Chapter 6

Addiction and Self-Control

TED O'DONOGHUE AND MATTHEW RABIN

MANY OBSERVERS suspect that self-control problems and related time inconsistencies play an important role in the consumption of addictive products, leading people to develop and maintain addictions against their long-run interests. People often consume addictive products despite an expressed desire to quit. For many people, it would appear that the long-run harm caused by an addiction outweighs its short-run benefits. In extreme cases, people destroy their lives with harmful addictions. Our goal in this chapter is to carefully explore the role that self-control problems—and people's awareness of those problems—play in harmful addictions. To do so, we develop a formal model of the decision to consume addictive products that explicitly incorporates a time-inconsistent taste for immediate gratification.

Economists have proposed rational choice models of addictive behavior (Becker and Murphy 1988; Becker, Grossman and Murphy 1991, 1994). These models characterize how consuming harmful addictive products can decrease future well-being while at the same time increasing the desire for those products in the future. Because these models consider only time-consistent agents, however, they a priori rule out the possibility of self-control problems.

Like the rational choice models of addiction, our model assumes that the choice to consume an addictive product is volitional, in the sense that people balance their current desire for the addictive product against their perceptions of the future consequences of current consumption. Our model is quite different, and less extreme, than rational choice models, however, because it assumes that people may be overattentive to their immediate gratification (that is, they may have self-control problems) and

may have incorrect beliefs about their future behavior (that is, they may not anticipate future self-control problems).

In the next section, we lay out our formal model. We assume that in each period people can either take a hit or not take a hit.[1] We incorporate two crucial characteristics of harmful addictive products. First, they involve *habit formation:* The more of the product people have consumed in the past, the more they desire that product now. Second, they involve *negative internalities:* The more of the product people have consumed in the past, the lower is their overall well-being now (regardless of current behavior).[2] The combination of habit formation and negative internalities implies that as people consume more and more of an addictive product, they get less and less pleasure from its consumption, yet they may continue to consume the product because refraining becomes more and more painful.

We incorporate self-control problems into the model by assuming that people have time-inconsistent intertemporal preferences. We apply a simple model of time-inconsistent preferences, originally proposed by Edmund S. Phelps and Robert A. Pollak (1968) in the context of intergenerational altruism, and later employed by David Laibson (1994a) to capture self-control problems within individuals: Relative to time-consistent preferences, people always give extra weight to well-being *now* over well-being at any future moment. These preferences give rise to self-control problems because at any moment people pursue immediate gratification more than they would have preferred if asked at any previous moment.

In addition to the implications of having self-control problems, we also focus on the implications of whether people are aware of their own future self-control problems. We examine two extreme assumptions: *Sophisticated* people are fully aware of their future self-control problems and therefore know exactly how they will behave in the future; and *naive* people are fully *un*aware of their future self-control problems and therefore believe they will behave in the future exactly as they currently would like themselves to behave. By systematically comparing *sophisticates, naifs,* and time-consistent agents (whom we refer to as *TC*s), we can examine the role of self-control problems in addiction and delineate how predictions depend both on self-control problems per se and on assumptions about foresight.

We begin with a stationary model of addiction, in which the temptation to hit can depend on the addiction level but otherwise remains constant over time, which allows us to identify some basic insights. We first ask what is the direct implication of self-control problems by comparing *TC*s and naifs. In the stationary model, naifs are always more likely to hit than *TC*s. Since naifs are unaware of future self-control problems, they perceive that they will behave exactly like *TC*s in the future and

therefore perceive the same future consequences of current indulgence as do *TC*s. Given their overattentiveness to immediate gratification, however, naifs are more likely to hit than *TC*s. Clearly, this intuition is far more general than the model of stationary preferences: In essentially any model of addiction, self-control problems combined with an unawareness of future self-control problems will cause people to consume more of an addictive product than they would like to consume from a long-run perspective.

We next ask what are the implications of being aware of future self-control problems by comparing naifs and sophisticates. We identify two effects. First, sophistication about future self-control problems can make people pessimistic about future behavior (that is, they believe in general that they will hit more often than they would if they had no self-control problem). We refer to this phenomenon as the *pessimism effect*. Second, sophistication about future self-control problems may make people realize that they will resist future temptations only if they resist temptation today. We refer to this phenomenon as the *incentive effect*. Because the habit formation property of addictive products implies that current indulgence has larger future costs the more people expect to refrain in the future, pessimism about future behavior tends to exacerbate overconsumption due to self-control problems. The incentive effect, in contrast, tends to mitigate overconsumption due to self-control problems. Hence, whether sophisticates hit more or less often than naifs depends on the relative magnitudes of the pessimism and incentive effects.

Of course, since the incentive effect is driven by future restraint, it can be operative only if there is some future period where people would refrain in the face of pure pessimism. Consider the implications of this point in a stationary model. If in period 1 people would hit when "unhooked" in the face of pure pessimism, then in all periods they would hit when unhooked in the face of pure pessimism, and therefore the incentive effect cannot be operative. In contrast, if in period 1 people would refrain when unhooked in the face of pure pessimism, then in all periods they would do so, and therefore the incentive effect can be operative. This logic implies that if people are initially unhooked, the incentive effect can be operative if and only if people would refrain without it. Since the pessimism effect makes sophisticates more likely to hit than naifs, we can therefore conclude that sophisticates are more likely than naifs to become addicted starting from being unhooked.

This logic does not imply that sophisticates are more likely to hit than naifs once hooked. Even if people would hit when hooked in the face of pure pessimism, refraining may reduce their addiction level to a point at which they would refrain in the face of pure pessimism, in which case the

incentive effect would be operative. Indeed, in our model sophisticates are always more likely than naifs to quit an established addiction.

We then consider nonstationary environments. First, we consider a model of youth, wherein the intrinsic temptation to hit is high early in life but declines as people get older. Second, we consider a weekend-weekday model, wherein the temptation to hit alternates between high (on weekends) and low (on weekdays). Third, we briefly discuss temporary temptations arising from traumatic events such as a divorce or a death of a loved one. Some examples in these nonstationary environments illustrate that the result that naivete helps people avoid harmful addictions is very special to the stationary environment. There are two reasons for this reversal. First, sophisticates may consume less in nonstationary environments because the incentive effect becomes operative in a broader array of circumstances. In particular, the incentive effect being operative merely requires that people refrain in the face of pure pessimism when the temptation to consume is lowest. For instance, in the youth model this means that people refrain when unhooked in the face of pure pessimism *in their old age;* and in the weekend-weekday model this means that people refrain when unhooked in the face of pure pessimism *on weekdays*. Second, naifs may consume more in nonstationary environments because of their aforementioned tendency not to quit an established addiction. When it is optimal to give in to high temptations and later quit, naifs often give in to high temptations and then never quit. Because we suspect nonstationary environments are more prevalent, we tentatively interpret such results to say that "sophisticated self-control problems" are not a major source of harmful addictions. If self-control problems help explain severely harmful addictions, we suspect they do so only in conjunction with some degree of naivete.

We extend our model to incorporate different types of "variable myopia." First, we consider *consumption-induced myopia*—we suppose that self-control problems may depend on recent consumption. When people are sober, they might have very mild self-control problems. Once they have had a few drinks, however, they may suddenly have significant self-control problems. Second, we consider exogenous variation in the taste for immediate gratification. These extensions allow us to further highlight the importance of fully understanding one's self-control problems. Consumption-induced myopia (in addition to basic self-control problems) always makes naifs consume more of an addictive product but may induce sophisticates—because of their fear of addiction—to consume even less of the addictive product than if they had no self-control problem. With an exogenously varying taste for immediate gratification, naifs—while consuming more than they would if they had no self-control problem—may consume

too little relative to sophisticates, because they undertake repeated, costly, unsuccessful attempts to quit their addiction under the naive belief they can stay unhooked.

We conclude the chapter by comparing our model of addiction to time-consistent models of addiction. We feel that studying self-control as it relates to addiction is an obviously appropriate line of research because self-control problems seem to exist and seem to be important. We also conjecture that research on addiction might be improved if researchers choose to investigate self-control problems rather than solely investigating the extreme time-consistent model. We then conclude by discussing what we suspect is on most people's minds when studying addiction—the degree to which people hurt themselves by becoming addicted. Rational choice models do not and cannot address the question of when and how people systematically hurt themselves by becoming addicted—except to assume the question away a priori. Especially because we illustrate at the end of the chapter that even modest self-control problems can hurt people severely, we feel that formulating models as a means for understanding when and how people might hurt themselves is an important agenda.

The Basic Model

We consider a discrete-time model with periods 1, ···, T, wherein we consider both $T < \infty$ and $T = \infty$.[3] We vastly simplify the model by assuming that in each period, t, consumption of an addictive product, a_t, is either 0 or 1: Each period people can either take a hit or not take a hit, wherein $a_t = 1$ if they take a hit and $a_t = 0$ if they refrain. Furthermore, we assume that the good is free. Our focus on free products helps highlight the fact that people may avoid addictive products because they lead to unpleasant long-run consequences, rather than because of the purchase price per se. It also simplifies notation and analysis.

Each period, people merely choose whether to hit this period (and cannot commit to any future choices).[4] Choice is rational or volitional in the sense that people balance their current desire for the good against the future consequences of consumption, given their current beliefs about their future behavior. Hence, whenever people take a hit, they are doing what currently seems to them to be the best course of action, with the important caveat that they may be overweighting their current well-being relative to their future well-being. In this sense, our model does not abandon the economic paradigm of considering human choice as balancing the benefits and costs of a course of action. As discussed in the introduction, however, our model is quite different from the rational choice models of addiction because we allow people to have

self-control problems and incorrect beliefs about their future behavior. As a result, our model does not necessarily imply that people will follow their most preferred lifetime path of behavior.

The crucial feature of addictive products is that past consumption affects current well-being. Gary S. Becker and Kevin M. Murphy (1988) provide a model of "instantaneous utility functions" to capture this feature, and we adopt (a translation of) their model. People's instantaneous utility for a given period represents how much pleasure they experience that period. Suppose that all effects of past consumption on period-t instantaneous utility can be captured in a single summary statistic, which we denote by k_t. We often refer to k_t as people's *addiction level* in period t. People's instantaneous utility in period t is given by $u_t(a_t, k_t)$—that is, how much pleasure they experience in period t depends both on whether they hit and on their addiction level.

In general, people's addiction level will be a function of their past consumption. Gary S. Becker and Kevin M. Murphy (1988) assume $k_t = \gamma k_{t-1} + a_{t-1}$ for some $\gamma \in [0, 1]$. For simplicity, we limit attention here to the case $\gamma = 0$, which implies that $k_t = a_{t-1}$. If people hit last period (that is, $a_{t-1} = 1$), then they are hooked this period (that is, $k_t = 1$); and if people refrained last period (that is, $a_{t-1} = 0$), then they are unhooked this period (that is, $k_t = 0$). Limiting attention to the case $\gamma = 0$ is of course unrealistic. Assuming $\gamma = 0$ implies that there are only two addiction levels, being hooked and being unhooked. Moreover, it implies that a single period of restraint gets people completely unhooked and that a single period of indulgence gets people completely hooked. These assumptions make sense only if periods are somewhat lengthy. Even so, it turns out that our main results and intuitions will hold for any $\gamma \in [0, 1]$. Hence, although the reader should not take our model too literally, we believe the model does reveal some more general insights.

Suppose the period-t instantaneous utility function takes the form shown in table 6.1. We often drop the subscript t from k_t and a_t when there is no danger of confusion. An important concept for the analysis will be the *current temptation to consume* the addictive product, by which we mean the instantaneous utility from taking a hit relative to that from not taking a hit. With the formulation in table 6.1, the temptation to consume in period t given addiction level k is $f_t(k) - g_t(k)$. Of course, the decision whether to hit relies on more than merely the current temptation to consume, since people care about how current consumption affects future instantaneous utilities. This trade-off between the current temptation to consume and the future costs of such consumption is the crux of the choice to become addicted.

We consider two characteristics of addictive products. The first is that they can be habit forming: The more people have consumed in the

Table 6.1 Instantaneous Utility Function I

Condition	Utility from Hitting: $u_t(1, k_t)$	Utility from Refraining: $u_t(0, k_t)$
When unhooked ($k_t = 0$)	$f_t(0)$	$g_t(0)$
When hooked ($k_t = 1$)	$f_t(1)$	$g_t(1)$

past, the larger is their current temptation to consume; for example, smoking cigarettes at age sixteen increases the temptation to smoke a cigarette at age seventeen. Formally:

DEFINITION 1 A product is *habit forming* if for all t, $f_t(1) - g_t(1) > f_t(0) - g_t(0)$.

In addition to being habit forming, addictive activities often generate negative internalities: The more people have consumed in the past, the smaller is their current well-being (no matter their current behavior); for example, smoking cigarettes at age sixteen reduces pleasure at age seventeen both from smoking and from not smoking.[5] Formally:

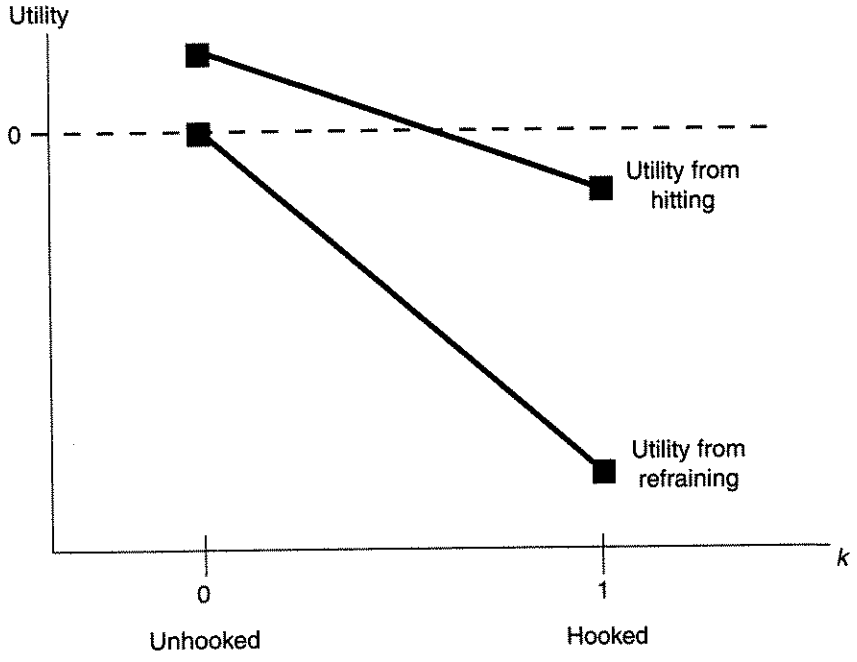
DEFINITION 2 A product has *negative internalities* if for all t, $f_t(1) < f_t(0)$ and $g_t(1) < g_t(0)$.

Negative internalities include health problems due to overeating or oversmoking, as well as the "tolerance" that is exhibited for many drugs.[6] Of course, activities can generate negative internalities without being habit forming (for example, eating cheesecake); and a habit-forming activity need not generate negative internalities (for example, jogging). "Addictive products" are usually considered both to be habit forming and to generate negative internalities, and that is the case we study in this chapter. Figure 6.1 illustrates what the instantaneous utility function might look like for such a good.

Our formulation allows for instantaneous utilities to vary across time. For simplicity, we assume that any nonstationarities arise from variations in the utility from hitting—that is, we assume $g_t(k)$ is independent of t. We can then without loss of generality express the period-t instantaneous utility function in terms of three parameters, as shown in table 6.2.

The formulation in table 6.2 normalizes the instantaneous utility from refraining when unhooked to be zero. Then f_t represents the temptation to hit when unhooked, ρ represents the magnitude of the negative internality, and $\sigma - \rho$ represents the magnitude of the habit formation. Any nonstationarities in the instantaneous utility function are captured by a

Figure 6.1 Instantaneous Utility



varying f_t. We should consider both the case of a stationary instantaneous utility function—so $f_t = f_0$ for all t—and the case of a nonstationary instantaneous utility function.

Throughout our analysis, we assume that people correctly predict how current consumption affects future instantaneous utility functions. Our analysis therefore ignores the possibility that people simply underestimate the addictive nature of products they consume. For the instantaneous utility function in table 6.2, this would mean that people underestimate ρ or σ. Although we suspect that this possibility might be quite important for addictive behavior—plausibly more important than self-control problems—our goal in this chapter is to study the implications of self-control problems alone.

Although the previous discussion characterizes instantaneous utilities for addictive products, in any given period people care not only about their current instantaneous utility but also about their future instantaneous utilities. This is captured by people's intertemporal preferences.

Table 6.2 Instantaneous Utility Function II

Condition	Utility from Hitting: $u_t(1, k)$	Utility from Refraining: $u_t(0, k)$
When unhooked ($k = 0$)	f_t	0
When hooked ($k = 1$)	$f_t - \rho$	$-\sigma$

Evidence suggests that people have self-control problems: People tend to pursue immediate gratification in a way that they do not appreciate from a long-run perspective. For example, suppose people are presented with a choice between doing seven hours of an unpleasant task on April 1 versus eight hours on April 15. We suspect that if asked on February 1 (that is, from a long-run perspective), virtually everyone would prefer the seven hours on April 1. Yet if given the same choice on April 1, most people would choose to put off the work until April 15.[7]

The standard economics model, in contrast, assumes that intertemporal preferences are *time consistent:* People's relative preference for well-being at an earlier date over a later date is the same no matter when they are asked. In the example above, such time consistency would require that, irrespective of the specific choice, people make the same choice on February 1 and April 1. The standard economics model therefore, a priori, rules out self-control problems.

A small set of economists and psychologists has over the years proposed formal models of time-inconsistent preferences and self-control problems.[8] Edmund S. Phelps and Robert A. Pollak (1968) put forward an elegant model of intertemporal preferences in the context of intergenerational altruism, which David Laibson (1994a) later used to capture self-control problems within individuals.[9] If u_τ is the instantaneous utility people get in period τ, then their intertemporal preferences at time t, U^t, can be represented by the following utility function:

For all t,

$$U^t(u_t, u_{t+1}, \cdots, u_T) \equiv \delta^t u_t + \beta \sum_{\tau=t+1}^{T} \delta^\tau u_\tau. \tag{6.1}$$

By assuming that both β and δ are greater than zero but no greater than one, these intertemporal preferences capture the idea that at each moment people care about their future well-being but typically less than they care about their current well-being. For $\beta = 1$, these preferences are time consistent, wherein the parameter δ represents "time-consistent" impatience. For $\beta < 1$, however, these preferences are time inconsistent, wherein the parameter β parsimoniously captures the degree to which people pursue immediate gratification: While β plays

no role in determining people's willingness to trade off well-being among future periods, it determines how much more they care about their current well-being than their well-being in all future periods.

When people have self-control problems, an important issue arises: Are they aware of these self-control problems? Our analysis considers two extreme assumptions: *Sophisticated* people are fully aware of their future self-control problems and therefore know exactly how they will behave in the future; and *naive* people are fully *un*aware of their future self-control problems and therefore believe they will behave in the future exactly as they currently would like to behave in the future.[10] Since we wish to compare people with self-control problems to people without self-control problems, our analysis also examines time-consistent agents, whom we refer to as *TC*s.

To formalize our predictions about how the three types behave, we assume people follow "perception-perfect strategies," which in this environment implies that people choose to hit today if and only if hitting today is optimal given their current preferences and their current beliefs about how they will behave in the future.[11]

To capture people's beliefs about how they will behave in the future, we define a *strategy* α to be a function that specifies what people would do in all situations. In other words, for all k and t, $\alpha(k, t)$ is the action people would pursue in period t when their addiction level is k. For example, if $\alpha(0, t) = 0$ and $\alpha(1, t) = 1$, then people would refrain in period t if unhooked, and people would hit in period t if hooked.

Let $U_t(k_t, \alpha)$ be people's period-t continuation (long-run) utility as a function of their addiction level in period t, k_t, and their strategy, α. Long-run utility represents intertemporal preferences from some prior perspective, so that self-control problems (that is, β) are irrelevant. People's long-run preferences are represented by equation (6.1) when $\beta = 1$, and therefore *TC*s, naifs, and sophisticates have identical long-run utilities. A useful way to write $U_t(k, \alpha)$ is

$$U_t(k, \alpha) = \begin{cases} f_t(k) + \delta U_{t+1}(1, \alpha), \text{ if } \alpha(k, t) = 1 \\ g_t(k) + \delta U_{t+1}(0, \alpha), \text{ if } \alpha(k, t) = 0. \end{cases}$$

Consider people in period t who are contemplating the consequences of their current behavior on their future intertemporal utility. Suppose they perceive that they will follow strategy α^p beginning in period $t + 1$, in which case they believe that if they hit this period then their intertemporal utility beginning next period will be $U_{t+1}(1, \alpha^p)$, and they believe that if they refrain this period then their intertemporal utility beginning

next period will be $U_{t+1}(0, \alpha^p)$. Hence, they perceive the (undiscounted) benefit of restraint to be $U_{t+1}(0, \alpha^p) - U_{t+1}(1, \alpha^p)$.

We now have a formalization of the choice of whether to hit today: People hit in period t if and only if, given their perceptions of future behavior α^p, the current temptation to hit $f_t(k) - g_t(k)$ is larger than the (discounted) future benefit from current restraint $U_{t+1}(0, \alpha^p) - U_{t+1}(1, \alpha^p)$. For simplicity, we assume people hit when indifferent. Notice that (given our very special assumptions) the benefit from restraint is independent of whether people are currently hooked, whereas the temptation to hit is higher if people are currently hooked. This means that for all three types, in any period people hit when unhooked only if they also hit when hooked.

*T*Cs are time consistent, so for each (k, t) their continuation strategy maximizes their continuation utility. The implication of time consistency in the framework discussed in the preceding paragraph is that *T*Cs correctly perceive their future behavior and that they discount the future benefit from current restraint by δ. Hence, we define perception-perfect strategies for *T*Cs as:

DEFINITION 3 *A perception-perfect strategy for T*Cs is a strategy α^{tc} that satisfies for all $k \geq 0$ and for all t, $\alpha^{tc}(k, t) = 1$ if and only if $f_t(k) - g_t(k) \geq \delta(U_{t+1}(0, \alpha^{tc}) - U_{t+1}(1, \alpha^{tc}))$.

At any point in time, naifs believe they will behave like *T*Cs beginning with the next period. Hence, in any period, naifs perceive that they will follow strategy α^{tc} beginning with the next period. Since naifs discount the future benefit of current restraint by $\beta\delta$, we define perception-perfect strategies for naifs as:

DEFINITION 4 A *perception-perfect strategy for naifs* is a strategy α^n that satisfies for all $k \geq 0$ and for all t, $\alpha^n(k, t) = 1$ if and only if $f_t(k) - g_t(k) \geq \beta\delta\ (U_{t+1}(0, \alpha^{tc}) - U_{t+1}(1, \alpha^{tc}))$.

Sophisticates, like *T*Cs, predict exactly how they will behave in the future. Sophisticates, like naifs, also discount the future benefit of current restraint by $\beta\delta$. Hence, we define perception-perfect strategies for sophisticates as

DEFINITION 5 A *perception-perfect strategy for sophisticates* is a strategy α^s that satisfies for all $k \geq 0$ and for all t, $\alpha^s(k, t) = 1$ if and only if $f_t(k) - g_t(k) \geq \beta\delta(U_{t+1}(0, \alpha^s) - U_{t+1}(1, \alpha^s))$.

In each period, *T*Cs and naifs are really just choosing an optimal future consumption path. *T*Cs will always stick to the behavior path

chosen in the first period. Naifs, in contrast, will often revise their chosen behavior paths as their preferences change from period to period. Sophisticates are in a sense playing a game against their future selves. Hence, their behavior will partly reflect "strategic" reactions to bad behavior by future selves that they cannot directly control and partly reflect attempts to induce good behavior from future selves.

Stationary Preferences

In this section, we analyze a stationary model of addiction:

(A1) Assume that $f_t = f_o$ for all t.

Assumption A1 says that the instantaneous utility function $u_t(a, k)$ depends on the current level of addiction k but not on the specific period t. As we shall see, this assumption is rather important. In many cases it is quite unrealistic: It assumes, for instance, that the first hit of a cigarette or cocaine yields the same pleasure to a twenty-year-old as it does to a sixty-year-old. Nonetheless, as a base case and to clarify certain issues, we maintain this assumption for this section.

We begin with a three-period example that provides some intuition and also illustrates how to solve for the perception-perfect strategies for *TC*s, naifs, and sophisticates. Suppose people live for three periods, which we interpret as youth, middle age, and old age. In any given period, people are currently hooked if $k = 1$ (that is, because they hit last period) and unhooked if $k = 0$ (that is, because they refrained last period). Finally, suppose that people's preferences in each of the three periods can be represented with the following instantaneous utilities:

EXAMPLE 1: Suppose $f_o = 10$, $\rho = 18$, and $\sigma = 25$.

Table 6.3 displays example 1. Consider how *TC*s with $\delta = 1$ would behave. *TC*s hit no matter what in their old age, since the instantaneous utility from hitting is larger than the instantaneous utility from refraining whether hooked or unhooked. In their middle age, *TC*s decide whether to hit knowing they will hit no matter what in their old age. It is straightforward to show that they refrain no matter what in their middle age; for example, when hooked in middle age, refraining yields intertemporal utility $(-25) + 10 = -15$, while hitting yields utility $(-8) + (-8) = -16$. In their youth, *TC*s know they will refrain in their middle age and hit in their old age no matter what they do now, and they prefer to refrain (because refraining yields $0 + 0 + 10 = 10$ while hitting yields $10 + (-25) + 10 = -5$). Hence, *TC*s with $\delta = 1$ refrain in their youth and middle age but then hit in their old age.

Table 6.3 Example 1

Condition	Utility from Hitting: $u(1, k)$	Utility from Refraining: $u(0, k)$
When unhooked ($k = 0$)	10	0
When hooked ($k = 1$)	−8	−25

Consider next naifs with $\delta = 1$ and $\beta = ½$. Naifs always believe they will behave like *TC*s in the future, and therefore in their youth naifs believe they will refrain in middle age and hit in old age no matter what they do now. Although having a self-control problem creates an increased desire to hit for naifs, with $\beta = ½$ naifs manage to refrain while young, because they perceive that refraining yields $0 + (½)\,0 + (½)\,10 = 5$ while hitting yields $10 + (½)\,(-25) + (½)\,10 = 2.5$. In their middle age, naifs are aware that they will hit no matter what in their old age. Now the self-control problem leads naifs to hit no matter what: Even when unhooked, hitting yields $10 + (½)\,(-8) = 6$ while refraining yields $0 + (½)10 = 5$. Finally, in their old age, naifs, like *TC*s, hit no matter what. Hence, naifs refrain in their youth but hit in both their middle age and old age.

In this example, naifs indulge in the addictive activity more than *TC*s. This result turns out to be quite general: Self-control problems combined with a belief that in the future they will not have such problems always leads people to overconsume addictive products. Indeed, the following result follows directly from definitions 3 and 4: For any contingency, if *TC*s hit, then naifs hit, and therefore if naifs refrain, then *TC*s refrain.

LEMMA 1. For any k and t, if $\alpha^{tc}(k, t) = 1$ then $\alpha^{n}(k, t) = 1$.

Now consider sophisticates with $\delta = 1$ and $\beta = ½$. In their middle age, sophisticates correctly perceive that, like *TC*s, they will hit no matter what in their old age. Given this belief, it is in fact optimal to hit no matter what in their middle age (the comparison is identical to that for naifs). In their youth, sophisticates realize that they will hit for the rest of their lives no matter what they do now. As a result, it is optimal to hit during their youth as well, because hitting yields $10 + (½)\,(-8) + (½)\,(-8) = 2$, while refraining yields $0 + (½)\,10 + (½)\,(-8) = 1$. Hence, sophisticates hit throughout their lives.

In this example, sophisticates indulge in the addictive activity more than naifs. Although this result may seem surprising, it reflects how

sophisticates' correct pessimism about future behavior can lead to increased consumption in the realm of addiction. In their youth, sophisticates know they will hit no matter what during middle age, whereas naifs optimistically *and incorrectly* believe they will surely refrain during middle age. The habit-forming property of addictive goods implies, however, that the more people expect to hit in the future, the smaller is the future benefit of refraining now. As a result, having (correctly) pessimistic beliefs about future behavior can make sophisticates more likely to indulge than naifs.

Example 1 illustrates some basic intuitions of the stationary model. We now show that these intuitions hold more generally. To do so, we focus on the case where there is an infinite horizon ($T = \infty$). We do so for two reasons. First, it is expositionally easier to describe the results for an infinite horizon. Second, this assumption is closer in spirit to the rational choice models of addiction and yields more realistic results.

In an infinite-horizon model with stationary instantaneous utilities, *T*Cs and naifs both follow a stationary strategy, wherein behavior depends only on the current addiction level k and not the specific period t. In any period, both *T*Cs and naifs choose today's behavior by determining their optimal lifetime path of behavior beginning from today. Given an infinite horizon, stationary instantaneous utilities, and our assumption that people hit when indifferent, for any t there is a unique optimal lifetime path of behavior, and this path depends on the current addiction level k but not the current period t. This logic is summarized in the following lemma:

> LEMMA 2. Under stationary instantaneous utilities and $T = \infty$, (1) there is a unique perception-perfect strategy for *T*Cs, α^{tc}, and this strategy is stationary; and (2) there is a unique perception-perfect strategy for naifs, α^{n}, and this strategy is stationary.

Since there are only two addiction levels (that is, people can be hooked or unhooked), and since people would never hit when unhooked but refrain when hooked, there are three relevant stationary strategies that *T*Cs and naifs might follow: They might hit no matter what; they might refrain no matter what; or they might refrain when unhooked but hit when hooked.

For sophisticates, there can be multiple perception-perfect strategies when there is an infinite horizon. However, there is a unique perception-perfect strategy for sophisticates when there is a finite horizon (given the assumption of hitting when indifferent). Throughout this chapter, we focus on perception-perfect strategies for an infinite

horizon that correspond to the unique finite-horizon perception-perfect strategy as the horizon becomes long.[12] This restriction rules out a perpetual one-shot-is-all-I-get mentality, wherein people think to themselves, "If I can just refrain today then I'll refrain always, whereas if I hit today I'll hit forever after." More precisely, we rule out this mentality when it can be supported *only* by infinite-horizon reasoning (analogous to folk-theorem-type equilibria in infinitely repeated games), because a variant of such a mentality can arise in a stationary finite-horizon model.

When there is a long, finite horizon, the crucial question that determines the behavior of sophisticates is whether they would hit when unhooked in the second-to-last period while knowing that they would hit no matter what in the last period. If the answer is yes, then they will hit no matter what in the second-to-last period, and they face the same decision in the third-to-last period. As a result, everything unravels, and they hit no matter what in all periods. Suppose the answer is no, so that they refrain when unhooked in the second-to-last period. Since the benefit from restraint cannot be smaller than when they hit for sure next period, in this case sophisticates must always refrain when unhooked; that is, α^s satisfies $\alpha^s(0, t) = 0$ for all t. In this case, the behavior of sophisticates when hooked is unclear—they might hit when hooked in all periods, they might refrain when hooked in all periods, or they might hit when hooked every τ periods for some $\tau > 1$ (in which case, α^s is nonstationary). We summarize this logic in the following lemma:

LEMMA 3. Under stationary instantaneous utilities and $T = \infty$, α^s satisfies either (1) $\alpha^s(k,t) = 1$ for all k and t, or (2) $\alpha^s(0,t) = 0$ for all t.

Now consider observed behavior when people are initially unhooked (that is, $k_1 = 0$). Lemma 2 implies that both *T*Cs and naifs either always hit or never hit, and lemma 3 implies that sophisticates also either always hit or never hit. To compare the three types, we must determine when each type always hits. TCs are time consistent, and therefore they always hit if and only if they prefer always hitting to never hitting. Since (it can be shown by calculating some infinite sums) always hitting yields intertemporal utility $f_0/(1 - \delta) - \delta\rho/(1 - \delta)$, and never hitting yields intertemporal utility 0, *T*Cs always hit if and only if $f_0 \geq \delta\rho$. For naifs, we must determine beliefs about future behavior. If $f_0 + \sigma - \rho < \delta\sigma$, then *T*Cs would refrain forever even if they were currently hooked. Naifs who are unhooked therefore consider taking a single hit, thinking they will never hit again. The single hit is worthwhile if and only if $f_0 \geq \beta\delta\sigma$, in which case naifs always hit. If $f_0 + \sigma - \rho \geq \delta\sigma$, then *T*Cs would hit forever if they

were currently hooked, and hence naifs who are unhooked believe (correctly) that they are choosing between never hitting and always hitting. In this case, naifs always hit if and only if $f_o \geq \beta\delta\rho / (1 - \delta + \beta\delta)$. Finally, sophisticates always hit when initially unhooked if and only if they prefer hitting today given that they will hit for sure tomorrow. Hence, sophisticates always hit if and only if $f_o \geq \beta\delta\rho$. Summarizing,

*T*Cs always hit if and only if

$$f_o \geq \delta\rho.$$

Naifs always hit if and only if

$$f_o \geq \beta\delta\sigma, \text{ when } f_o < \rho - (1-\delta)\sigma.$$

$$f_o \geq \beta\delta\,\rho/(1-\delta+\beta\delta), \text{ when } f_o \geq \rho - (1-\delta)\sigma.$$

Sophisticates always hit if and only if

$$f_o \geq \beta\delta\rho.$$

Given $\beta < 1, f_o > 0$, and $\sigma > \rho > 0$, the following proposition derives from the above equations:

PROPOSITION 1. Under stationary instantaneous utilities and $T = \infty$, if $k_1 = 0$ (that is, people are initially unhooked): (1) if TCs always hit, then naifs always hit; and (2) if naifs always hit, then sophisticates always hit.

Part 1 of proposition 1 merely restates lemma 1: Naifs are always more likely to hit than *T*Cs. Part 2 of proposition 1 establishes that the surprising outcome of example 1—that sophisticates consume more of the addictive product than naifs—always holds in a stationary model when people are initially unhooked.

The result that sophisticates are more likely to hit than naifs, however, very much relies on people being initially unhooked. To illustrate, consider behavior in example 1 when $\beta = 2/3$ and $k_1 = 1$. With $\beta = 2/3$, both sophisticates and naifs hit in middle age if they are hooked but refrain in middle age if they are unhooked. Sophisticates correctly predict this behavior and, as a result, find it optimal to refrain while young even with $k_1 = 1$ in order to induce good behavior in middle age. Naifs, in contrast, believe they will refrain no matter what in middle age and therefore choose to hit while young for $k_1 = 1$. Hence, for $\beta = 2/3$ and $k_1 = 1$, sophisticates refrain in both youth and middle age, whereas naifs hit throughout their lives. (Proposition 1 is not violated, since for $\beta = 2/3$ and $k_1 = 0$, both sophisticates and naifs refrain in youth and in middle age.)

Hence, the example illustrates that, when initially hooked, sophisticates can be more prone to quit than naifs. In fact, this result holds more generally:

PROPOSITION 2. Under stationary instantaneous utilities and T-∞, if $k_1 = 1$ (that is, people are initially hooked): (1) If *TC*s always hit, then sophisticates always hit; (2) if sophisticates always hit, then naifs always hit.

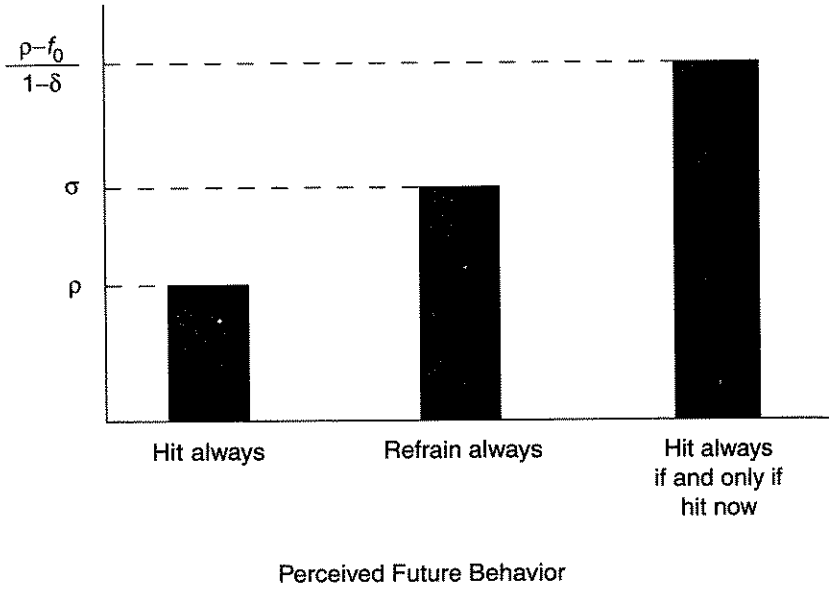
The different results in propositions 1 and 2 highlight the complex role of awareness about future self-control problems. In fact, there are two ways in which sophistication about future self-control problems can influence people's behavior. First, sophistication about future self-control problems can make people pessimistic about future behavior (that is, they believe that in general that they will hit more often than they would if they had no self-control problem). We refer to this phenomenon as the *pessimism effect*. Second, sophistication about future self-control problems may make people realize that they will resist future temptations only if they resist current temptation. We refer to this phenomenon as the *incentive effect*.

With some oversimplification, figure 6.2 illustrates the distinction between the pessimism effect and the incentive effect. Figure 6.2 shows the future benefit of current restraint as a function of three possible beliefs about future behavior: People might believe they will always hit in the future no matter what they do now; people might believe they will refrain always in the future no matter what they do now; and people might believe they will always hit in the future if they hit now but refrain always in the future if they refrain now. The figure assumes parameters such that *TC*s refrain in all contingencies.[13] This implies that naifs and *TC*s both perceive that in the future they will refrain no matter what, and therefore according to figure 6.2 the future benefit from current restraint is σ.

Pure pessimism reflects that while *TC*s and naifs perceive that they will refrain no matter what in the future, sophisticates may perceive that they will hit no matter what in the future, in which case, according to figure 6.2, the benefit from current restraint is ρ. Hence, pure pessimism about future behavior implies that sophisticates are more likely to hit than *TC*s or naifs (because the perceived benefit from restraint is smaller). Figure 6.2 makes clear that this result is driven by the habit-forming property of addictive products (that is, $\sigma > \rho$ is exactly equivalent to the product being habit forming). When a product is habit forming, the more often people will hit in the future, the less costly is hitting now.

Sophisticates may not be purely pessimistic; rather, they might be pessimistic about their future behavior when hooked but optimistic

Figure 6.2 Future Benefit of Current Restraint



about their future behavior when unhooked, in which case they perceive a need to refrain now in order to induce good behavior in the future. This is when the incentive effect is operative, and according to figure 6.2, in this case the benefit from current restraint is $(\rho - f_o)^0 / (1 - \delta) > \sigma$.[14] Hence, the incentive effect can imply that sophisticates are less likely to hit than *T*Cs or naifs. This result is driven by sophisticates' concern about *improper* future overconsumption (a concern that neither *T*Cs nor naifs would ever have). That is, sophisticates refrain when naifs or *T*Cs do not only if sophisticates are refraining to prevent improper future behavior.

The crucial question then is when does the incentive effect become operative; and since the incentive effect is driven by future restraint, the answer is, only if there is some future period where sophisticates will refrain when unhooked *in the absence of the incentive effect*.[15] It is this intuition that drives the different results in propositions 1 and 2 (and that we build on in our discussion of nonstationary preferences and variable myopia). In the stationary model, if in period 1 people would hit when unhooked in the face of pure pessimism, then in all periods they would hit when unhooked in the face of pure pessimism, and therefore the incentive effect cannot be operative. This means that whenever sophis-

ticates *need* the incentive effect to refrain when unhooked, it is inoperative. It follows that sophisticates are more likely than naifs to hit when unhooked (proposition 1). When people are initially hooked, in contrast, the incentive effect can be operative even when they would hit in the face of pure pessimism because refraining now will get them unhooked. It turns out that naifs refrain when hooked only if sophisticates refrain when unhooked, but then the incentive effect is operative and sophisticates are more likely to refrain when hooked than naifs (proposition 2).

Throughout this section, we explicitly and implicitly state that both sophisticates and naifs are "hurting" themselves with their behavior. Indeed, this notion can be formalized. Of course, in an environment in which people have different preferences at different times, we must specify what we are using for a welfare criterion. A conservative approach is to assume there are no true preferences and to consider Pareto comparisons (see, for example, Goldman 1979, 1980, and Laibson 1994a). Alternatively, Ted O'Donoghue and Matthew Rabin (1999) employ a less conservative approach, deeming the long-run preferences (that ignore any taste for immediate gratification) to be the true preferences, relevant for welfare analysis. In the examples of this section and throughout the rest of the chapter, however, sophisticates and naifs can hurt themselves by any criterion. Intuitively, if people get *inappropriately* addicted to a product, they are generating dissatisfaction in almost every period of their lives, and hence from all points of view addiction is perceived as undesirable. We return to this issue at the end of this chapter.

Nonstationary Preferences

Although the stationary model provides insight into how self-control problems and awareness of future self-control problems might affect addictive behavior, some of the results depend on the unrealistic assumption that the instantaneous utility function is constant over time. This assumption rules out the possibility that the desire to consume addictive products decreases as people get older. It also rules out the possibility of day-to-day fluctuations in the desire to consume addictive products—for example, the desire to consume may be greater on weekends than it is on weekdays, or the desire may be greater in response to certain traumatic events (as when abstaining alcoholics resume drinking during a crisis). In this section, we consider these possibilities in order to get a more complete picture of how self-control problems affect addictive behaviors.

As discussed earlier in the chapter, we model nonstationary instantaneous utilities by introducing variations in the utility from hitting.

Table 6.4. Example 2

Condition	Utility from Hitting: $u_t(1, k)$	Utility from Refraining: $u_t(0, k)$
In youth when unhooked	14	0
In youth when hooked	–4	–25
In middle age when unhooked	10	0
In middle age when hooked	–8	–25
In old age when unhooked	–5	0
In old age when hooked	–23	–25

In other words, we assume there is a sequence $(f_1, f_2, \ldots, f_T)$ such that $f_t(k) = f_t - \rho k$, and we assume that $g_t(k) = -\sigma k$ for all t. The stationary case assumes $f_t = f_o$ for all t. In this section, we consider various ways in which f_t may depend on t.

For many addictive products, the temptation to consume declines over the course of one's life. For example, if a twenty-year-old and a sixty-year-old have both never taken cocaine, it seems likely that the temptation to take a first hit is larger for the twenty-year-old. This difference might arise from forces such as peer pressure, or the young body's physical resilience, or merely the fact that an older person tends to lose interest in novel activities. Consider the following model of addiction:

(A2) Assume that $f_1 \geq f_2 \geq \ldots \geq f_T$.

To illustrate how this new assumption can change the results, consider the following variant of example 1:

EXAMPLE 2: Suppose $(f_1, f_2, f_3) = (14, 10, -5)$, $\rho = 18$, and $\sigma = 25$.

Table 6.4 illustrates example 2. Examples 1 and 2 have the same levels of habit formation and negative internalities—that is, ρ and σ are the same in the two examples. Moreover, examples 1 and 2 have identical instantaneous utilities for middle age. In example 2, however, people have a larger temptation to hit in their youth and a smaller temptation to hit in their old age. Indeed, the crucial feature of example 2 is that people hit in their old age if and only if they are hooked (in example 1, people hit in their old age no matter what).

In example 2, it is straightforward to show that *TC*s with $\delta = 1$ refrain throughout their lives; sophisticates with $\delta = 1$ and $\beta = \frac{1}{2}$ also refrain throughout their lives; and naifs with $\delta = 1$ and $\beta = \frac{1}{2}$ hit throughout

their lives. (The calculations are left to the reader.) Of particular interest is how modifying example 1 so as to incorporate a decreasing temptation over one's lifetime affects sophisticates and naifs in opposite directions. Sophisticates indulge less in example 2 than in example 1—in example 2 they never hit, whereas in example 1 they always hit—and naifs indulge more in example 2 than in example 1—in example 2 they always hit, whereas in example 1 they refrained in their youth.

That sophisticates indulge less in example 2 than in example 1 reflects the increased power of the incentive effect in the youth environment. In example 2, sophisticates hit in old age if and only if they are hooked. Knowing this, they hit in middle age if and only if they are hooked. In their youth, sophisticates correctly recognize that hitting now means also hitting in both middle age and old age, whereas refraining now means also refraining in both middle age and old age. Since even in their youth they perceive always hitting to be *worse* than always refraining, sophisticates choose to refrain in their youth. In their youth sophisticates would most like to hit now and refrain thereafter; but they choose to refrain in their youth in order to induce good behavior in the future (that is, because of the incentive effect).

That naifs indulge more in example 2 than in example 1 reflects how the youth environment can be problematic for naifs, who give in to large youthful temptations under the false belief that they will later quit. In example 2, in their youth, naifs (like sophisticates) would most like to hit in their youth and refrain thereafter. Since naifs do not foresee future self-control problems, they choose to follow this path in their youth; but they end up never quitting and, therefore, suffer a lifetime of addiction.

We now describe behavior more generally in the youth environment. The interesting case is that in which people have an increased temptation to hit while young but eventually the temptation falls to a more normal level. We refer to this phenomenon as people maturing:

DEFINITION 6 Suppose there exists some $\tau \geq 2$ such that $f_1 \geq f_2 \geq \ldots \geq f_\tau = f_{\tau+1} = \ldots = f_T$. Then we say people become mature in period τ.

With this definition in hand, we can state a general proposition regarding youth models:

PROPOSITION 3. Suppose that once a person becomes mature, she will refrain even in the face of pure pessimism. Then (1) in all situations, sophisticates hit only if naifs hit; and (2) sophisticates always hit only if they prefer (from a period-1 perspective) to always hit rather than never to hit.

The crucial condition in proposition 3 is that *there is eventually some period* in which people will refrain when unhooked even in the face of pure pessimism. We feel that this is a realistic condition for many addictive products—eventually people will lose interest in the product as long as they are unhooked at that time. The results in proposition 3 reflect that this condition is exactly the condition for when the incentive effect is operative in the youth environment. Part 1 states that in this case sophisticates are less likely to hit than naifs in all situations. Part 2 states that in this case sophisticates cannot suffer a *costly* lifelong addiction, because they choose to hit throughout their lives only if that is optimal from a period-1 perspective.

These results stand in stark contrast to the results in the stationary model. In the stationary model, the incentive effect is operative if and only if *in the first period* people would refrain when unhooked in the face of pure pessimism, and as a result sophisticates can suffer a very harmful lifelong addiction because of a feeling that addiction is inevitable. In the youth model, in contrast, as long as the temptation to consume eventually falls to the point at which people would choose to refrain even in the face of pure pessimism, the inevitability of addiction vanishes, and as a result sophisticates are less likely to hit than naifs and unlikely to suffer harmful lifelong addictions.

Although proposition 3 suggests that sophisticates will not suffer a lifelong addiction when doing so is particularly costly, sophisticates may engage in costly misbehavior in their youth (provided they will indeed quit once mature). For example, suppose people are sure they will quit drinking as soon as they graduate from college (that is, when they become mature). Knowing this, they may drink no matter what in the last semester at college, which can lead them to drink no matter what in the second-to-last semester of college, and so on. As a result, they may start drinking in the first semester of college knowing full well that they will drink throughout college and then quit, even though from the perspective of the first semester of college they would prefer not to drink at all in college. Two comments about such youthful misbehavior are in order. First, it can clearly be quite costly if maturity comes late in life. Even so, we feel that for many addictive products maturity does set in at a reasonable age. Second, whether such youthful misbehavior occurs depends critically on whether people would quit if hooked once mature. If not, then misbehavior during youth is quite dangerous and therefore unlikely. If so (for example, the college example above), then misbehavior during youth is quite safe and therefore likely.[16]

Finally, we note that although we have no formal results concerning the behavior of naifs, there is reason to believe that naifs are likely to do quite poorly in the youth environment. Recall that in the stationary model

Table 6.5 Example 3

Condition	Utility from Hitting	Utility from Refraining
On weekend when unhooked	13	0
On weekend when hooked	–2	–18
On weekday when unhooked	10	0
On weekday when hooked	–5	–18

naifs have a tendency not to quit once hooked *even when it is well worth their while.* In the youth environment this becomes a real problem whenever the optimal plan is to hit in one's youth and later quit. Indeed, even in cases in which naifs would refrain forever after reaching maturity unhooked, naifs may form a very harmful lifelong addiction. For example, naifs may indulge in some addictive activity every week during college, planning to quit as soon as they graduate, and then indulge every week after graduation for the rest of their lives, each week planning to quit *next* week.

In addition to generally declining over one's lifetime, the temptation to engage in addictive activities may also fluctuate from day to day. The temptation to consume alcohol, for instance, may be larger on weekends than it is on weekdays. Consider the following model of addiction:

(A3) Assume that $f_t = f_o + X$ for $t \in \{1, 3, 5, \ldots\}$ and that $f_t = f_o$ for $t \in \{2, 4, 6, \ldots\}$.

This assumption says that each period is either a weekend (odd-numbered periods) or a weekday (even-numbered periods), and the temptation to hit is larger on weekends. Consider the following example:

EXAMPLE 3: Suppose $f_o = 10$, $\rho = 15$, $\sigma = 18$, and $X = 3$.

Table 6.5 illustrates example 3. Given the instantaneous utilities in example 3, *TC*s never hit when $T = \infty$ and $\delta = .99$.[17] For δ close to one, *TC*s choose the behavior that maximizes their long-run per-week payoff. If *TC*s choose to always hit, then they have payoff –2 on weekends and –5 on weekdays for a per-week payoff of –7. If they choose to never hit, their per-week payoff is zero. If they choose to hit on weekends and refrain on weekdays (that is, to consume in moderation), then they have payoff 13 on weekends and –18 on weekdays for a per-week payoff of –5. Hence, *TC*s choose to never hit.

Next consider how naifs and sophisticates behave, now assuming that $\beta = .7$. Naifs choose to hit in all periods. For naifs (and sophisticates), on any specific weekend the optimal lifetime plan of behavior is to hit today

and never again, regardless of whether they are currently hooked. On any specific weekday, the optimal lifetime plan is to hit today and never again if they are currently hooked and never to hit if they are currently unhooked. Naifs therefore hit on weekends whether or not they are hooked and hit on weekdays when they are hooked. As a result, naifs hit in all periods. Sophisticates, in contrast, only hit every other weekend (that is, sophisticates follow the behavior path: Hit, refrain, refrain, refrain, hit, refrain, refrain, refrain, hit, . . .). In other words, sophisticates consume the addictive good in much smaller amounts than naifs. This example is precisely the type of situation where the incentive effect helps out sophisticates. Naifs hit on the first weekend planning to get unhooked during the upcoming weekday, but once hooked they are not able to resist even the weekday temptation. Sophisticates realize that for certain weekdays they will be able to control themselves only if they are unhooked, and thus they have an extra incentive to refrain even in the face of a larger temptation the preceding weekend.

To understand the specific cycle that sophisticates follow, we must ask when the incentive effect will be particularly strong. Suppose there is some weekend when sophisticates hit whether or not they are hooked.[18] On the preceding weekday, there is no incentive effect. Even so, given the smaller weekday temptation, sophisticates hit if and only if they are hooked. The incentive effect is therefore operative on the preceding weekend, and as a result they are able to resist the higher weekend temptation when unhooked. Since they hit when hooked on that weekend, the incentive effect is even stronger on the preceding weekday: Restraint induces further restraint in each of the next two periods, whereas hitting induces further hits in each of the next two periods. For this particular example, the incentive effect is now strong enough that sophisticates refrain on that weekday whether or not they are hooked. However, this means there is no incentive effect to overcome the larger temptation on the preceding weekend, so sophisticates hit on that weekend whether or not they are hooked, restarting the cycle.

Example 3 and other similar examples further highlight our main theme in this section: Restricting attention to stationary instantaneous utilities is very misleading, because it ignores a number of realistic situations wherein sophistication is likely to help people with self-control problems and wherein naivete can really hurt people with self-control problems. Indeed, in a more general model of periodically changing utilities we hypothesize that sophisticates may consume even less than *TC*s. (In example 3, we have made assumptions such that *TC*s refrain altogether.)

Gary S. Becker and Kevin M. Murphy (1988) and Gary S. Becker, Michael Grossman, and Kevin M. Murphy (1991, 1994) discuss the role

of traumatic events (such as divorce, the death of a loved one, being fired) in causing people to consume addictive products. Within their stationary model, however, they are limited to formalizing traumatic events as discrete shocks to people's addiction level. By allowing for nonstationary instantaneous utilities, we can better endogenize traumatic events, formalizing them as short-term increases in the temptation to consume. Indeed, a model with traumatic events might be qualitatively similar to the youth model and the weekend-weekday model. For instance, we can reinterpret youth as the period of time directly following the traumatic event in which the temptation to consume an addictive good is high and maturity as the point at which the person has "recovered" from the traumatic event. Alternatively, we can imagine life as being full of traumatic events, in which case the weekend-weekday model could be interpreted as capturing the repeated fluctuations between normal times (that is, weekdays) and traumatic times (that is, weekends).

Predicting the effects of traumatic events in light of our other nonstationary models suggests that traumatic events are most likely to lead to severe addictions for naifs. Even when they do not want a lifelong addiction, naifs may end up with one because they consume when the temptation is high, thinking they will just quit once they recover. Of course, traumatic events can also cause TCs to get addicted when they would not in the absence of such events—but only if the shock is so severe that they prefer a lifelong addiction at the moment they first hit. We do not have a good empirical sense for how important such events are in inducing addiction (and, more likely, relapse), but if they *are* important, we suspect that any attempt to infer either the implicit discount rate or marginal utility of consuming the addictive product during such events would be more suggestive of naive self-control problems than a nonmyopic rational choice decision to begin a long-term addiction.[19]

Variable Myopia

In our discussions of stationary preferences and nonstationary preferences, we analyze behavior assuming that the extent of people's self-control problems does not vary at all over time. While observed propensity to succumb to temptation can vary because of changes in the scale of temptation—and indeed it is the role of habit formation in altering these trade-offs that is the crux of the role that self-control problems play in addiction—our examination of stationary and nonstationary preferences assumes that the degree of myopia itself is constant. We now consider two examples in which β varies over time. These examples further buttress our general impression that severely harmful addictive behavior

is more likely to arise from naive self-control problems than from sophisticated self-control problems.

In the sections on stationary and nonstationary preferences we assume that past consumption of addictive products affects current behavior only through its effects on instantaneous utilities. For many addictive products—particularly mind-altering substances such as alcohol—there is a second mechanism through which past consumption can influence current behavior: Very recent consumption might in fact increase the magnitude of self-control problems. For example, sober people may have only modest self-control problems, but once they start drinking alcohol, they may develop severe self-control problems. When drunk, they may virtually ignore the long-run consequences of their behavior and just pursue immediate gratification. We refer to this phenomenon as *consumption-induced myopia.*

To introduce consumption-induced myopia into the model, suppose that if people refrained last period, then their intertemporal preferences are described by equation 6.1 with $\beta = \beta_0$; but if people hit last period, then their intertemporal preferences are described by equation 6.1 with $\beta = \beta_1 < \beta_0$. In other words, people are especially myopic when they have consumed in the preceding period and are currently hooked.[20] We assume that time-consistent people are unaffected by consumption-induced myopia, and therefore the behavior of *TC*s will again represent the benchmark of how naifs and sophisticates would like to behave from a long-run perspective.

The assumptions of naivete and sophistication are essentially the same in this environment as in the basic model. In any given period, naifs believe that they will behave like *TC*s in the future. Sophisticates, on the other hand, are completely aware of their self-control problems, including the effects of consumption-induced myopia, and they therefore correctly predict future behavior.[21]

To see how consumption-induced myopia might matter, consider a nonstationary weekend-weekday example:

EXAMPLE 4: Suppose $f_o = 10$, $\rho = 15$, $\sigma = 18$, and $X = 8$.

Table 6.6 displays example 4. Example 4 is identical to example 3 except that the weekend temptation is larger. For $T = \infty$ and $\delta = .99$, it is straightforward to show that *TC*s always hit on weekends and refrain on weekdays. For the case $\beta_0 = \beta_1 = .9$, it is straightforward to show that sophisticates and naifs both behave exactly like *TC*s, so naifs and sophisticates both consume in moderation: Hit on weekends and refrain on weekdays.

Table 6.6 Example 4

Condition	Utility from Hitting	Utility from Refraining
On weekend when unhooked	18	0
On weekend when hooked	3	−18
On weekday when unhooked	10	0
On weekday when hooked	−5	−18

Now consider $\beta_1 < \beta_0 = .9$. For simplicity, we focus on $\beta_1 = 0$, which means that the consumption-induced myopia is severe. For naifs, β_1 is irrelevant to their decision when unhooked since it affects neither their current preferences nor their predictions of future behavior. Given $\beta_0 = .9$, naifs hit on weekends when unhooked and refrain on weekdays when unhooked; but β_1 is very relevant for naifs' decisions when hooked because it is incorporated into their current preferences. For $\beta_1 = 0$, naifs hit when hooked on both weekends and weekdays. Hence, naifs with consumption-induced myopia always hit. Extrapolating from our model, we can interpret this example as naifs becoming alcoholics not because they immediately start out drinking every day but because they start out drinking immoderately on nights they had intended to drink moderately. Then, because they become more and more hooked on alcohol, eventually they will start drinking every day.

For sophisticates, unlike naifs, β_1 can influence behavior when unhooked, since sophisticates correctly predict how they will behave when hooked, and this prediction can influence current behavior. For $\beta_1 = 0$, sophisticates of course always hit once they start hitting, just like naifs.[22] When unhooked, however, sophisticates anticipate—and disapprove of—their future behavior resulting from hitting on a weekend and therefore *never* hit. Note that sophisticates consume less of the addictive product than *TC*s. We call such an outcome *preemptive abstinence*, and tautologically this abstinence is not ideal: It *would* be preferable to drink moderately, but sophisticates recognize that their true choice is between total abstinence and total addiction, and their choice of abstinence is preferable to the total addiction to which naifs succumb.[23]

A second noteworthy aspect of this example, related to the preemptive abstinence, concerns comparative statics on β_1. For sophisticates, lowering β_1 decreases consumption—it can move sophisticates from consuming in moderation to not consuming at all. This contrasts with both naif behavior in the consumption-induced myopia model and either naif or sophisticated behavior in the unitary myopia model. In both those cases, people always consume more on average in

Table 6.7 Example 5

Condition	Utility from Hitting: $u(1, k)$	Utility from Refraining: $u(0, k)$
When unhooked ($k = 0$)	2	0
When hooked ($k = 1$)	−1	−5

response to intensifying the average self-control problem. Naifs never try to preempt self-control problems and hence can only respond to increases in such problems by succumbing more often. In the unitary myopia model, for all examples with sophistication that we have investigated, the direct effect of a stronger taste for immediate gratification always swamps the indirect effect of preemptive abstinence.

While the previous example assumes myopia may depend on recent behavior, myopia might also depend on exogenous forces. A period of depression may induce a lack of concern for the future consequences of one's actions. Various cues in the environment—such as seeing somebody else smoke—may induce a temporary temptation to consume the product that does not necessarily correspond to the enjoyment one will derive from the activity. Finally, consumption-induced myopia for one addictive product might affect the level of myopia for another addictive product. Just as people who are drunk may lose inhibition in drinking more, they may also lose the inhibition to smoke. Hence, in studying addiction to cigarettes, the exogenous event of whether people are drunk may lead to variations in ability to refrain from smoking.[24]

Consider the following stationary example:

Example 5: Suppose $f_o = 2$, $\rho = 3$, and $\sigma = 5$.

Table 6.7 illustrates example 5. In example 5, consider an infinite horizon with $\delta = .99$. To model the time variance of myopia in a simple and extreme way, suppose that $\beta_t = 1$ for odd t and $\beta_t = 0$ for even t. *TC*s refrain always, since any other course of action yields a negative average utility. Hitting always yields utility profile $2, -1, -1, -1, \ldots$, and the cost of refraining when hooked (−5) outweighs the benefit of hitting when unhooked (2), so any pattern of moderate consumption also will not be attractive.

Naifs and sophisticates both hit in even periods, whether hooked or unhooked, since in these periods they do not attend at all to their future well-being. What do they do in odd periods?

Sophisticates refrain in the first period but then hit in all future odd periods. In every odd period after the first period, sophisticates will find themselves currently hooked. Since they realize they will hit in even periods, they correctly anticipate that their choice is between hitting every other period versus hitting every period. Hence, their choice of utility profiles in all odd periods after the first is between (–5,2, –5,2, . . .) from alternating and (–1, –1, –1, . . .) from always hitting. Hitting always is preferable to repeatedly suffering the pain of withdrawal—only to repeatedly become addicted again.

Naifs likewise refrain in the first period, but naifs *will* make the mistake of repeatedly trying to quit their habit because they naively think that they will stay unhooked, in which case they perceive it as worthwhile to pay the cost of withdrawal.[25]

While sophisticates consume more than naifs in this example, sophisticates are in fact behaving more in their long-term interest. Both are consuming more than is optimal, but the harm from consumption is very much not monotonic in consumption—if people simply will not be able to control themselves often enough, they may in fact be better off living with their addiction than trying to eliminate it.[26] The more general point is that in a world of variable myopia, misguided attempts to quit addictions, followed by relapse, may represent another significant problem for naifs.

Conclusion: Self-Control Versus Rational Choice

Our goal in this chapter has been to outline some simple models of the relationship between self-control problems and addictive behavior. Researchers who use mathematical models to study human choice—mostly economists—traditionally approach intertemporal choice problems by assuming time consistency. By focusing on self-control problems, therefore, we depart from this traditional approach. We conclude by discussing some of the advantages of our self-control model of addiction relative to rational choice models of addiction.

Throughout, we have not addressed the issue of whether self-control problems can lead to behaviors that cannot be explained with time-consistent preferences. In fact, such smoking guns—*qualitative* predictions that are inconsistent with rational choice theory—are difficult to come by in our highly stylized and simplified models. In these models, only a few types of behavior can arise, and most of these behaviors could arise from time-consistent preferences.[27] One might ask, then, why it is worthwhile to study a self-control model of addiction. We feel there are a number of reasons.

The most obvious reason is simple realism. The evidence overwhelmingly supports the existence of a time-inconsistent taste for immediate gratification, and we conjecture that almost all social scientists, policy makers, and humans in general believe in their hearts that people have self-control problems. It is true that time consistency is a simpler assumption (and more familiar to economists), and it is clearly warranted to investigate human behavior with simplifying assumptions—in a sense this is one of the strengths of economics. But, it is also clearly warranted to investigate human behavior with more realistic assumptions, particularly in arenas such as addiction wherein common intuition is that a facet of human nature ignored by economists may matter.

Related to the issue of realism, we predict that models incorporating self-control problems (especially, we conjecture, models that include an element of naivete) will be better calibrated than rational choice models and hence make sounder *quantitative* predictions. We do not have empirical evidence for this conjecture, but to illustrate our reasoning we present a simple calibration exercise within our framework: We demonstrate how very patient people with very small self-control problems can get addicted in situations in which time-consistent people would get addicted only if they were to discount the future at an implausibly heavy rate.

Formally, we ask what discount factor $\hat{\delta}$ would time-consistent people need to have to match the behavior in a given example of people with (β, δ) preferences. Consider a stationary infinite-horizon model with a period length of one week—for example, each week people decide whether to indulge in an addictive activity. Consider people who have a (long-run) *yearly* discount factor .95 (that is, $\delta^{52} = .95$), where δ is the weekly discount factor. In addition, these people have very small self-control problems: They have an extra bias for this week's well-being over next week's well-being of only 1 percent (that is, $\beta = .99$). If these people were to make a one-shot decision concerning well-being this week versus well-being during a week one year from now, they would look very patient: Their discount factor for this range would be .9405. But, suppose these people must decide each week whether to consume an addictive product characterized by the following instantaneous utilities:

EXAMPLE 6: Suppose $f_0 = 10$, $\rho = 10.1$, and $\sigma = 10.1$.

Table 6.8 illustrates example 6. It is straightforward to show that these people always hit in this situation, whether they are sophisticated or naive. How impatient would time-consistent people have to be to always hit in this situation? It can be shown that time-consistent people with discount factor $\hat{\delta}$ always hit only if $\hat{\delta} \leq .99$., and since $\hat{\delta}$ is the per-period

Table 6.8 Example 6

Condition	Utility From Hitting: $u(1, k)$	Utility from Refraining: $u(0, k)$
When unhooked ($k = 0$)	10	0
When hooked ($k = 1$)	−10.1	−10.1

discount factor, this implies that time-consistent people always hit only if they have a yearly discount factor ($\hat{\delta}^{52}$) smaller than .6. Hence, people with yearly discount factor .95 and very small self-control problems of $\beta = .99$ behave in a way that is consistent with a rational choice model of addiction but only for implausibly low yearly discount factors smaller than .6. Moreover, more extreme calibration results arise if we consider a smaller period length or larger self-control problems. For instance, consider $\delta^{52} = .95$ as before, but now suppose $\beta = .95$. In this case, there exist instantaneous utilities such that sophisticates or naifs always hit whereas time-consistent people would always hit only for a yearly discount factor smaller than .07, which is ludicrously small.

The crucial intuition driving these calibration results is the incremental nature of most addictive behavior. At each point in time, people choose whether to indulge now, and the cumulative effect of these decisions determines whether people get and remain addicted. With self-control problems, a sequence of incremental decisions can lead to behavior very different from how people would behave if committing up front to a lifetime path of behavior. In a rational choice model, in contrast, the incremental nature of addiction is irrelevant. If people know exactly what the future holds, and have no self-control problems, then people become addicted only if that is the optimal lifetime path of behavior.

Indeed, this incremental decision-making intuition suggests ways that our self-control model of addiction might yield qualitatively distinct predictions in more complicated environments. For example, consider the possibility of nonlinear pricing, such as having a yearly fee in conjunction with a per-unit price. Rational choice models would suggest, for instance, that a (monopolist) tobacco company could increase profits by using such a two-part tariff, since presumably consumers are getting some surplus. In contrast, our self-control model of addiction suggests that such two-part tariffs are very much the wrong pricing strategy. For sophisticates, the yearly fee may be the commitment device needed to not become addicted. For naifs, our model suggests that at any point in time they may expect to consume very little (because they are planning to quit soon), and therefore naifs also would be unwilling to pay the yearly fee. Hence,

in richer models, allowing for self-control problems may in fact yield qualitatively distinct predictions.

The final—and in our view probably the most important—reason for studying self-control problems is that they predict very different *welfare* implications than the rational choice model. As discussed at the end of the section about stationary preferences, our model, unlike rational choice models, implies that people are hurting themselves with severe addictions.[28] To further illustrate this point, we reconsider the calibration example above. Suppose instantaneous utilities are as in example 6, but now consider a finite horizon ($T < \infty$) and $\delta = 1$. It can easily be shown that people with self-control problems (with magnitude $\beta = .99$), whether sophisticated or naive, will hit every period—irrespective of T. What is their stream of utilities for doing so? It is 10 in the first period, and $-.1$ for every period thereafter. For a one-shot instantaneous utility of 10, they experience a total negative utility for the last $T-1$ periods of their lives of $(T-1)(0.1)$. Obviously, if the number of periods in their lives becomes arbitrarily large, they suffer an arbitrarily large negative lifetime utility. Even from the period-1 perspective, where they receive their one-shot instantaneous utility of 10 and discount the future by $\beta = .99$, this outcome is clearly an unattractive option relative to never hitting. In other words, from *any* perspective self-control problems are causing severe harm.

It is perhaps unclear whether self-control problems will turn out, empirically, to be a major facet of cigarette and alcohol consumption, and other forms of addiction. Further investigation is required, extending and generalizing models such as those we present in this chapter (most notably, to allow for variable consumption levels and to consider the effects of prices) so as to make them testable. Models that investigate self-control problems are necessary, though, if economists or other researchers using formal models intend for their research to be deemed relevant by those who think it plausible that (on average) people are too addicted to harmful products for their own good.

We are grateful to David Laibson and other participants at the conference on addiction, and to an anonymous referee, for useful feedback; and to Doug Almond and especially Erik Eyster for research assistance. For financial support, we thank the National Science Foundation (Award 9709485), and Rabin thanks the Russell Sage and the Alfred P. Sloan Foundations. This project was started while the authors were visiting the Math Center at Northwestern University, and we are grateful for its hospitality and financial support. A draft of this chapter was completed while Rabin was a Fellow at the Center for Advanced Study in the Behavioral Sciences, supported by National Science Foundation Grant SBR-960123. He is extremely grateful for the center's hospitality and the NSF's support.

Notes

1. Although we assume that consumption each period is a binary choice (rather than a continuous choice), our model is essentially a simplified form of the Gary S. Becker and Kevin M. Murphy (1988) and Gary S. Becker, Michael Grossman, and Kevin M. Murphy (1991, 1994) models.
2. Negative internalities may include future health, career, or personal problems, as well as tolerance.
3. See Ted O'Donoghue and Matthew Rabin (1998) for a more general formulation and analysis of the model we develop in this chapter. Readers can also refer to that work for proofs of generalized versions of the results presented here, for which we have omitted proofs.
4. Conspicuously absent from our model is the ability to use external commitment devices. Alcoholics sophisticated about their self-control problems may, for instance, choose to check themselves into the Betty Ford Clinic. Note that naifs would not use external commitment devices since they always believe they will behave themselves in the future.
5. Products could also generate *positive internalities*, wherein past consumption increases current well-being (for example, jogging). We borrow the term internalities from Richard J. Herrnstein et al. (1993), who define an internality as a within-person externality. The temporal internality we consider is merely one possible type of internality. Since we assume people fully understand how current consumption affects future well-being, we are in fact assuming that people internalize the internality; more generally, this need not be the case.
6. Note that such tolerance can be dissociated from habit formation: If $f_t(1) < f_t(0)$ and $f_t(1) - g_t(1) < f_t(0) - g_t(0)$, then people get less pleasure from consuming *and* are less tempted to do so. While self-control still has a role to play in consuming such nonaddictive but harmful products, we conjecture that self-control problems are less costly in such contexts. In any event, we do not analyze such situations in this chapter.
7. For some recent discussions of empirical evidence of time inconsistency, see Richard H. Thaler (1991) and Richard H. Thaler and George Loewenstein (1992).
8. See George Ainslie (1991, 1992), George Ainslie and Nick Haslam (1992a, 1992b), George Ainslie and Richard Herrnstein (1981), Shin-Ho Chung and Richard Herrnstein (1967), Kris Kirby and Richard Herrnstein (1995), and George Loewenstein and Drazen Prelec (1992). For formal economic models of time-inconsistent preferences more generally, see for instance Robert H. Strotz (1956), Edmund S. Phelps and Robert A. Pollak (1968), Robert A. Pollak (1968), and Steven M. Goldman (1979, 1980).
9. This model has since been used by David Laibson (1995, 1997), Ted O'Donoghue and Matthew Rabin (1998, 1999, forthcoming), Carolyn Fischer (1997), and others.

10. These assumptions (and the labels) were originally laid out by Robert H. Strotz (1956) and Robert A. Pollak (1968). Most papers studying time-inconsistent preferences assume sophistication (for example, Laibson [1994a, 1995, 1997], Fischer [1997]). George Akerlof (1991) and Ted O'Donoghue and Matthew Rabin (1998, 1999, forthcoming) also consider naive beliefs.

11. The term "perfect" is a play on the standard game-theoretic notion of perfect equilibrium and here reflects that people believe that their future behavior will be rational. The term "perception" allows for people to have correct or incorrect beliefs about their own future behavior.

12. For both *T*Cs and naifs, the unique infinite-horizon perception-perfect strategy corresponds to the unique finite-horizon perception-perfect strategy as the horizon becomes long.

13. This is the most interesting case, since if *T*Cs hit in all contingencies then so do naifs and sophisticates, and if *T*Cs hit when hooked then so do naifs and sophisticates.

14. This inequality follows from the assumption that *T*Cs would refrain when hooked. If *T*Cs would hit when hooked, the inequality would be reversed. Moreover, the discerning reader will notice that in that case the incentive effect being operative means that sophisticates perceive the same benefit from restraint as *T*Cs and naifs (and they all hit when hooked).

15. This conclusion relies on our restricting our attention to infinite-horizon, perception-perfect strategies that correspond to a perception-perfect strategy for some long, finite horizon.

16. This intuition corresponds to the standard game-theoretic result that making outcomes worse in some contingencies can help people because they may now avoid getting into those contingencies.

17. We choose $\delta = .99$ for this example because of our interpretation of period length as half of a week. For such a period length, any time-consistent discount factor must be close to one. (Indeed, even $\delta = .99$ implies a somewhat small yearly discount factor of .59.)

18. Recall that we restrict attention to perception-perfect strategies corresponding to the unique perception-perfect strategy for a finite horizon as the horizon becomes long. For a finite horizon, we suppose the last period is a weekend, and of course people hit whether or not they are hooked on this weekend.

19. Gary S. Becker and Kevin M. Murphy (1988) invoke traumatic events such as divorce to explain how people might start consuming an addictive product, but do not present any formal analysis of that decision. Athanasios Orphanides and David Zervos (1995) and Ruqu Wang (1997) more directly consider the decision to become addicted. Both papers emphasize the case in which people are uncertain as to how addictive a product is and experiment to find out. The logic of this section suggests naifs could suffer severe addictions in that environment because they experiment with overopti-

mistic beliefs about ease of quitting an accidental addiction. Sophisticates may suffer from the reverse problem. They may underexperiment because of a fear of getting addicted. Our general theme arises again: Sophisticates are unlikely to suffer an unwanted severe lifelong addiction, whereas naifs are far more likely.

20. Here again we emphasize that the dichotomous weekend-weekday model should not be taken too literally, and also draw attention to the restrictiveness of our assumption that people become immediately unaddicted after one period of restraint. Both aspects of our model exaggerate the resemblance of consumption-induced myopia to habit formation, when in more general models they would be much more distinct. Consumption-induced myopia implies that very recent consumption leads to more consumption of the addictive product—well beyond the habit formation plausibly induced by the recent consumption. It also dissipates immediately upon short-term cessation of consumption. If people start drinking heavily at eight o'clock in the evening, by ten o'clock they may be binge drinking without any regard to consequences. This will be true despite the fact that the two hours of drinking has not in any way made them alcoholics (indeed, the myopia induced by two hours of heavy drinking is likely to be much more intense for a novice than an experienced—and alcoholic—drinker). Both our assumption of $\gamma = 0$ (that addiction depends solely on the previous periods consumption) and our use of two- and three-period models leads to an artificial conflation of the two phenomena. Even within this simplistic model, however, one important distinction *does* show up: The *welfare* implications of consumption are very different if it comes from intensified myopia rather than habit formation. In our model, and in life, an alcoholic often benefits enormously in terms of current well-being from taking another drink; persistent consumption by addicts can sometimes be rationalized by cost-benefit analysis. The hypothesis of consumption-induced myopia may be that people consume a product that brings them virtually no pleasure, even in the short run. Indeed, although we focus on the habit-forming aspect of addictive products, products that induce myopia by altering one's perspective may be vastly overconsumed even if they are not at all addictive.

21. In addition to naifs and sophisticates, there are some natural hybrids to consider in this modified environment; for example, people might know β_0 but incorrectly believe $\beta_1 = \beta_0$. We doubt the plausibility of a sophisticated drunk; but allowing sophisticates to be naive while drunk would not affect our example below and would probably yield qualitatively similar predictions in more general settings. What is crucial is that sophisticates *when sober* anticipate the loss of control when drunk.

22. For less extreme values of β_1, sophisticates may stop hitting for values that naifs do not.

23. Such preemptive abstinence does not require consumption-induced myopia. Indeed, preemptive abstinence can arise in nonstationary models of the type discussed earlier.

24. A number of important issues, beyond the scope of this chapter, are raised by examples described above. First, it is not clear that these examples all really correspond to variations in β rather than variations in the marginal instantaneous utility of consuming the product. Although our impression is that alcohol-induced propensity to smoke cigarettes is not about a change in the utility function, it is far less clear that cues that make some activity salient do not directly affect the experienced well-being from engaging in the activity. Similarly, smoking, eating, or taking mind-altering drugs may be more utility enhancing when people are depressed than when they are not. We have not analyzed the variant-utility case sufficiently to know its implications but suspect it would be similar in many ways to the variable-myopia model.

    A second issue concerns the degree to which changes in myopia from some of these sources are genuinely exogenous; just as people (if sophisticated) may avoid drinking out of fear of drinking to excess, so too people may avoid it out of fear of smoking to excess. Similarly, people may sensibly try to avoid certain cues that might set off addictive behavior—avoiding being around other smokers if they are trying to quit smoking. For work that discusses some of these issues, and departures from the simple discounting model of self-control problems, see David Laibson (1994b) and George Loewenstein (1996).

25. If the taste for immediate gratification in even periods were sufficiently strong, of course, they would (fortunately for them) procrastinate in attempting to withdraw. This example does not rely on the extreme assumption that there is no self-control problem in even periods; so long as $\beta_t > .8$ for t even naifs would repeatedly try to quit.

26. This pattern, and the suspicion of its suboptimality, is well known in weight control: Huge numbers of people "successfully" lose weight on diets only to regain it. We do not know the extent to which this phenomenon results from the type of logic described in this simple example. Of course, none of our models apply per se to overconsumption of food. Although obesity resulting from overconsumption of food is clearly an example of a negative internality, the habit formation aspect of addiction that we emphasize in our model is not present—or at least it is far more subtle. Nonetheless, especially since we do not carefully formulate in this chapter which results come from habit formation and which come from the negative internality, we believe it would be useful to apply similar analysis to the case of eating and other nonaddictive activities.

27. It is also the case that rational choice models of addiction tend not to make qualitative predictions that are inconsistent with self-control models of addiction. Essentially all qualitative implications emphasized in rational choice models of addiction are also consistent with our self-control model of addiction. For instance, extensions of our model (and all other reasonable models we can imagine) would be consistent with the prediction that demand for addictive products decreases with the price of those products—

which is perhaps the main empirical finding of Gary S. Becker, Michael Grossman, and Kevin M. Murphy (1991,1994).

28. We remind the reader that it is not obvious what the welfare criterion should be (as noted in our discussion about stationary preferences). Although we do not formalize any of the welfare claims made in this section, we are confident that variants of all our claims can be articulated using any reasonable welfare criterion.

References

Ainslie, George. 1991. "Derivation of 'Rational' Economic Behavior from Hyperbolic Discount Curves." *American Economic Review* 81: 334–40.

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

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

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

Ainslie, George, and Richard J. Herrnstein. 1981. "Preference Reversal and Delayed Reinforcement." *Animal Learning and Behavior* 9: 476–82.

Akerlof, George A. 1991. "Procrastination and Obedience." *American Economic Review* 81: 1–19.

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

Becker, Gary S., Michael Grossman, and Kevin M. Murphy. 1991. "Rational Addiction and the Effect of Price on Consumption." *American Economic Review* 81: 237–41.

———. 1994. "An Empirical Analysis of Cigarette Addiction." *American Economic Review* 84: 396–418.

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

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

Goldman, Steven M. 1979. "Intertemporally Inconsistent Preferences and the Rate of Consumption." *Econometrica* 47: 621–26.

———. 1980. "Consistent Plans." *Review of Economic Studies* 47: 533–37.

Herrnstein, Richard J., George Loewenstein, Drazen Prelec, and William Vaughan. 1993. "Utility Maximization and Melioration: Internalities in Individual Choice." *Journal of Behavioral Decision Making* 6: 149–85.

Kirby, Kris, and Richard J. Herrnstein. 1995. "Preference Reversals Due to Myopic Discounting of Delayed Reward." *Psychological Science* 6: 83–89.

Laibson, David. 1994a. "Essays in Hyperbolic Discounting." Department of Economics, Massachusetts Institute of Technology.

———. 1994b. "A Cue Theory of Consumption." Department of Economics, MIT.

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

———. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112: 443–77.

Loewenstein, George. 1996. "Out of Control: Visceral Influences on Behavior." *Organizational Behavior and Human Decision Processes* 65: 272–92.

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

O'Donoghue, Ted, and Matthew Rabin. 1998. "Addiction and Present-Biased Preferences." Cornell University and University of California, Berkeley.

———. 1999. "Doing It Now or Later." *American Economic Review* 89: 103–24.

———. Forthcoming. "Incentives for Procrastinators." *Quarterly Journal of Economics.*

Orphanides, Athanasios, and David Zervos. 1995. "Rational Addiction with Learning and Regret." *Journal of Political Economy* 103: 739–58.

Phelps, Edmund S., and Robert A. Pollak. 1968. "On Second-Best National Saving and Game-Equilibrium Growth." *Review of Economic Studies* 35: 185–99.

Pollak, Robert A. 1968. "Consistent Planning." *Review of Economic Studies* 35: 201–8.

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

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

Thaler, Richard H., and George Loewenstein. 1992. *Intertemporal Choice."* In *The Winners Curse: Paradoxes and Anomalies of Economic Life,* edited by Richard H. Thaler. New York: Free Press.

Wang, Ruqu. 1997. "The Optimal Consumption and the Quitting of Harmful Addictive Goods," Queens University.

The Russell Sage Foundation

The Russell Sage Foundation, one of the oldest of America's general purpose foundations, was established in 1907 by Mrs. Margaret Olivia Sage for "the improvement of social and living conditions in the United States." The Foundation seeks to fulfill this mandate by fostering the development and dissemination of knowledge about the country's political, social, and economic problems. While the Foundation endeavors to assure the accuracy and objectivity of each book it publishes, the conclusions and interpretations in Russell Sage Foundation publications are those of the authors and not of the Foundation, its Trustees, or its staff. Publication by Russell Sage, therefore, does not imply Foundation endorsement.

Library of Congress Cataloging-in-Publication Data

Addiction : entries and exits / edited by Jon Elster.
p. cm.
Includes bibliographical references and index.
ISBN 0-87154-235-8
1. Addicts—Psychology. 2. Substance abuse—Etiology. 3. Self-control.
I. Elster, Jon, 1940– .
RC564.A282 1999
616.86—DC21 99-31479
CIP

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials. ANSI Z39.48-1992.

RUSSELL SAGE FOUNDATION
112 East 64th Street, New York, New York 10021
10 9 8 7 6 5 4 3 2 1

Contents