolic Discounting and Golden Eggs." *Quarterly Journal of Economics* 112(2): 443-477.

------. 1998. "Life-Cycle Consumption and Hyperbolic Discount Functions." *European Economic Review* 42: 861-871.

Laibson, David, Andrea Repetto, and Jeremy Tobacman. 1998. "Self-Control and Saving for Retirement." *Brookings Papers on Economic Activity* 1: 91-196.

Loewenstein, George, and Drazen Prelec. 1992. "Anomalies in Intertemporal Choice: Evidence and an Interpretation." *Quarterly Journal of Economics* 107(2): 573-597.

O'Donoghue, Ted, and Matthew Rabin. 1999*a*. "Doing It Now or Later." *American Economic Review* 89(1): 103-124.

------. 1999*b*. "Incentives for Procrastinators." *Quarterly Journal of Economics* 114(3): 769-816.

------. 1999*c*. "Procrastination in Preparing for Retirement." In *Behavioral Dimensions of Retirement Economics*, edited by Henry Aaron. Washington D.C. and New York: Brookings Institution Press & Russell Sage Foundation.

------. 1999*d*. "Addiction and Self Control." In *Choice Over Time*, edited by George Loewenstein and Jon Elster. New York: Russell Sage Foundation.

------. 2000*a*. "The Economics of Immediate Gratification." *Journal of Behavioral Decision Making* 13(2): 233-250.

------. 2000*b*. “Addiction and Present-Biased Preferences.” Unpublished paper. Cornell University and U.C. Berkeley.

------. 2001. “Choice and Procrastination.” *Quarterly Journal of Economics* 116(1): 121-160.

Phelps, E. S. and Robert A. Pollak. 1968. “On Second-Best National Saving and Game-Equilibrium Growth.” *Review of Economic Studies* 35: 185-199.

Pollak, Robert A. 1968. “Consistent Planning.” *Review of Economic Studies* 35: 201-208.

Prelec, Drazen. 1989. “Decreasing Impatience: Definition and Consequences.” Unpublished paper. Harvard Business School.

Read, Daniel. This volume. “Is Time-Discounting Hyperbolic or Subadditive?”

Strotz, Robert. 1956. “Myopia and Inconsistency in Dynamic Utility Maximization.” *Review of Economic Studies* 23(3): 165-180.

Thaler, Richard. 1991. “Some Empirical Evidence on Dynamic Inconsistency.” In *Quasi Rational Economics*. New York: Russell Sage Foundation.

Thaler, Richard, and George Loewenstein. 1992. “Intertemporal Choice.” In *The Winner’s Curse: Paradoxes and Anomalies of Economic Life*, edited by Richard Thaler. New York: Free Press.

Wertenbroch, Klaus. This volume.

Notes

1. This basic phenomenon goes by a variety of names in addition to "self-control problems" — e.g., "hyperbolic discounting", "present-biased preferences", and "preference for immediate gratification".

2. For an overview of this evidence, see Frederick, Loewenstein, and O'Donoghue (2002, this volume), and also Ainslie (1975, 1991, 1992), Ainslie and Haslam (1992*a*, 1992*b*), Loewenstein and Prelec (1992), Thaler (1991), and Thaler and Loewenstein (1992). See, however, Read (this volume) for a critique of much of the evidence for hyperbolic discounting, which is confounded with a subadditivity in discounting.

3. There is an older literature in economics, pioneered by Strotz (1956) and Pollak (1968), that analyzes general abstract models of time-inconsistent preferences. This literature does not make any assumptions about the form of time inconsistency — allowing for self-control problems but also other forms of time inconsistency — nor does it discuss much the implications of time inconsistency.

4. Strotz (1956) and Pollak (1968) carefully lay out these two assumptions (and develop the labels), but do not much consider the implications of assuming one versus the other.

5. As another example, the decision whether to buy a durable good — e.g., an automobile — can be interpreted as a long-term commitment, and hence awareness should not be important.

6. Recent research in psychology suggests that seemingly unrelated decisions may in fact be connected through limited "willpower". For instance, while the decision which movie to watch tonight may seem unrelated to the decision whether to eat dessert tonight, having to make the former decision first may affect the willpower available for the latter decision. For an overview, see Baumeister and Vohs (this volume).

7. Exactly when sophisticates complete the task depends on T. Doing the backwards induction, we can derive that sophisticates' plan must involve planning to complete the task in periods T, $T-3$, $T-6$, and so on, and planning to wait in all other periods. Hence, if, for instance, $T=63$, then sophisticates complete the task in period *3*; if $T=64$, then sophisticates complete the task in period *1*; and if $T=65$, then sophisticates complete the task in period *2*.

8. This logic is more carefully formalized in O'Donoghue and Rabin (2001).

9. If $T < 10$, then naifs do the activity in period *1*.

10. The three utility functions are from the family of constant-relative-risk-aversion (CRRA) utility functions, and have coefficients of relative risk aversion equal to *1/2*, *1*, and *3/2*, respectively. Pollak (1968) showed that awareness is irrelevant for logarithmic utility; more generally, within the family of CRRA utility functions, awareness exacerbates over-consumption for relative risk aversion less than unity, and awareness mitigates over-consumption for relative risk aversion greater than unity. Also note that the degree of over-consumption is decreasing in the degree of risk aversion (amount of curvature in the utility function).

11. These preferences are the same as the preferences represented by $U^0(u_1,u_2,...,u_T) \equiv \beta\sum_{\tau=1}^{T}\delta^{\tau-t}u_\tau$. Also, for "small" self-control problems (β close to 1), long-run preferences are quite similar to period-*1* preferences $U^1(u_1,u_2,...,u_T) \equiv u_1 + \beta\sum_{\tau=2}^{T}\delta^{\tau-t}u_\tau$, and yield similar welfare conclusions.

12. While such commitments are often discussed in the literature, most formal models incorporate rather limited commitment technologies. As a result, such models may over-state the harm caused sophisticated self-control problems.

13. The Pareto criterion's unwillingness to designate the first outcome as inefficient *holds even for time-consistent agents*, illustrating that the Pareto criterion is not what is used for welfare analysis in standard models of intertemporal choice. Indeed, the long-run-utility criterion (or perhaps the closely related period-*1*-utility criterion) is equivalent to what is used in standard analyses.

14. While Akerlof (1991) does not frame his analysis in terms of self-control problems, his model implicitly corresponds to *(β,δ)* preferences.

15. Because in year 4 the person merely consumes whatever is left over, the last substantive decision occurs in year 3.

Table 1: Year-1 consumption (c_1) given \$100,000 to allocate over three years, assuming $r = 0\%$, $\delta = 1$, and $\beta = .7$.

Utility Function	Naifs	Partial Naifs		Sophisticates
	$\hat{\beta} = 1$	$\hat{\beta} = .9$	$\hat{\beta} = .8$	$\hat{\beta} = .7$
$u(c) = c^{1/2}$	$c_1 =$ \$50,505	$c_1 =$ \$50,574	$c_1 =$ \$50,812	$c_1 =$ \$51,271
$u(c) = \ln c$	$c_1 =$ \$41,667	$c_1 =$ \$41,667	$c_1 =$ \$41,667	$c_1 =$ \$41,667
$u(c) = -c^{-1/2}$	$c_1 =$ \$38,809	$c_1 =$ \$38,801	$c_1 =$ \$38,776	$c_1 =$ \$38,725

Table 2: Year-1 consumption (c_1) given \$100,000 to allocate over four years, assuming $r = 0\%$, $\delta = 1$, and $\beta = .7$.

Utility Function	Naifs	Sophisticates	Boundedly Rational Sophisticates
$u(c) = c^{1/2}$	$c_1 =$ \$40,486	$c_1 =$ \$41,781	$c_1 =$ \$41,229
$u(c) = \ln c$	$c_1 =$ \$32,258	$c_1 =$ \$32,258	$c_1 =$ \$32,258
$u(c) = -c^{-1/2}$	$c_1 =$ \$29,717	$c_1 =$ \$29,601	$c_1 =$ \$29,650