Learning Self-Control*

S. Nageeb Ali†
University of California, San Diego

November, 2009

Abstract

This paper examines how a decisionmaker who is only partially aware of his temptations learns about them over time. In facing temptations over time, individuals use their experiences to forecast future self-control problems and choose the appropriate level of commitment. I demonstrate that rational learning can be perpetually partial and need not result in full sophistication. The main result of this paper characterizes necessary and sufficient conditions for learning to converge to full sophistication. This result has implications for commonly studied self-control environments such as savings and addiction.

Keywords: self-control, partial awareness, Bayesian learning, sophistication, experimentation, dual-selves, self-confirming equilibrium

*This paper is a revised version of Chapter 3 of my dissertation. I thank Ricardo Alonso, Manuel Amador, Susan Athey, Roland Bénabou, Dan Benjamin, Doug Bernheim, Aislinn Bohren, Juan Carrillo, Vince Crawford, Stefano DellaVigna, Ignacio Esponda, Drew Fudenberg, Ben Ho, Shachar Kariv, Navin Kartik, Botond Köszegi, Troy Kravitz, Jon Levin, Ulrike Malmendier, Shamim Sinnar, Joel Sobel, Joel Watson, Tom Wiseman, and especially Paul Niehaus for useful feedback and discussions. I acknowledge financial support from the UCSD Academic Senate, the UCSD Hellman Fund, the Stanford Institute for Economic Policy Research, and the Institute for Humane Studies.

†Email: snali@ucsd.edu; Web: http://econ.ucsd.edu/∼snali; Address: 9500 Gilman Drive, Department of Economics, La Jolla, CA 92093-0508.

Contents

"An individual who finds himself continuously repudiating his past plans may learn to distrust his future behavior, and may do something about it."
— Strotz (1955)

1 Introduction

Questions of temptation and costly self-control are at the forefront of psychology and economics. An important choice in analyzing self-control is in deciding how aware a decisionmaker is of his temptations since different assumptions of self-awareness naturally translate into different behavioral predictions. The assumption most commonly made, referred to as *sophistication*, is that the decisionmaker perfectly anticipates future self-control problems. Yet, sophistication is often seen as an inaccurate model of a decisionmaker's awareness, especially in the short- and medium-run when the decisionmaker may be inexperienced. An alternative assumption is that of *naivete*, whereby a decisionmaker anticipates having perfect self-control in the feature. Bridging the gap between these two extreme assumptions, O'Donoghue and Rabin (2001) propose a framework of partial naivete in which a decisionmaker with (β, δ) preferences assigns probability 1 to a particular level of present bias $(\hat{\beta})$ more moderate than his true present-bias $(\hat{\beta} > \beta)$.

With a few exceptions, most models of study imperfect self-control make one of the above assumptions. That derived behavior is sensitive to the specification of beliefs has been widely recognized, including by those that first modeled self-control (Strotz, 1955; Pollak, 1968). Nevertheless, the practice has been to treat a decisionmaker's belief about his self-control as exogenous and derive dynamic choice given that exogenous belief. While this practice makes analysis tractable, it raises two conceptual issues. First, in models of partial sophistication, a decisionmaker's beliefs may be incompatible with what is observed over time, and so ancillary assumptions are necessary for analysis.[1] Second, given the many different awareness assumptions that could be made, it has been difficult to assess which assumption is appropriate for a particular environment and why. Fudenberg (2006), in his recent discussion of behavioral economics, voices this concern.

> "I think that behavioral economics would be well served by concerted attempts to provide learning-theoretic (or any other foundations) for its equilibrium concepts. At the least, this process might provide a better understanding of when the currently used concepts apply...."

This paper proposes a simple framework to address these concerns. In this approach, beliefs and choices are derived endogenously and jointly evolve based on the decisionmaker's experience. Endogenizing beliefs in this way allows one to pose and answer the question of whether

[1]For example, a partially naive decisionmaker may see behavior that contradicts his view of the world in each and every period. The literature typically addresses this issue by assuming that the DM's beliefs are not revised, which implicitly assumes that the DM does not receive feedback about payoffs or the past history.

and when sophistication closely approximates the decisionmaker's self-awareness once he has had many opportunities to learn. Although one may tempted to conclude that learning should always engender sophistication, I demonstrate that tradeoffs between commitment and flexibility inherent in self-control environments may actually impede learning. The main result of this paper is to offer a necessary and sufficient condition—across self-control environments—under which such possibilities arise, and derive when instead, learning inexorably leads to sophistication. This condition can be easily checked in applications to assess the appropriateness of sophistication in an environment, and as such, these results have implications for how we should think about self-awareness in different settings.

As one application of this approach, I analyze a standard buffer-stock consumption-savings environment in which a decisionmaker has a temptation to overconsume in each period and can commit by purchasing illiquid assets. This setting has been the focus of many papers in the quasi-hyperbolic literature (e.g. Laibson, 1997; Barro, 1999), almost all of which assume sophistication. I demonstrate that this canonical environment satisfies the condition for sophistication: thus, a decisionmaker who learns about his tendency to overconsume from his past choices eventually chooses commitment as if he could perfectly forecast his temptation to overconsume. This learning-theoretic justification for focusing on sophistication in savings environments tells us that in this important setting, sophistication emerges naturally from experience.

In contrast to this, I also demonstrate that the condition for sophistication may fail in standard applications. Building on the seminal contribution of Bernheim and Rangel (2004), I analyze the behavior of a decisionmaker who learns about how easily he becomes addicted. I demonstrate that such a decisionmaker can make long-run choices that depart from how a sophisticated individual would make choices, and that this analysis can provide insight into why certain interventions—such as temporary subsidies to drug rehabilitation—have long-lasting implications on rehabilitation.

The general framework that I develop builds on the Planner-Doer approach to self-control: decisions are made through the conjunction of a single long-run Planner (with dynamically consistent preferences) and a myopic Doer who is a short-run player or behavioral type. Prior research using such models, particularly Fudenberg and Levine (2006), has demonstrated that such models offer analytically simple and tractable frameworks to understand self-control in a variety of settings,[2] and this paper illustrates how this approach can be used to address questions of awareness. The Planner represents the forward-looking rational individual that chooses how much to commit by investing in illiquid assets, signing contracts etc. In contrast, the Doer represents the instinctive response of the individual that makes the daily choices, and is presented with stimuli, situations, and choices; as such, the Doer should be thought of

[2] Thaler and Shefrin (1981) were the first to propose dual-self models to model imperfect self-control. There have been a number of other recent papers—Benabou and Pycia (2002), Bernheim and Rangel (2004), Benhabib and Bisin (2005), Loewenstein and O'Donoghue (2007), Brocas and Carrillo (2008)—that use dual-self models. Dekel and Lipman (2007), Chatterjee and Krishna (2009), and Ambrus and Rozen (2009) offer axiomatizations of dual-self models.

as a short-run player or behavioral type. While the behavior of the Doer may be cultivated through good habits, the Planner may be uncertain about the extent to which the Doer resists temptation, and attempts to extract this information from the Doer's behavior.

To fix ideas, I describe the example explored in Section 2: suppose that an individual faces a choice between eating fish and steak in each period. In the beginning of the period, the Planner makes a reservation either at a restaurant that features simply fish and the other that features both fish and steak. The payoff from eating fish is constant over time and across the two restaurants, but that from eating steak depends on an i.i.d. taste shock that is realized after the Planner selects the menu. Eating fish is *ex ante* optimal, but there are contingencies in which steak may be preferred. As in Kreps (1979), the Planner would strictly prefer the flexibility associated with the larger menu were self-control not an issue; however, once the decisionmaker is at the restaurant, the choice of what to eat (if there is a choice) is made by the Doer. The Doer's preferences are partially aligned with that of the Planner, and partially reflect a temptation to eat steak in contingencies in which the Planner prefers fish (because of its associated long-term benefits). The Planner is uncertain about the strength of the Doer's temptation and would benefit from learning it over time since that helps him plan in the future. Were the decisionmaker to know that in the restaurant, he could resist temptation, he may strictly prefer the larger menu to utilize his informational advantage. In contrast, if he believes that he would be too easily tempted by the opportunity to eat steak, the decisionmaker may find it *ex ante* optimal to select the smaller menu, which offers commitment value.

Learning about one's ability to resist temptation involves costly experimentation (and, in this case, is isomorphic to a two-armed bandit): the Planner observes the Doer's self-control *only* when the Planner chooses the larger menu and exposes the Doer to temptation. Accordingly, once the Planner becomes sufficiently pessimistic about the Doer's type, he chooses the smaller menu since the value of flexibility and learning no longer outweigh the expected sacrifice of the Doer succumbing to temptation. Notably, the Planner may make this decision with positive probability even when the Doer is a "good type" who resists temptation. The Planner can therefore choose to remain perpetually biased even when given infinite possibilities to learn.

The direction of the skewness in these beliefs merits discussion. The Planner cannot perpetually overestimate the Doer's self-control and undercommit relative to how he would choose in the full information benchmark. To see why, notice that any belief that rationalizes flexibility will allow the Planner to continue to passively learn and therefore update his beliefs. Thus, if the Doer is unable to resist temptation, the Planner almost-surely learns this over time, and eventually makes the same commitment choice that he would in the complete-information benchmark. While such optimism necessarily erodes over time and is not sustainable in the long-run, it may be present in the short and medium-run especially if the Planner begins with a prior that greatly overestimates the Doer's ability to resist temptation.

On the other hand, the Planner may perpetually underestimate the Doer's self-control even if he began with an overoptimistic prior: once the Planner has observed the Doer capitulating to temptation too often, flexibility appears too costly in the short-term, and a Planner will choose to actively learn for as long as the future value of learning outweighs its cost. So long as the Planner is not perfectly patient, he is willing to undertake a finite number of trials after which he may commit to the singleton menu and never revise his belief thereafter. Thus, partial awareness and learning can endogenously lead to perpetual *overregulation* whereby individuals choose rigid commitments or lifestyles that they would not were they more self-aware.

A feature of the setting described above is that the Planner chooses between full flexibility or full commitment (through singleton menus), but has no ability to partially commit. In contrast, many settings feature a variety of partial commitments that retain some flexibility for the decisionmaker while providing a measure of commitment. For example, in the savings environment, an individual can purchase illiquid assets to ensure some minimal savings without relinquishing flexibility altogether. Contractual mechanisms, such as recent innovations like *stickK.com*,[3] can penalize certain choices and provide incentives to counteract temptation while retaining some flexibility. Social mechanisms—through promises, shame, and peer groups—also provide partial commitment insofar as they change the benefits and costs of particular actions, but are not completely binding. Apart from external partial commitments, a decisionmaker may also rely on internal commitments, such as *costly self-control* (Gul and Pesendorfer, 2001; Fudenberg and Levine, 2006), which influence the choice from a menu without relinquishing flexibility altogether. Apart from the static benefit of intermediate commitment, the framework here demonstrates that these partial commitments play a role in the decisionmaker's long-run behavior and beliefs.

The main result of this paper identifies exactly when partial commitments induce efficient learning regardless of the Planner's patience. The necessary and sufficient condition for such learning to emerge, *Full Commitment Monotonicity* (FCM), is a condition on how fully informed Planners behave, and relates to the richness of the set of partial commitments. In application, this condition can be assessed through either the modeler's understanding of the environment or carefully designed choice experiments. The condition naturally translates into a monotonicity condition, which may or may not be satisfied depending upon the range of commitment technologies. As mentioned above, I explore the implication of these results in three different settings that have been studied in prior work assuming sophistication: consumption-savings with illiquid assets, addiction with cue-triggered compulsions, and a costly self-control environment. Within these settings, I illustrate how to check the validity of FCM, and derive the implications within that theory.

The results of this paper underscore the insights that emerge from deriving the decision-maker's understanding of his temptations from his environment. Based on these results, one may expect that in settings that feature a wide range of partial commitments, people may know more about their temptations than in settings in which partial commitments are less effective. From a normative perspective, this role of partial commitments has implications for the design

[3] *StickK.com* is a recent contractual innovation designed to "...help people achieve their goals and objectives by enabling them to form Commitment Contracts." (http://www.stickk.com/about.php).

and pricing of commitment. Prior work that assumes a particular belief for a decisionmaker and derives dynamic choices holding that belief fixed is silent on how beliefs and choices evolve over time and vary by setting, and from that perspective, this study generates a new perspective on self-control and self-awareness.

Some readers may be troubled by the disappearance of overoptimism at the limits of Bayesian learning, although to my knowledge, there is no evidence to suggest that people persistently undercommit. While I use the approach developed here to study asymptotic behavior, the framework is sufficiently flexible and tractable to model a decisionmaker's initial response and these initial responses may reflect initial optimism. Indeed, there are numerous reasons to believe that decisionmakers should begin with optimistic priors and learn slowly,[4] and so findings in the field may be consistent with short- and medium-run behavior. Moreover, to the extent that decisionmakers often have to simultaneously learn about their self-control and the benefits or costs of different actions, one should expect learning in a multidimensional environment to be particularly slow. Indeed, I show through example that such environments present an identification challenge that impedes learning severely and cause a decisionmaker to undercommit relative to a full information benchmark. As such, there are natural reasons to expect learning to be slow in certain settings, as in health, when the decisionmaker is simultaneously learning about self-control and the benefits and costs of actions.

Section 2 provides a simple example of the impediment to learning introduced by imperfect self-control; the example reduces to a standard two-armed bandit. Section 3 considers a richer model with partial commitments, in which experimentation takes a more general form. In that section, I derive the main characterization of this paper. Section 4 applies the framework to savings behavior, addiction, and the possibility for costly self-control. Section 5 discusses the results of this paper in light of the related literature and offers the example of the identification challenged induced by multidimensional learning. The Appendix contains all proofs.

2 A Simple Example

To illustrate the mechanism for incomplete learning in the most transparent way, I analyze a simple example. Consider an infinitely-lived individual who chooses between undertaking an activity ($a_t = 1$) or not ($a_t = -1$) in each period t in $\{0, 1, 2, 3, ...\}$. Undertaking the activity in period t generates a deterministic reward $b \in \left(\frac{1}{2}, 1\right)$ but involves a stochastic cost c_t uniformly drawn from $[0, 1]$. This activity can represent, for example, the choice to consume fish from the example described in the introduction or to exercise, in which case b captures the discounted long-run gain from the activity.

The decision in each period is made through the conjunction of two systems, the *Planner* and the *Doer*, who act sequentially in each period. The Planner is a long-run agent with an

[4] Optimism could emanate from a number of different sources, if a positive belief about one's attributes has some intrinsic value (Brunnermeier and Parker, 2005; Kőszegi, 2006; Dal Bo and Tervio, 2008) or induces motivation (Benabou and Tirole, 2002; Compte and Postlewaite, 2004).

exponential discount factor $\delta \in (0,1)$, and obtains $(b - c_t)$ if $a_t = 1$ and 0 otherwise. In the 1^{st} sub-period of each period t, the Planner is forced to make a commitment choice prior to the realization of c_t. He chooses between full commitment to undertaking the activity ($z_t = 1$), full commitment to not undertaking the activity ($z_t = -1$), and *flexibility* ($z_t = 0$) which delegates the decision to the Doer who acts in the 2^{nd} sub-period. When the Planner chooses to commit to an action, the action is undertaken regardless of the Doer's type or the realized cost; one way to interpret such commitments is that the Planner chooses a singleton menu that constrains the subsequent choice entirely.

The Doer in each period is not a strategic actor but can be thought of as a behavioral type or short-run player. Nature selects the type, θ, of the Doer from $\{\underline{\theta}, \bar{\theta}\}$ where $0 \leq \underline{\theta} < \bar{\theta} \leq 1$, at the beginning of the game and this type persists throughout the game. When the Planner is flexible, the Doer of type θ observes c_t and chooses $a_t = 1$ if and only if θb exceeds c_t. Thus, the Doer is tempted towards inactivity, and undertakes the activity in fewer contingencies than the Planner would wish to do so. This temptation may emerge from a manifestation of a present-bias when costs and rewards are temporally separated. The perspective here is that the Doer being myopic (representing the individual's *instincts*) does not account for how its choices influence the Planner's decisions.

The Planner faces a tradeoff between commitment and flexibility: while he would like to trust the Doer's informational advantage, flexibility opens the door for temptations to guide that choice. Since the payoff from flexibility depends on the Doer's susceptibility to temptation, it is potentially valuable for the Planner to try to learn about θ so as to make the right decision in the future. I assume that the Planner would prefer to commit to undertaking the activity if he were perfectly confident that $\theta = \underline{\theta}$ and would prefer to remain flexible if he were confident that $\theta = \bar{\theta}$.[5]

He begins with a prior $\mu_0 = \Pr\left(\theta = \bar{\theta}\right)$ that ascribes positive probability to both types. For purposes of simplicity, suppose that all that he observes over time is behavior and not past realizations of costs (I consider more general informational structures in Section 3). Since full commitments override the Doer, the Planner learns about θ only when he chooses to be flexible. I use h^t to summarize the relevant history for the Planner when acting at time t, and μ_t to denote the Planner's posterior belief that the Doer's type is $\bar{\theta}$. Given a prior belief μ, the Planner updates his beliefs to μ^+ and μ^- when he is flexible and observes the Doer choose $a_t = 1$ and $a_t = -1$ respectively. The dynamic decision problem of the Planner is described by the value function

$$V(\mu) = \max \left\{ \begin{array}{l} b - E[c] + \delta V(\mu), \\ \sum_{\theta \in \{\underline{\theta}, \bar{\theta}\}} \Pr^{\mu}(\theta) \left(\theta b \left(b - E[c | c \leq \theta b] + \delta V(\mu^+)\right) + (1 - \theta b)\, \delta V(\mu^-)\right) \end{array} \right\}, \quad (1)$$

where the 1^{st} term is the value of committing to undertake the action, and the 2^{nd} term is

[5] The relevant condition is $b \in \left(\frac{1}{2-\underline{\theta}}, \frac{1}{2-\bar{\theta}}\right)$; when this does not hold, the Planner's optimal commitment choice is independent of his belief about the Doer's type.

the value of flexibility. The Planner never commits to $a = -1$ since *ex ante*, this is dominated by committing to $a = 1$. Standard arguments ensure that the value function exists, is unique, continuity, and non-decreasing in μ, and thus, the optimal decision takes the form of a simple threshold rule:

Proposition 1. *There exists $\mu^* \in (0, 1)$ such that for all $\mu < \mu^*$, the Planner's optimal choice is to commit to the activity and for $\mu \geq \mu^*$, the Planner's optimal choice is flexibility.*

When the Planner is more optimistic about the Doer's ability to resist temptation or values the option to learn, he is more willing to remain flexible. Because of the option-value associated with flexibility and learning, a forward-looking Planner is willing to choose flexibility even if he expects that committing yields greater short-run payoffs. For less optimistic beliefs, when $\mu < \mu^*$, the benefits of flexibility and learning do not offset the expected costs of temptation and therefore the Planner chooses to commit. Since the choice of commitment shuts down the channel for learning, once he chooses to commit in one period, he finds it optimal to commit in every subsequent period. Because the choice to commit is endogenous, eventual beliefs are endogenous, and therefore evolve differently for each type of the Doer (assuming that the prior μ_0 exceeds μ^*).

Consider the case in which the Doer is of type $\underline{\theta}$. The Planner's beliefs will fluctuate so long as he remains flexible and his beliefs remain bounded away from putting probability 1 on $\bar{\theta}$. The Martingale Convergence Theorem ensures that the Planner's beliefs must eventually settle down but that convergence to the completely incorrect belief of $\mu = 1$ does not happen. Therefore, the Planner must eventually commit, making the same choice that he would in the complete-information benchmark; initial partial awareness does not forge any long-lasting differences from the standard benchmark.

Now suppose that the Doer is of type $\bar{\theta}$. In this case as well, the Martingale Convergence Theorem implies eventual convergence of the Planner's beliefs. Convergence obtains in two distinct ways: either the Planner's beliefs converge to the truth ($\mu_t \to 1$) or the Planner's belief μ_t falls below μ^*, leading him to commit. Both events occur with positive probability. Recall that in the complete-information environment, a Planner who knew that his self-control problem is $\bar{\theta}$ would never choose to commit. Thus, relative to this benchmark, partial awareness and learning introduce a possibility for long-run inefficiency and overcommitment. The preceding ideas are summarized below.

Theorem 1. *For any type $\theta \in \{\underline{\theta}, \bar{\theta}\}$, a Planner eventually chooses to commit with strictly positive probability.*

1. *If the Doer is of type $\underline{\theta}$, then almost-surely, the Planner eventually chooses to commit.*

2. *If the Doer is of type $\bar{\theta}$, then the Planner either chooses to commit or he learns the Doer's type. Both events occur with strictly positive probability if $\mu_0 \geq \mu^*$.*

This result illustrates how any beliefs that induce excessive flexibility eventually vanish: if the Doer is of type $\underline{\theta}$, any belief of the Planner that rationalizes flexibility is refined over time as the Planner naturally infers the Doer's type from its choices. Almost-surely, this information leads the Planner to conclude that the Doer's type does not warrant flexibility, and therefore, the Planner eventually chooses to commit. While overoptimism disappears eventually, it may persist in the short and medium-run, especially if individuals begin with such priors.

What is the source for the friction from underestimating self-control? The principal challenge that the Planner faces is that when he optimally chooses to commit (believing that he faces a Doer of type $\underline{\theta}$), his commitment choice leaves no opportunity for him to distinguish between different types. Partial commitments offer an opportunity for him to escape this experimentation trap, as I explore below.

For simplicity, suppose that the individual has risk-neutral and additive preferences, and consider a one-period bond contract in which the Planner pays a lump-sum amount L and a prize x is returned to the decisionmaker in the event that the Doer chooses $a = 1$. There are various ways in which one could model how such a contract affects the Doer; for simplicity, suppose that a Doer of type θ chooses to take the action if $\theta b + x$ exceeds the cost c. In this case, a fully informed Planner regardless of the Doer's type can implement the first-best by setting x to equal $(1 - \theta)\, b$, and paying a prior lump-sum amount of $(1 - \theta)\, b^2$. This bond contract leads to zero expected transfers and induces the Doer to choose $a = 1$ whenever the Planner would like to do so.

Not only does such a bond contract allow a fully informed Planner to achieve first-best, but it also helps a partially aware Planner learn efficiently. Through such a contract, regardless of the Planner' beliefs, it is now never optimal to restrict the Doer's choice set. Even when the Planner becomes quite confident that the Doer's type is $\underline{\theta}$, he is better off with a bond contract that retains flexibility. Over time, the Planner is able to observe the Doer's behavior and therefore continue learning; almost-surely, his beliefs converge to the truth.

The possibility for partial commitments to dynamically improve learning has implications for commitment design. Recent years have seen the advent and study of commitment contracts that incentivize good behavior and mitigate self-control problems.[6] The following section analyzes more generally when partial commitments lead to efficient learning. The reader particularly interested in the application of these results can skip to Section 4 in which I highlight the implications of partial commitments (and their absence) in canonical self-control environments.

3 When is Learning Adequate?

The general model developed in this section builds on the above example in three ways. First, the Doer's type is drawn from a continuum. Second, the Planner has a range of commitment

[6] Internet-based services such as *stickK.com* permit an individual to sign a contract (specifying a monitor) that involves penalties if a particular target is not met. Giné et al. (2009) find that a bond commitment contract helps smokers cope with their addiction.

technologies that may include partial commitment options. Third, the Planner observes signals about past taste shocks.

The timing of the game is unchanged. The Planner has a stage-payoff function in which $a_t = -1$ leads to a payoff of 0 and $a_t = 1$ leads to a payoff of u_t. This payoff, u_t, is drawn from $\mathcal{U} \equiv [\underline{u}, \bar{u}]$ where $\underline{u} < 0 < \bar{u}$, distributed i.i.d. with cdf $F(\cdot)$ that has a continuous and strictly positive density and a strictly positive expectation. The Planner cannot directly choose actions but chooses commitment from a range of commitment choices z_t in Z, a closed subset of $[-1, 1]$ that contains $\{-1, 0, 1\}$. Z is a parameter of the model, representing a range of commitments or ability to exert self-control. My analysis in this section focuses on how different assumptions on Z translate into the potential for full-learning. Varying the commitment level z within Z affects the actions of the Doer: $z_t \in \{-1, 0, 1\}$ correspond to the full commitment and full flexibility choices from the simple example, while intermediate choices may induce the Doer to take particular actions in some but not all contingencies, as described below.

As before, the Doer is a myopic agent that does not account for how its actions affect the Planner's choices. The Doer's behavior is governed by its type θ drawn from $\Theta = [\theta_l, \theta_h]$ according to measure μ_0 that has a continuous and strictly positive density on Θ. The Doer observes the taste shock u_t, and chooses $a_t = 1$ if $W(u_t, \theta, z)$ is non-negative and $a_t = 0$ otherwise. The payoff function W is continuous and strictly increasing in each argument. Under full flexibility ($z = 0$), a Doer of type θ_l never undertakes the activity while type θ_h always undertakes the activity. Observe that this formulation permits uncertainty about the direction of temptation.[7] Full commitments override the choices of the Doer insofar as they determine the eventual action regardless of the taste shock and type: for all u, $W(u, \theta_l, 1) \geq 0 > W(u, \theta_h, -1)$; I interpret such full commitments as the choice of a singleton menu. A special case that the reader can keep in mind as a convenient example is $W(u, \theta, z) = u + 2\theta + 4z$, in which $\Theta = \mathcal{U} = [-0.5, 0.5]$. I denote a full information environment $\Gamma = (\mathcal{U}, F; W, \Theta)$ as a constellation of parameters that satisfy the above conditions.

After each period t, the Planner observes a_t (knowing his own choice z_t) and obtains some information about u_t. The setting of Section 2 considered an extreme case where the Planner learns nothing about u_t; one could equally consider the other extreme in which the Planner always observes u_t. Spanning these two extremes, one could analyze an intermediate environment—*imperfect attribution*—in which the Planner observes u_t with noise. More generally, suppose that the Planner does not observe u_t but obtains a signal $y_t(a_t, u_t)$ that comes from a space $Y = \mathcal{U}$, and is distributed according to $\tilde{F}(y|u, a)$.[8] For $t > 0$, $h^t = ((z_0, a_0, y_0), ..., (z_{t-1}, a_{t-1}, y_{t-1}))$ denotes the relevant history for the Planner in period t, and μ_t denotes the posterior beliefs of θ conditional on history h^t.

For much of the paper, I make the following assumption on observability.

[7] That temptation can exist in either direction is consistent with evidence from Kivetz and Simonson (2002) and Ameriks et al. (2007).

[8] Other papers within the Bayesian learning paradigm that consider coarse observation functions include Rubinstein and Wolinsky (1994), Dekel et al. (2004), and Esponda (2008).

Assumption 1. *$\tilde{F}(y|u,a)$ has full support on Y and has a density that is strictly positive and jointly continuous in (u,y).*

Assumption 1 simplifies the analysis although it is not necessary: it ensures that the support of signals is independent of the Doer's type, and therefore, a single signal cannot be perfectly revealing. The assumption plays a role in the clean characterization of when partial learning emerges in Theorem 2. Apart from its technical role, there are substantive reasons to be interested in imperfect attribution, as detailed in this footnote.[9] In Section 3.4.2, I illustrate what happens when Assumption 1 fails and instead, u_t is fully observable *ex post*. The main result holds in the absence of Assumption 1, but the characterization and proof are more intricate.

Before proceeding with the analysis, I make a number of remarks about the model.

Partial Commitment. I model a Planner's ability to partially commit as his influence over how the Doer behaves for each taste-shock. An interpretation of this commitment is that the Planner can use obstacles/instruments that nudge the Doer in a particular direction by making it more costly to take a particular action. These instruments include financial commitment contracts (like those offered by *stickK.com*), self-rationing,[10] the role of peer groups and promises,[11] and costly self-control. Partial commitment may also correspond to a habit-formation process whereby the Planner changes the Doer's preferences through the management of cues. Since the efficacy of any particular "therapy" is learned only through experience, the Planner necessarily engages in experimentation.

A different formulation for partial commitment is one in which actions come from a rich set, the Planner chooses a strict but non-singleton subset of actions, and the Doer makes choices from that smaller menu. While a general treatment of this setting is left for future research, I explore an application in this spirit in Section 4.1 when I study a canonical consumption-savings decisions in which the Doer chooses actions from a continuum and the Planner sets an upper-bound on possible consumption.

Relationship with Example. The model of Section 2 is a special case in which $u_t = b - c_t$, $\mathcal{U} = [b-1, b]$, $\Theta = \left\{\underline{\theta}, \bar{\theta}\right\}$, and $W(u, \theta, z) = u - (1-\theta)\, b + z$, where $z \in Z = \{-1, 0, 1\}$ in

[9] First, instead of observing the taste shock directly, the Doer may observe merely a noisy signal of it; equivalent results emerge in an imperfect monitoring model in which the Doer obtains a noisy signal of u_t, and the Planner observes u_t but not the signal of the Doer. Second, to the extent that past circumstances and taste shocks are representations of subjective states and emotions, it may be difficult for the Planner to obtain perfect information about past circumstances. Various studies highlight the difficulties individuals face in recalling past feelings and motives, and in separating situational factors from one's innate character. Baumeister et al. (1994) argue for the need for an individual to track her actions to minimize attribution errors, and Ameriks et al. (2003) describe it as a "skill," tracing its impact on wealth accumulation. Moreover, if taste-shocks are meant to encapsulate hedonic states, Kahneman et al. (1997) describe the difficulty individuals face in recall.

For these reasons, imperfect attribution has also been studied in related work, where the assumption is that a decisionmaker learns about his self-control using a "revealed preference" approach (Benabou and Tirole, 2004; Battaglini et al., 2005; Dal Bo and Tervio, 2008).

[10] See Wertenbroch (2003) for a discussion of the evidence on *self-rationing*, whereby individuals choose to not stock tempting objects so as to make their more expensive to obtain.

[11] See Battaglini et al. (2005) and Carrillo and Dewatripont (2008).

the absence of a bond contract, or $Z = [-1, 1]$ when there is a continuum of bond contracts to incentivize actions.

Bandit Structure. Since Z is any closed subset of $[-1, 1]$–including intervals, discontinuities, and isolated points—the Planner faces an experimentation environment similar to but richer than a standard multi-armed bandit. Accordingly, while I use insights from the bandit / experimentation framework, these results are distinct from those derived in prior work.

3.1 Adequate Vs. Partial Learning

The central issue analyzed is whether the Planner eventually attains the same payoffs that he would in the complete-information benchmark where the Doer's type is known. The complete-information payoff is defined as follows: for each θ and z in Z, $\pi(\theta, z) = E_u\left[u\mathbf{1}_{W(u,\theta,z)\geq 0}\right]$ corresponds to the static payoffs from commitment choice z when the Doer's type is θ. Denote the *complete-information payoff* when the Doer is of type θ by $\pi^*(\theta) = \max_{z\in Z}\pi(\theta, z)$ and let $z^*(\theta)$ be the smallest of the maximizers of $\pi(\theta, z)$. This payoff corresponds to the Planner's optimal balance between flexibility and commitment when the Planner is certain about the Doer's type.

When the Planner is uncertain about θ and has beliefs μ, he optimizes with respect to his beliefs. Denote the payoff in a single period from a commitment choice z by $m(\mu, z) = \int_\theta \pi(\theta, z)\, d\mu$; however, when the Planner is even slightly patient, he also values learning. Given a commitment choice z and prior μ, the Planner may revise his beliefs when observing the Doer undertake action a and signal y; $Q(\mu, z)$ denotes the probability distribution over posteriors induced by (μ, z), and $P(\Theta)$ denotes the set of Borel probability measures on Θ. The Planner solves the Bellman Equation,

$$V(\mu;\delta) = \max_{z\in Z}\left\{m(\mu, z) + \delta\int_{P(\Theta)} V(\tilde{\mu};\delta)\, dQ(\mu, z)\right\}. \qquad (2)$$

Conditional on a true θ, the Planner's beliefs, commitment choices, and therefore, value function evolve stochastically; the object of interest is their long-run distribution. We can partition the set of types into those for which learning is fully efficient, yielding the full information payoffs, from those for which learning may entail some long-run inefficiencies.

Definition 1. *Learning is* ***adequate*** *for a type θ if the Planner's payoffs when uncertain eventually converges to the complete-information benchmark:*

$$\Pr\left(\lim_{t\to\infty} V(\mu_t;\delta) = \pi^*(\theta)\,|\theta\right) = 1,$$

and otherwise, learning is ***partial*** *for type θ. Learning is* ***globally adequate*** *if the set of types for which learning is partial, Θ_P^δ, has μ_0-measure 0, and otherwise, learning is* ***inadequate****.*

Adequate learning is weaker than *complete learning* insofar as the eventual belief need not identify the type but the Planner nevertheless obtains the same payoffs as in the complete information benchmark. Globally adequate learning requires more than the mere existence of a strategy that ensures learning; indeed, it must be that such a strategy is optimal.

3.2 Distinguishability and Full Commitment Monotonicity

The possibility for globally adequate learning relies critically on the commitment set Z, and one can assess if Z is conducive to learning by examining how a fully informed Planner behaves. Before deriving the precise condition, I provide a heuristic explanation. For learning to be globally adequate, it must be the case that learning is free since were it costly at all, a Planner may stop learning once he is sufficiently confident. Since learning obtains whenever a Planner endows the Doer with partial flexibility (the empirical frequency of the Doer's choices identify its type), it is only when he chooses to completely constrain the Doer that learning is impeded. Therefore, for adequate learning to emerge, it is necessary that whenever the Planner chooses commitment levels z that force particular types θ to always choose $a = 1$, he can still distinguish it from those types for which he would wish to retain flexibility.[12] This property of *distinguishability* is fundamental to the possibility for globally adequate learning. Indeed, the consequence of its failure is apparent in Section 2 in which committing the weak type ($\underline{\theta}$) to select $a = 1$ precludes the strong type $\left(\bar{\theta}\right)$ from having any flexibility.

In the setting here, and in applications explored in Section 4, distinguishability naturally translates into a monotonicity condition. Since any commitment choice that forces a particular type to choose $a = 1$ also forces every higher type, adequate learning then fails whenever the Planner would prefer to provide flexibility to a high type but force a lower type to choose $a = 1$. To describe the condition precisely, let $\bar{z}(\theta) = \min\{z \in Z : W(\underline{u}, \theta, z) \geq 0\}$ denote the minimal level of commitment necessary to make type θ always undertake the action. The set $\hat{\Theta} = \{\theta \in \Theta : z^*(\theta) = \bar{z}(\theta)\}$ contains those Doer types that a fully informed Planner forces to choose $a = 1$ regardless of the taste-shock.

Definition 2. *The set of partial commitments, Z, satisfies Condition **FCM** (Full-Commitment Monotonicity) for an environment Γ if $\theta \in \hat{\Theta}$ implies that for all $\theta' > \theta$, $\theta' \in \hat{\Theta}$.*

Under FCM, the set of types for which the Planner chooses to provide the Doer with no flexibility must be an interval of the form $[\theta, \theta_h]$ if it is non-empty. A sufficient but not necessary condition for FCM is that a fully informed Planner never forces a type to always choose $a = 1$. A commitment set fails FCM for Θ if a fully informed Planner would wish to fully commit for some type but remain flexible for a higher type. Since the Planner's preference for commitment is defined by how the Doer maps taste-shocks into actions, whether an environment satisfies FCM is invariant to the representation of the Doer's preferences.[13]

[12] Since u has a strictly positive expectation, the Planner always prefers forcing the Doer to take $a = 1$ rather than forcing it to choose $a = -1$.

[13] Specifically, define two environments $\Gamma = (\mathcal{U}, F; W, \Theta)$ and $\Gamma' = (\mathcal{U}', F'; W', \Theta')$ to be *equivalent* if $(\mathcal{U}, F) = (\mathcal{U}', F')$ and there exists an isomorphism $\iota : \Theta \to \Theta'$ such that $W(u, \theta, z) \geq 0$ if and only if $W'(u, \iota(\theta), z) \geq 0$. Then it is straightforward to show that Γ satisfies FCM if and only if Γ' satisfies FCM.

Section 2 highlights this contrast: the basic commitment set fails FCM since committing type $\underline{\theta}$ entails providing no flexibility to type $\bar{\theta}$. On the other hand, once the bond commitment contract is introduced, the environment satisfies FCM since in that case, $\hat{\Theta}$ is empty.

FCM is distinct from standard notions of richness insofar as a commitment set may fail FCM while a strict subset satisfies it. The following examples illustrate this aspect of FCM: suppose that $W(u, \theta, z) = u + 2\theta + 4z$, and $\Theta = \mathcal{U} = \left[-\frac{1}{2}, \frac{1}{2}\right]$.

Example 1 (FCM with no commitment)**.** *Suppose $Z = [-1, 1]$: then a fully informed Planner can set $z^*(\theta) = -\frac{\theta}{2}$ to align every type's preference perfectly while retaining flexibility for each type. Thus, $\hat{\Theta} = \emptyset$ and FCM is trivially satisfied.*

Example 2 (FCM with commitment)**.** *Suppose $Z = \{-1\} \cup [0, 1]$: then a fully informed Planner sets $z^*(\theta) = \max\left\{-\frac{\theta}{2}, 0\right\}$. Observe that for any type θ above $\frac{1}{4}$, this choice of commitment induces that type to always choose $a = 1$ and thus, $\hat{\Theta} = \left[\frac{1}{4}, \frac{1}{2}\right]$. In this case, $\hat{\Theta}$ is non-empty but nevertheless satisfies FCM.*

Example 3 (FCM not preserved under expansion)**.** *Suppose that commitment level $\left\{-\frac{1}{4}\right\}$ is added to the previous Z: then the fully informed Planner would set $z^*\left(\frac{1}{4}\right) = 0$ to force a Doer of that type to always choose $a = 1$, while setting $z^*\left(\frac{1}{2}\right) = -\frac{1}{4}$ providing flexibility to type $\frac{1}{2}$.*

3.3 Main Results

The following result states the connection between FCM and learning.

Theorem 2.

1. *If the set of partial commitments satisfies FCM for Γ, then for every δ, learning is globally adequate.*

2. *If the set of partial commitments does not satisfy FCM for Γ, then for every δ, learning is inadequate. The set of types for which learning is partial is independent of δ and is $\Theta_P = \left\{\theta \in \Theta \backslash \hat{\Theta} : \theta > \theta' \text{ for some } \theta' \in \hat{\Theta}\right\}$.*

FCM is necessary and sufficient for globally adequate learning regardless of the Planner's patience or prior. When FCM is satisfied, even when the decisionmaker has slight patience, his eventual choices correspond to how he would behave were he fully sophisticated about his imperfect self-control. As proven in Lemma 9, eventual beliefs μ_∞ almost-surely place probability 1 on the true θ if θ is not in $\hat{\Theta}$, and otherwise put probability 1 on $\hat{\Theta}$. In both cases, the eventual commitment choice yields the same payoffs as those in the complete-information benchmark. Moreover, under FCM, globally adequate learning may obtain even if the decisionmaker is not sufficiently forward-looking to actively learn and experiment: a *passive learner* who myopically best-responds to his beliefs eventually behaves as if he perfectly understood his temptation. As such, the introduction of partial commitments that yield FCM can have strong implications for individuals regardless of their patience.

In contrast, when FCM fails, then so long as the Planner is not perfectly patient ($\delta = 1$), learning is partial. The set of types for which learning is partial is invariant to the Planner's patience or prior, though the probability of the failure of learning does depend on both. When the Planner's eventual choice diverges from the full information benchmark for some θ in Θ_P, it does so in a systematic way: the Planner chooses a commitment level z that exceeds $\bar{z}(\theta)$, and thereby, fully commits the type to $a = 1$ when partial commitment is optimal.

The characterization of Θ_P and its invariance to the discount factor and prior is an implication of Assumption 1. Failing Assumption 1, if the Planner can observe u_t perfectly *ex post*, then the partial learning set Θ_P^δ depends on δ and μ_0. However, the remainder of Theorem 2 holds as before: FCM remains necessary and sufficient for globally adequate learning. In Section 3.4.2, I prove the analogue of Theorem 2 when u_t is observed perfectly by the Planner after the Doer's choice.

From a methodological perspective, the above result informs us that only when FCM is satisfied does a steady-state analysis that assumes that the Planner is fully sophisticated have learning-based foundations. In environments in which it is satisfied, this result give us greater confidence in assuming sophistication for experienced individuals. At the same time, the result gives us cause to suspect that individuals may not be fully sophisticated in environments in which FCM may fail, and indicates that interventions may help even experienced individuals have a greater understanding of their temptations.

I omit a detailed description of how a failure of FCM leads to partial learning since its essence is contained in Section 2; a general proof is in the Appendix. To illustrate how FCM leads to globally adequate, consider Examples 1 and 2. Globally adequate learning fails only when the Doer is induced to always choose $a = 1$; in only that case is the Planner unable to identify the Doer's type from its empirical frequency. Observe that once the Planner's belief converges to any limit, his limiting action must be a static best-response to that limiting belief since all opportunities to learn are exhausted. Therefore, to demonstrate that globally adequate learning obtains, it suffices to demonstrate that under FCM, any limiting belief that rationalizes a limiting commitment choice must put positive probability only on those types for which such a choice is optimal.

Consider Example 1: in this case, an uncertain Planner has no reason to choose commitment levels outside $\left[-\frac{1}{4}, \frac{1}{4}\right]$. Suppose that globally adequate learning fails and that the Planner observes that the Doer always chooses $a = 1$ under some commitment level z. The Planner can then infer that θ is in $\left[\frac{1}{4} - 2z, \frac{1}{2}\right]$. A deviation to a commitment level of $z - \frac{1}{8}$ is strictly profitable: for a positive measure of types, this deviation strictly improves payoffs and for all other types, this deviation does not affect payoffs. Iterating this process, the Planner always decreases commitment until the Doer chooses both actions with positive probability. Then the Planner can eventually identify the Doer's type from the empirical frequency of its choices. In this example, that globally adequate learning obtains may not be surprising since $\hat{\Theta}$ is empty.

To that end, consider Example 2: here the Planner would wish to force every type above $\frac{1}{4}$ to always choose $a = 1$ but prefers relinquishing some flexibility to lower types. With commitment level z in $\left[0, \frac{1}{4}\right]$, if the Planner finds that the Doer always chooses $a = 1$, he can infer that θ is in $\left[\frac{1}{4} - 2z, \frac{1}{2}\right]$. Reducing the commitment level to $\max\left\{z - \frac{1}{8}, 0\right\}$ improves the Planner's static payoffs: for every type $\theta \geq \frac{1}{4}$, the payoffs remain unchanged, and for any type lower than that, his payoffs potentially improve strictly. Thus, once again, iterating the best-response involves decreasing the level of commitment until the Planner reaches the point at which he ascribes probability 1 to θ in $\left[\frac{1}{4}, \frac{1}{2}\right]$ or to some type θ less than $\frac{1}{4}$. In either case, his limiting action leads to identical payoffs as in the complete information benchmark.

The notion of FCM hinges on how different types in Θ respond to partial commitments chosen from Z. Yet, a designer of commitments—policymaker or firm—may not know Θ or W precisely. The result below demonstrates that only the full set of commitments ensures globally adequate learning for all possible environments that satisfy the above assumptions.

Theorem 3. *Learning is globally adequate for all environments* Γ *if and only if* $Z = [-1, 1]$.

3.4 Remarks

3.4.1 Identifying Awareness

The necessary and sufficient condition for adequate learning, FCM, is defined in terms of how a fully sophisticated Planner copes with the full range of possible temptations. As in the applications in Section 4, it is relatively straightforward to check distinguishability conditions in full-information models that we typically use; in contexts such as savings or addiction, our model of the temptation and commitment environment inherently reveal whether FCM would be satisfied. Accordingly, when specifying the full information model, a modeler often implicitly assumes FCM or its absence.

When a modeler or policymaker does not have a particular full-information model, checking for FCM may be challenging since how a decisionmaker copes with different levels of temptation is not directly observed. Nevertheless, within some settings, an individual's long-run preference for commitment can help identify his eventual awareness.[14] I illustrate this by considering the simple framework here, assuming that a policymaker can observe choices from the commitment set Z. Supposing that the Planner chooses commitment as studied above, his inferred preference ranking over commitment would lead the policymaker to conclude the following.

1. If a decisionmaker's most preferred commitment is some $z^* \in Z$, and this commitment is strictly preferred over every other, then the decisionmaker is partially committing and therefore, learns adequately in the limit. However, if the decisionmaker weakly prefers the greatest possible commitment ($z = 1$) over every other commitment choice, then he is fully committing and so he may or may not be sophisticated.

[14]That a preference for commitment can reveal awareness is described by Gul and Pesendorfer (2005); their setting is different, and their focus is not on questions of learning.

2. When the decisionmaker may or may not be sophisticated, the policymaker can learn about the decisionmaker's awareness by *exposing* him to different commitments. If temporary subsidies to commitments not in the most preferred set have long lasting implications on the decisionmaker's commitment choices, then learning was inadequate.[15]

3.4.2 Perfect Ex Post Observability

Assumption 1 plays an important role in Theorem 2: it ensures that any observation (z, a, y) that occurs with "positive probability" when the Doer is of type θ also occurs with "positive probability" when the Doer is a different type θ'.[16] Without this common support assumption, a single observation can lead the Planner to attribute probability 0 to certain types. In this section, I study learning without Assumption 1: specifically, I suppose after the Doer's choice, the Planner observes u_t perfectly in each period. I demonstrate that the equivalence between FCM and globally adequate learning holds in this environment, although the characterization of the types for which learning is partial depends on patience and the Planner's prior.

The following example highlights some of the issues involved.

Example 4. *Similar to Section 2, consider a 3 type model in which* $b = \frac{5}{8}$, $\Theta = \left\{\frac{1}{5}, \frac{1}{2}, \frac{4}{5}\right\}$, *and the Planner has a choice between full flexibility and commitment. In the full information benchmark, the Planner chooses to fully commit when the Doer is of type* $\frac{1}{5}$, *and otherwise chooses full flexibility. Therefore, FCM fails since the Planner commits for the lowest type but remains flexible for both higher types. If costs* c_t *were unobserved (as in Section 2), learning would be partial for both higher types. However, when costs are ex post observable, learning may be adequate for type* $\theta = \frac{4}{5}$ *but not necessarily for type* $\theta = \frac{1}{2}$. *The logic is as follows.*

Suppose the Planner has a uniform prior: his one-period static payoff from full flexibility exceeds that from commitment, and consequently, his initial commitment choice is full flexibility. If $c_0 \in \left[0, \frac{1}{8}\right]$, *then he learns nothing about the Doer since all 3 types choose* $a_0 = 1$ *in response to such costs. Similarly, if* $c_0 \in \left(\frac{1}{2}, 1\right]$, *all 3 types choose* $a_0 = 0$ *revealing no information. However, if* $c_0 \in \left(\frac{5}{16}, \frac{1}{2}\right]$, *then if he observes the Doer undertaking the activity, he learns that* $\theta = \frac{4}{5}$ *immediately since that would be the only type to choose that action. Finally, if* $c_0 \in \left(\frac{1}{8}, \frac{5}{16}\right]$, *then after seeing the Doer taking the costly action, the Planner believes that* $\theta \in \left\{\frac{1}{2}, \frac{4}{5}\right\}$. *Consequently, for all of the outcomes that occur with positive probability when* $\theta = \frac{4}{5}$, *the Planner always chooses to remain flexible. Almost-surely, he learns the Doer's type when* $\theta = \frac{4}{5}$. *At the same time, partial learning is still possible for* $\theta = \frac{1}{2}$: *there exists* $\bar{\delta} < 1$ *such that for* $\delta < \bar{\delta}$, *the Planner eventually chooses to commit with positive probability and never learns* θ.

[15] Indeed, as described in Section 4.2, temporary subsidies to drug rehabilitation are often believed to have effects after the subsidy has ended (Kaper et al., 2005, 2006). Given the experimentation challenge that individuals face in finding the appropriate therapy for their temptation, it is natural that learning is inadequate.

[16] Technically, any observation that is within the support of the induced distribution under type θ is also within the support of the induced distribution under type θ'. Since y draws values from a continuum, each particular observation has probability 0.

The example highlights the intricacies that emerge without Assumption 1. The following weaker result emerges in this environment.

Theorem 4.

1. *If the set of partial commitments satisfies FCM for* Γ*, then for every* $\delta > 0$*, learning is globally adequate.*

2. *If the set of partial commitments does not satisfy FCM for* Γ*, then for every* $\delta > 0$*, learning is inadequate. The set of types for which learning is partial depends on the Planner's patience,* δ*, and the Planner's prior,* μ_0*.*

The equivalence between FCM and globally adequate learning is nearly identical to that of Theorem 2. The sufficiency of FCM for globally adequate learning with perfect ex post observability follows from Theorem 2: learning strategies that are efficient when information is coarse continue to be efficient when information is precise. The necessity of FCM however requires a distinct and more intricate argument. Under perfect observability, a Bayesian Planner updates his prior by truncating the support after each observation; the argument in this context relies on the particular form that updating takes.

3.4.3 Relationship to Self-Confirming Equilibrium

The dynamic process of learning that I have studied has a natural relationship with the steady-state solution concepts of *self-confirming* equilibria (SCE), developed by Fudenberg and Levine (1993a), in which players' beliefs must be consistent with experiences only on the path of play.[17] Indeed, one way to view these results may be as a foundation for studying SCE in dynamic self-control environments.[18] However, despite the close resemblance to SCE, the dynamics studied herein are necessary to draw the distinction between FCM and its absence.

To illustrate, when FCM is satisfied and $\hat{\Theta}$ is non-empty, there exist self-confirming equilibria in which the Planner overcommits and partially learns: suppose that the Planner holds a belief μ that assigns probability 1 to $\hat{\Theta}$ and responds to these beliefs by choosing $z = 1$. Such a belief-action pair is a SCE regardless of the Doer's true type θ. In contrast to this SCE, when the Planner begins with a smooth prior in the dynamic model, he almost-surely does not ascribe probability 0 to the complement of $\hat{\Theta}$ if the Doer's true type is not in $\hat{\Theta}$. This gives the Planner

[17] Self-confirming equilibria have been studied and developed further in Dekel et al. (1999, 2004), and applied to a behavioral context by Esponda (2008). A summary is offered by Fudenberg and Levine (2009a).

[18] The appropriate analogue is a static Bayesian game between the Planner and Doer in which each has the payoff functions specified earlier, and the Planner knows how different types of the Doer behave. A Nash equilibrium in this game corresponds to the Planner putting probability 1 on the true Doer type and best-responding to it. A self-confirming equilibrium corresponds to the Planner having a possibly non-degenerate belief μ of the Doer's type, the Planner takes a best-response z, and the Planner's information feedback does not cause him to update the belief μ.

a strict incentive to choose lower commitment levels, and such choices lead to adequate learning. As such, the limiting choices of the dynamic model refines the set of SCE.[19]

Moreover, while it is known that self-confirming equilibria correspond to the results of myopic experimentation (Fudenberg and Levine, 1993b), the steady-state model does not inform what restrictions (if any) are necessary for the Planner's patience to yield partial learning in the absence of FCM. In contrast, the dynamic framework reveals that inadequate learning transpires for all discount factors, and if Assumption 1 is satisfied, permits one to characterize the set of types for which learning may be partial independently of the Planner's patience.

4 Applications

4.1 Consumption-Savings

Individuals are often tempted to overconsume and sometimes demand commitment to curb this temptation; accordingly, imperfect self-control has been studied in many papers on consumption and savings.[20] As demonstrated by prior work, access to an illiquid asset endows the decisionmaker with a powerful instrument for partial commitment.[21] I show that such partial commitments yield full learning for a decisionmaker who is uncertain about his tendency to overconsume. In particular, applying the framework that I developed in Section 3, I show that an analogue of FCM is automatically satisfied in this environment thereby allowing us to leverage those positive results in this more complex setting.

The simple savings environment that I study is a decisionmaker who faces a tradeoff between commitment and flexibility in each period. His expected payoff from a consumption stream $\{c_t\}_{t=0}^{\infty}$ is $E\left[\sum_{t=0}^{\infty}\delta^t u_t U\left(c_t\right)\right]$, where for simplicity, U is a CRRA utility function with a coefficient of relative risk aversion $\sigma > 0$. The DM is uncertain about the value of consumption in each period, represented by the i.i.d. taste shock u_t, which is drawn continuously from $[\underline{u}, \bar{u}]$ with cdf F and where $0 < \underline{u} < \bar{u}$ and $E[u] = 1$. The assumption of CRRA utility allows me to restrict attention to consumption paths that are linear in wealth.

Consumption choices are made as follows: each period is divided into two sub-periods. Wealth at the beginning of the period is denoted by y_t which can be invested by the Planner in illiquid assets or kept in a liquid form in the first sub-period. The Planner selects a minimal savings rate, s_t, by investing an amount $s_t y_t$ in illiquid assets. In the second sub-period, the Doer selects consumption c_t in the second sub-period that is feasible—$c_t \leq (1 - s_t)\, y_t$—and invests the remainder. As before, the Planner is the long-run decisionmaker whereas the Doer is a short-run behavioral type who simply over-consumes relative to the Planner's preferences.

[19] The exact issue is a failure of lower hemi-continuity in the Planner's best-response $z(\mu)$: consider a sequence of beliefs $\{\mu_n\}_{n=1}^{\infty}$ that converges (in the weak topology) to μ_∞, a measure with support $\hat{\Theta}$. Suppose further that for each n, $\hat{\Theta}$ is a strict subset of $Supp\,(\mu_n)$. Then $z = 1$ is one of a continuum of exact best-responses for belief μ_∞ but is not an exact best-response for any belief μ_n.

[20] Harris and Laibson (2003) survey this burgeoning literature. Closely related recent papers are Amador et al. (2006), Brocas and Carrillo (2008), and Fudenberg and Levine (2006, 2009b).

[21] In particular, see Laibson (1997), Laibson et al. (1998), and Amador et al. (2006).

Formally, let $c(u_t)$ denote the Planner's optimal policy if he could contract perfectly on the taste shock u_t and had wealth 1.[22] Relative to this, a Doer of type θ would like to consume $\theta c(u_t)$ at the same wealth level, and selects the feasible consumption closest to this ideal. Doer types θ are in $\left[1, \bar{\theta}\right]$ drawn according to μ_0.

I make two assumptions on the capital markets: first, that the Planner has access to perfect capital markets at the investment stage so the initial wealth y_0 includes all of the consumer's future income. Second, I assume that returns on savings occur at gross rate R regardless of whether these savings were made by the Planner or Doer. Accordingly, the wealth in period t is generated from the returns from saving in the prior period: $y_t = R(y_{t-1} - c_{t-1})$. Both of these assumptions serve to simplify the exposition but are not critical to the analysis. Finally, to ensure that the transversality condition is satisfied, I assume that $\delta R^{1-\sigma} < 1$.

Since my goal is to establish the inevitability of adequate learning, I make the most stark of informational assumptions: the Planner observes only past consumption and not past taste-shocks. Since learning emerges here, it continues do so in less austere observational environments. Unlike the analysis of Section 3, commitment consists of specifying menus that constrain the Doer's choice rather than affecting the Doer's payoffs.[23] The specification of a consumption-cap induces separating and bunching regions and affects the Planner's inferences: for choices that are strictly below the consumption-cap, the Planner recognizes that the consumption corresponds to the Doer's ideal point (given the taste-shock). In contrast, when consumption is at the cap, then the Planner recognizes that with probability 1, the cap was binding, and therefore, the Planner has less ability to distinguish the Doer's ideal point. Since the Planner's optimal cap depends on his beliefs and wealth, solving the model with partial awareness is a challenging exercise.

Nevertheless, one can characterize asymptotic behavior using simply the full-information benchmark and tools from Section 3. Lemma 12 in the Appendix demonstrates that for each type θ, a fully informed Planner has a stationary solution, which consists of purchasing the same fraction of illiquid assets, $s(\theta)$, in each period. This induces a consumption-cap and bunching for high taste shocks but retains flexibility for the Doer to choose lower consumption for lower taste shocks. If the maximal *consumption rate*, $1 - s(\theta)$, is less than the Doer's ideal consumption rate for all taste-shocks, then the consumption-cap is binding for all taste-shocks. Solving for the the optimal binding consumption path yields that for all θ, $s(\theta)$ takes a maximum value of $s^* = \delta^{1/\sigma} R^{(1-\sigma)/\sigma}$.

Based on the optimal binding consumption path, one can then partition Θ into those types for which full commitment is optimal and those types for which some flexibility is retained. Consider those types θ' that consume more than $(1 - s^*)$ even with the lowest possible taste-shock $\underline{u}$. For such a type, it is optimal to set $s(\theta)$ to equal s^*, and thereby commit these

[22] In the Appendix, I show that $c^*(u) = \frac{u^{1/\sigma} C}{u^{1/\sigma} C + \delta^{1/\sigma} R^{(1-\sigma)/\sigma}}$, where C is the solution to $E\left[\left(C u^{1/\sigma} + \delta^{1/\sigma} R^{(1-\sigma)/\sigma}\right)^{\sigma}\right] = 1$.

[23] As such, the notion of commitment also differs from that in Laibson (1997) and Barro (1999), who abstract from the value of flexibility, and consider a decisionmaker's ability to commit to a particular consumption path.

types to consume $(1-s^*)\,y_t$ regardless of the taste-shock u_t. In contrast, for a type θ for which $\theta c\,(\underline{u}) < 1 - s^*$, the Planner purchases a smaller fraction of illiquid assets, and thereby retains greater flexibility for the Doer. Accordingly, the only types that are fully committed to a particular consumption path are those in $\hat{\Theta} = \{\theta \in \Theta : \theta c\,(\underline{u}) \geq 1 - s^*\}$, and all other types are offered some partial flexibility. Figure 1 below illustrates these consumption-caps: θ is a type that is accorded partial flexibility, whereas type θ' is accorded no flexibility since $\theta' c\,(\underline{u})$ exceeds s^*.

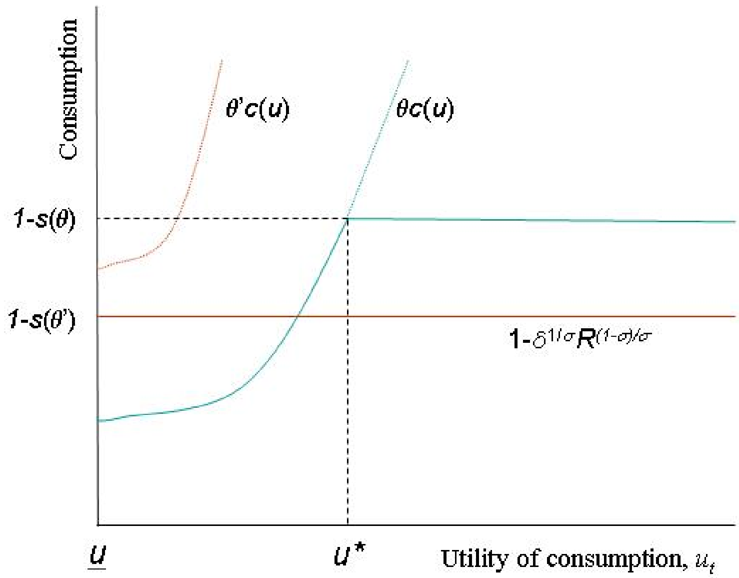

Figure 1

The above figure illustrates consumption caps when the Planner knows the Doer's type. Solid lines indicate the Doer's consumption for each taste-shock and the dotted curve is the Doer's ideal consumption.

The appropriate analogue of FCM and distinguishability are thus naturally satisfied in this setting: when the Planner prefers to fully commit a type, he also prefers to fully commit all higher types. Moreover, when the Planner sets a consumption-cap of $(1-s^*)$, any type $\theta \notin \hat{\Theta}$ is bound by this consumption-cap with probability strictly less than 1, as can be seen in the above figure. In the figure, type θ' in $\hat{\Theta}$ makes choices at the consumption-cap $(1-s^*)$ with probability 1, the type $\theta \notin \hat{\Theta}$ chooses below that consumption cap whenever u_t is less than u^*. Thus, even when selecting this commitment path, the Planner can distinguish types in $\hat{\Theta}$ (for which such full-commitment is optimal) from types not in $\hat{\Theta}$ (for which the Planner retains some flexibility for the Doer). In either case, learning is adequate.

Theorem 5. *Learning is globally adequate in this consumption-savings framework.*

The possibility for globally adequate learning implies that eventually, the decisionmaker sets commitment choices in a fully sophisticated manner. Thus, the assumption that decision-makers are sophisticated about their tendency to overconsume has a strong learning-theoretic justification. Of course, while learning, the Planner may make (*ex post*) inefficient choices that adversely affect his future income distribution relative to the full information benchmark, but biases in the Planner's beliefs eventually disappear. Partial commitments are indeed necessary for this positive learning: if the Planner faced a choice between full flexibility and full commitment to consuming a particular fraction of income in each period, the possibility for persistent biases would re-emerge.

The full information model to which behavior eventually converges shares many predictions of sophistication in the quasi-hyperbolic savings model—such as inducing a demand for commitment and the tendency to undersave relative to a normative benchmark—but is analytically distinct. The full information model analyzed here has a unique solution and a constant commitment rate with CRRA utility (similar to Fudenberg and Levine 2006, 2009b). In contrast, the quasi-hyperbolic savings model in discrete-time has multiple sophisticated solutions (Krusell and Smith, 2003; Harris and Laibson, 2003) and thus, learning-theoretic foundations for sophistication in the quasi-hyperbolic model would have the additional challenge of justifying the equilibrium selection.

4.2 Addiction

Several studies on addiction have highlighted how individuals respond in different ways to drug use, depending on social, psychological, and genetic factors. Wagner and Anthony (2002) find that a mere 16% of habitual cocaine users actually become addicted during their first 10 years, and O'Brien and Anthony (2005) estimate that only 6% become addicted during their first $1-2$ years. In light of this substantial heterogeneity, individuals may not know how they would respond to addictive substances, and learn about this only through experience. I explore the possibilities for learning by building on the model of Bernheim and Rangel (2004).[24]

So as to save space, I relegate the formal details of this model to the Appendix and informally describe what unfolds as the decisionmaker learns. So as to illustrate different mechanisms for partial learning as transparently as possible, I make stark assumptions. Throughout, a type θ_G is less susceptible to addiction and a type θ_B is more susceptible.

Example 5 (Overavoidance)**.** *Suppose that if the DM were of type θ_G, a fully-informed DM would leisurely consume drugs but that for type θ_B, a fully-informed DM is easily addicted and so is best off avoiding drugs altogether. Suppose that avoidance of cues can eliminate the possibility for cue-conditioned consumption altogether. Then a partially aware DM may experiment with drugs, and be triggered sufficiently often that he decides that he is too easily susceptible to addiction and should perpetually avoids drugs. Such commitment may emerge even if the DM's could leisurely consume substances.*

The possibility for overavoidance corresponds closely to the overcommitment result obtained in earlier sections. That individuals may avoid drugs when partially aware is intuitive: while

[24] Orphanides and Zervos (1995) and Wang (2007) also focus on learning and experimentation but in the context of the rational addiction framework. While experimenting with drugs can induce *ex post* regret, long-run learning is complete in their framework.

there is a chance that one may not become addicted, the risks may simply be too great to induce any experimentation. Therefore, beliefs that exaggerate these risks may persist. Consistent with this case, Lander (1990) finds that even trained medical practitioners overestimate the likelihood with which patients become addicted to pain management medication, and choose to withhold such medication for fear of inducing addiction. Viscusi (1990) discusses how individuals overestimate the dangers of smoking.[25]

At the same time, partial awareness and learning can induce over-consumption of drugs: suppose a DM is uncertain about the efficacy of rehabilitation.

Example 6 (Insufficient Rehabilitation)**.** *Suppose that for type θ_B, costly rehabilitation is ineffective insofar as he returns to a high addictive state soon thereafter with high probability, but that for type θ_G, rehabilitation is optimal. A partially aware DM who finds himself in state S will only experiment with rehabilitation for finitely many periods before (incorrectly) concluding that he becomes too easily addicted after rehabilitation to justify its expense.*

That the cost of rehabilitation and therapy may prevent individuals from experimenting sufficiently often to learn about its efficacy is intuitive. Consistent with this possibility, Kaper et al. (2005, 2006) find that subsidizing smoking cessation therapies indeed impacts long-run abstinence even after the subsidization program ended. One reasonable interpretation is that individual addicts are unsure about which therapy may successfully help them quit smoking and learn this through the subsidization program.[26]

Similar to the challenge of insufficient rehabilitation, the following highlights the challenge of zero-tolerance beliefs, described by Baumeister et al. (1994) and others: the belief that one's addictive tendencies will snowball becomes a self-fulfilling prophecy.

Example 7 (Zero-tolerance beliefs)**.** *Suppose that a decisionmaker is persuaded to (falsely) believe that an initial lapse will cause one to have little future self-control. Then, a former addict who avoids drugs, but experiences a single lapse believes that there is no value to avoiding drugs in higher addictive states. Accordingly, such a DM, after a single lapse, will expose himself in every subsequent period (although avoidance may have been optimal in a full-information setting).*

When an individual holds zero-tolerance beliefs, he anticipates a slippery slope following an initial lapse. Thus, such an individual shows little restraint: he exposes himself to drugs and never learns how much self-control he may have possessed had he avoided them.

[25] Carrillo and Mariotti (2000) discuss a different source for such biased beliefs: namely that a decisionmaker may strategically choose to believe that such temptations are dangerous to induce commitment.

[26] In a similar vein, but in the context of fitness, Charness and Gneezy (2009) find that a temporary incentive to exercise has strong post-intervention effects. While they interpret this in support of habit-formation, a complementary explanation is that some individuals may have been sufficiently underconfident about their future self-control (and future attendance) that exercise did not appear valuable in absence of incentives. However, during the incentive period, these individuals learned that they possessed sufficient self-control to exercise regularly, and this confidence in the future motivates them to exercise.

4.3 Costly Self-Control

Fudenberg and Levine (2006, 2009b) offer a simple but powerful Planner-Doer framework in which the Doer's preferences are completely characterized by the present and the Planner exerts costly self-control so as to induce the Doer to take actions more consistent with the Planner's preferences. Here, I analyze how a Planner who is uncertain about the cost or efficacy of self-control learns over time how much self-control to use, building on the binary action environment from Section 2 and Fudenberg and Levine (2006, Section 4).

Consider a decisionmaker who can completes an infinite sequence of tasks. Completing a task in any period ($a_t = 1$) leads to a *future* benefit with a present value of b. Not completing it ($a_t = -1$) has the benefit of leisure c_t, where c_t is uniform drawn from $[0, 1]$. I assume that b is in $\left(\frac{1}{2}, 1\right)$, generating a tradeoff between commitment and flexibility.

Prior to the realization of c_t, the Planner chooses a non-empty menu from $\{-1, 1\}$. After the realization of the benefit of leisure, the Planner can exert costly self-control if it is necessary to influence the Doer's choice. When the Planner exerts self-control of magnitude $\Gamma \geq 0$, self-control is *successful* for a Doer of type γ if Γ exceeds γc_t; otherwise, it *fails*. The Planner's marginal payoff from completing a task at date t is $b - \Gamma$ and the marginal payoff from not completing the task is $c_t - \Gamma$. I assume that γ is in an interval $\left[\underline{\gamma}, \bar{\gamma}\right]$.

The full-information model, in which the Planner knows γ, corresponds to that considered by Fudenberg and Levine (2006). If γ is sufficiently low, then the Planner chooses the menu $\{-1, 1\}$ and has an associated cutoff c^* such that he exerts costly self-control γc_t if and only if c_t is less than c^*. Such a choice has the benefit of flexibility, since the Planner can exert costly self-control depending on the taste-shock. In contrast, when γ is sufficiently high, the Planner may prefer to commit through the singleton menu; while this relinquishes flexibility, it obviates costly self-control. The critical value of γ that determines whether the Planner prefers flexibility or commitment is $\gamma^* = \frac{(1-b)^2}{2b-1}$.

When the Planner is uncertain about γ, he may use too much or too little self-control to influence the Doer. Employing too little self-control leads the agent to not complete the task while the costs of self-control are still borne. In contrast, too much self-control guarantees that the task is completed, but at unnecessary expense. Over time, the Planner learns about the efficacy of self-control based on the past history. As before, it is possible to assess eventual beliefs and choices simply by studying the full-information model.

Theorem 6. *If $\bar{\gamma}$ is at most γ^*, then FCM is satisfied: a fully-informed Planner always chooses a flexible menu and then to exercise self-control. Thus, learning is globally adequate. In contrast, if $\bar{\gamma}$ exceeds γ^*, then FCM is violated: a fully-informed Planner may choose to commit for sufficiently high values of γ but would prefer flexibility for lower values of γ. Thus, learning is inadequate.*

The above result highlights when a decisionmaker learns to use self-control and commitment appropriately. When flexibility and self-control are always preferable to commitment, learning is fully efficient, whereas when commitment may be preferable, then long-run inefficiencies may persist in the limit.

5 Discussion

This paper offers a simple and tractable non-equilibrium approach to study how beliefs, self-control, and commitment choices jointly evolve. I derive a condition on full information settings that can reveal exactly when sophistication has strong learning-based foundations. When FCM is satisfied, even a passive learner eventually makes choices as if he were in the full-information benchmark. As such, learning-based foundations for sophistication in these settings, such as a savings environment, can help address the criticism that it involves an unrealistic level of rationality.[27] In contrast, when FCM fails, then learning fails for a decisionmaker who is not perfectly patient. In these settings, an assumption of sophistication precludes certain behavioral predictions and policy interventions can facilitate greater awareness.

Generating a non-equilibrium approach to study evolving self-awareness requires the choice of a modeling framework, and there is little reason at this stage to think that the approach here is the best possible way to do so. Nevertheless, it provides a useful starting point to study a number of different canonical applications and emerge with some general insights. One aspect of a dual-selves environment that makes it particularly attractive for such a purpose is that such models are analytically simple and generally have unique solutions when a decisionmaker is sophisticated, as emphasized in prior work. Accordingly, the task of checking whether the *unique* solution under partial awareness converges to the *unique* solution under sophistication is straightforward. In contrast, intrapersonal games even under exogenous beliefs often have multiple solutions and are challenging to solve. These difficulties are exacerbated by the introduction of partial awareness and learning since these inevitably raise questions of self-signaling in an intrapersonal game.

When learning fails at the limit, I find that it occurs in a particular direction, namely that individuals underestimate their self-control. It is important to understand the extent to which these results can be reconciled with evidence from the field. Measuring self-awareness is difficult, not least because individual choices are driven by many features—limited memory or attention, signaling considerations, temptations in the choice of menu—absent in standard self-control models (including this one). Some of the evidence on self-awareness comes from individuals' contractual choices and their subsequent decisionmaking. DellaVigna and Malmendier (2006) demonstrate that many individuals choose monthly gym memberships but given their usage of

[27] Rubinstein (2005) offers such a criticism:

> "Sophistication is unrealistic since it suffers from the problems of subgame perfection. An agent is super-rational in the sense that he perfectly anticipates his future selves and arrives at equilibrium between them. Present-bias is a realistic phenomenon, but the combination of the β, δ preferences with naivete or sophistication assumptions makes the model even more unrealistic than time consistency models."

the gym, would have paid less per visit had they selected the pay-per-visit option. Similarly, Shui and Ausubel (2004) find that individuals accept introductory credit card offers with lower interest rates for a shorter duration rather than a higher interest rate with a longer duration though they may be better off with the latter choice. In both cases, behavior is consistent with individuals being sophisticated about their self-control—valuing the commitment offered by a monthly membership or a shorter duration period—as well as them overestimating their self-control. In a model with overlapping generations, it is quite likely that both effects are present insofar as less experienced customers who have not had opportunities to learn may overestimate self-control. DellaVigna and Malmendier (2006) argue that delays in cancelling contracts are more consistent with partial naivete. However, even when a decisionmaker does not overestimate self-control, he has an incentive to delay canceling since there is an "option value" from learning, particularly for the first-time customers that comprise this sample. Such individuals may be learning both about their self-control and the costs and benefits of going to a health club (as also argued by Fudenberg and Levine 2006).

Indeed, in such a setting, one can show that learning about self-control can be particularly slow and cause someone to undercommit for a long period of time. To highlight this mechanism simply, consider the two-type model from Section 2, and suppose for simplicity that $\delta = 0$. Instead of the distribution of costs being known by the Planner, suppose that it depends on the realization of a random variable ξ; in particular, let c_t be distributed uniformly with support $[\xi, 1+\xi]$, and suppose that ξ itself is 0 with probability $1-\lambda_0$ and $\left(\bar{\theta}-\underline{\theta}\right)b$ with probability λ_0. This particular support is made for simplifying reasons and is unnecessary for the analysis.[28] Over time, it would be natural for the Planner to learn about the disutility of taking the action, but to make the point transparently, suppose that information about past costs does not emerge for a long time. Within this context, a Planner who observes the Doer choose $a = 1$ with frequency $\underline{\theta}b$ cannot distinguish his lack of self-control from a strong disutility of taking the action. Prior beliefs then determines his choice to remain flexible or commit (in the long-run limit), and under certain conditions, such a Planner may perpetually undercommit relative to a full information benchmark.

Proposition 2. *If μ_0 and λ_0 are sufficiently high, and $(\underline{\theta}, 0)$ is the realized state, then a partially aware Planner always chooses flexibility. If, however, a sophisticated Planner knew that $\xi = 0$ or that $\theta = \underline{\theta}$, then almost-surely, the Planner would choose to commit.*

The assumption that a Planner never observes any ex post information about payoffs is critical:[29] when the Planner obtains informative signals about payoffs, this sequence of signals identify ξ in the long-run limit, eliminating the possibility for undercommitment. Such an informational restriction is implausible in the long-run but is likely to be true in the short and medium run. As such, in a health-club setting, the framework here predicts that results

[28] Identical results hold if ξ varies along a continuum and has values in its support that are $\left(\bar{\theta}-\underline{\theta}\right)b$ apart.

[29] As such, the issue is similar to the cases studied by Dekel et al. (2004) and Esponda (2008) in which payoffs are not observed, and to the identification challenges in Acemoglu et al. (2009).

consistent with overestimating self-control should persist over a long horizon when the Planner learns about the utility or disutility of actions slowly.

Finally, the nature of long-run beliefs does not imply that a decisionmaker should never undercommit. Commitment often has costs apart from financial or its loss of flexibility; in many contexts, such as addiction, the choice to commit is observable, and so a decisionmaker who recognizes his imperfect self-control may nevertheless choose not to commit so as to hide it from others. Moreover, to the extent that commitment requires self-control (as argued by Noor 2007), even a fully sophisticated decisionmaker may not commit. If a decisionmaker is tempted by choice of menu, his behavior may appear rationalized by partial naivete even when he has a perfect understanding of his self-control. As such, extensions to this framework to allow for these considerations arise can explain why individuals may not commit to the extent that they should, even when they learn about their temptations over time.

In spite of all the potential reasons for people not to commit, many recent studies highlight the demand for commitment that individuals exhibit, as surveyed in DellaVigna (2009). Consistent with a demand for commitment, Prelec and Loewenstein (1998), Huffman and Barenstein (2005), and Cadena and Keys (2008) find that individuals are averse to debt, even when the debt is interest-free. Baumeister et al. (1994) interpret obsessions, compulsions, and rigidity as a response to beliefs that one lacks control otherwise. Such a demand for commitment is consistent with both sophistication and underestimating self-control.

I have focused on only one channel for uncertainty to interact with commitment choices, namely that of costly experimentation. Naturally, my work builds on the rich literature in experimentation—especially Rothschild (1974), Easley and Kiefer (1988), and Aghion et al. (1991)—although to my knowledge, the relationship between FCM and adequate learning derived here is unique to this setting. Prior work on self-control and learning has studied issues that arise when information affect one's temptation;[30] such issues are complementary to my analysis since information helps the Planner make commitment choices but has no other effect on the Doer's actions. I view this paper as a starting point to understand how an individual learns how to exert self-control, and there are many directions for future research.

A Appendix

A.1 Section 2

I defer the proof of Proposition 1 to the Supplementary Appendix because it follows standard techniques.

[30]For example, Benabou and Tirole (2004) examine the possibility for a self-reputation mechanism to induce internal commitment and overregulation, in which temptation is mitigated by the adverse reputational effect it induces in future incarnations. That partial awareness can induce overregulation is a common feature of both our models. Bodner and Prelec (2003) and Dal Bo and Tervio (2008) also consider self-signaling games in which an individual cares intrinsically about the type of person that she is, and thereby chooses actions that signal her goodness and hide her true type. Carrillo and Mariotti (2000) explores how a decisionmaker may decide to stop learning, even when it is costless, so as to mitigate future commitment problems.

Proof of Theorem 1 on p. 8

Define the public likelihood ratio, $\lambda\left(h^t\right) \equiv \frac{\Pr\left(\theta=\bar{\theta}|h^t\right)}{\Pr\left(\theta=\underline{\theta}|h^t\right)}$, and denote the random variable which has realizations $\lambda\left(h^t\right)$ by λ_t. Let $\lambda^* = \frac{\mu^*}{1-\mu^*}$. Consider the case where $\theta = \underline{\theta}$. Observe that $\langle\lambda_t\rangle$ is a super-martingale:

$$E_t\left[\lambda_{t+1}|\theta=\underline{\theta}\right] = \begin{cases} \lambda_t & \text{if } \underline{\theta} > 0 \\ \left(1-\bar{\theta}b\right)\lambda_t & \text{if } \underline{\theta} = 0 \end{cases}$$

By Doob's Supermartingale Convergence Theorem for non-negative random variables (Billingsley, 1995, p. 468-469), there exists a non-negative random variable λ_∞ with support $[0,\infty)$ such that $\lambda_t \rightarrow_{a.s.} \lambda_\infty$.

Lemma 1. *Supp* $(\lambda_\infty) \subseteq [0, \lambda^*)$

Proof. Suppose that for some fixed $x \in (\lambda^*, \infty)$ that $x \in Supp\,(\lambda_\infty)$. Consider the ball $B_\varepsilon(x) = (x-\varepsilon, x+\varepsilon)$ where $\varepsilon < \min\left\{\left(\frac{\bar{\theta}-\underline{\theta}}{\bar{\theta}+\underline{\theta}}\right)x, \left(\frac{b\left(\bar{\theta}-\underline{\theta}\right)}{2-b\left(\bar{\theta}+\underline{\theta}\right)}\right)x, x-\lambda^*\right\}$. Since x is in the support of λ_∞, then λ_i is eventually in $B_\varepsilon(x)$ with positive probability. Observe that for all $\tilde{x} \in B_\varepsilon(x)$, when $\lambda_i = \tilde{x}$, the Planner chooses $z = 0$ since $\tilde{x} > \lambda^*$. The Planner then observes either the activity not being undertaken or the activity being undertaken. Consequently, $\lambda_{i+1} \in \left\{\left(\frac{1-\bar{\theta}b}{1-\underline{\theta}b}\right)\tilde{x}, \left(\frac{\bar{\theta}}{\underline{\theta}}\right)\tilde{x}\right\}$ if $\underline{\theta} \neq 0$ and $\lambda_{i+1} = \left(1-\bar{\theta}b\right)\tilde{x}$ if $\underline{\theta} = 0$; neither value is in $B_\varepsilon(x)$. Therefore, $\lambda_i \in B_\varepsilon(x) \Rightarrow \lambda_{i+1} \notin B_\varepsilon(x)$; thus, $x \notin Supp\,(\lambda_\infty)$, leading to a contradiction.

To prove $\lambda^* \notin Supp\,(\lambda_\infty)$: suppose the opposite. It must be that for all ε, λ_i is eventually in $B_\varepsilon(\lambda^*)$ with positive probability. Let $\varepsilon < \min\left\{\lambda^*\left(\frac{\bar{\theta}-\underline{\theta}}{\bar{\theta}+\underline{\theta}}\right), \lambda^*\left(\frac{b\left(\bar{\theta}-\underline{\theta}\right)}{2-b\left(\bar{\theta}+\underline{\theta}\right)}\right)\right\}$, and consider the set $B = [\lambda^*, \lambda^*+\varepsilon]$. Observe that along any sample-path, if $\lambda_i < \lambda^*$, then for every $k > i$, $\lambda_k = \lambda_i < \lambda^*$. Consequently, the only sample-paths along which $\langle\lambda_i\rangle$ converges to λ^* are those in which $\lambda_i \geq \lambda^*$ for every i, and $\lim_{i\to\infty}\lambda_i = \lambda^*$. For $\tilde{x}$ in B, if $\lambda_i = \tilde{x}$, the Planner must choose $z = 0$ since $\tilde{x} \geq \lambda^*$. Then, as before, $\lambda_{i+1} \in \left\{\tilde{x}\left(\frac{1-\bar{\theta}b}{1-\underline{\theta}b}\right), \tilde{x}\left(\frac{\bar{\theta}}{\underline{\theta}}\right)\right\}$, neither of which is in B. Q.E.D.

The above establishes that $\Pr\left(\lim_{t\to\infty}\lambda_t \in [0,\lambda^*)\,|\theta=\underline{\theta}\right) = 1$. Since the Planner chooses $z = 1$ for beliefs in that range, the first part of the result is established.

Consider $\rho\left(h^t\right) = \frac{\Pr\left(\theta=\underline{\theta}|h^t\right)}{\Pr\left(\theta=\bar{\theta}|h^t\right)}$, and let ρ_t be the random variable which has realization $\rho\left(h^t\right)$. Let $\rho^* = \frac{1-\mu^*}{\mu^*}$. Consider the case where $\theta = \bar{\theta}$ and observe that $\langle\rho_t\rangle$ is a martingale. By the Martingale Convergence Theorem, there exists a random variable ρ_∞ with $Supp\,(\rho_\infty) \subseteq [0,\infty)$ such that $\rho_t \rightarrow_{a.s.} \rho_\infty$.

Lemma 2. *Supp* $(\rho_\infty) \subseteq \{0\} \cup (\rho^*, \infty)$

Proof. The proof is analogous to that of Lemma 1. Q.E.D.

The above Lemma establishes that conditional on $\theta = \bar{\theta}$, $\mu_\infty \in \{1\} \cup [0, \mu^*)$ with probability 1. Moreover, observe that for a given μ_0 and $\rho_0 = \frac{1-\mu_0}{\mu_0}$, $\Pr\left(\lim_{t\to\infty} z_t = 1|\theta=\bar{\theta}\right) \geq$

$\left(1-\bar{\theta}b\right)^{\left\lceil \frac{\log p^{*}-\log\rho_{0}}{\log(1-\underline{\theta}b)-\log(1-\bar{\theta}b)}\right\rceil} > 0$. To see that $\Pr\left(\lim_{t\to\infty} z_t = 0|\theta = \bar{\theta}\right) > 0$: observe that since $\langle\rho_t\rangle$ is a martingale, $E\left[\rho_t\right] = \rho_0 \leq \rho^*$ for every t where the inequality follows from $\mu_0 \geq \mu^*$. By Fatou's Lemma, $E\left[\rho_\infty\right] \leq \lim_{t\to\infty} E\left[\rho_t\right] \leq \rho^*$, and as $(0, \rho^*] \cap Supp\left(\rho_\infty\right)$ is non-empty, $\Pr\left(\rho_\infty = 0\right) > 0$. Therefore $\Pr\left(\lim_{t\to\infty} z_t = 0|\theta = \bar{\theta}\right) > 0$.

A.2 Section 3

Some Preliminaries: Before proceeding to the theorems in the text, I highlight some properties of the complete-information solution. Recall that $\pi^*(\theta) = \max_{z\in Z}\pi(\theta, z)$ and $z^*(\theta)$ is the smallest maximizer of $\pi(\theta, z)$. Let $u(\theta, z)$ be the lowest value in $U_{\theta,z} = \{u \in \mathcal{U} : W(u, \theta, z) \geq 0\}$. For a set $A \subset \Theta$, let $\inf(A)$ and $\sup(A)$ denote the infimum and supremum of the set A.

Lemma 3. *If $\theta < \theta'$, then $z^*(\theta) \geq z^*(\theta')$.*

Proof. Suppose, towards contradiction, that $z = z^*(\theta) < z^*(\theta') = z'$. Observe that $\pi(\theta', z') - \pi(\theta', z) = \int_{u(\theta',z')}^{u(\theta',z)} udF > 0$ implies that $u(\theta', z) > 0$. Observe that $u(\theta, z) \geq u(\theta', z)$ and $u(\theta, z') \geq u(\theta', z')$. Accordingly, $\pi(\theta, z') - \pi(\theta, z) \geq 0$. If $\pi(\theta, z') = \pi(\theta, z)$, then it must be that $u(\theta, z) = u(\theta, z') \in \{\underline{u}, \bar{u}\}$. If $u(\theta, z) = \bar{u}$, then $\pi(\theta, z) = 0 < E[u] = \pi(\theta, 1)$ contradicting the assumption that $z \neq z^*(\theta)$. If $u(\theta, z) = \underline{u}$, then by monotonicity of W, $u(\theta', z) = \underline{u} < 0$ leading to a contradiction. Therefore, it must be that $\pi(\theta, z') > \pi(\theta, z)$ contradicting the supposition that $z = z^*(\theta)$.[31] Q.E.D.

Lemma 4. *Consider a convergent decreasing sequence $\{\theta_n\}_{n=1}^{\infty}$ such that $\theta_n \in \hat{\Theta}$ for each n. Then $\lim_{n\to\infty}\theta_n \in \hat{\Theta}$.*

Proof. Let $\theta_\infty = \lim_{n\to\infty}\theta_n$. By continuity of π and since Z is closed, $\lim_{n\to\infty}\bar{z}(\theta_n)$ exists and $\bar{z}(\theta_\infty) = \lim_{n\to\infty}\bar{z}(\theta_n)$. Lemma 3 establishes that because $\theta_\infty \leq \theta_n$ for each n, $z^*(\theta_\infty) \geq z^*(\theta_n)$. Recognizing that $z^*(\theta_n) = \bar{z}(\theta_n)$, and taking limits yields $z^*(\theta_\infty) \geq \lim_{n\to\infty}\bar{z}(\theta_n) = \bar{z}(\theta_\infty)$. Therefore, $\theta_\infty \in \hat{\Theta}$. Q.E.D.

For the solution when the Planner is partially informed, I draw on techniques and results from Easley and Kiefer (1988). An application of their Theorems 1-3 establish the existence and uniqueness of the value function, $V(\mu; \delta)$, and that the correspondence of maximizers is u.h.c. Let $z(\mu)$ be the selection of the minimum of the maximizers from this correspondence. For each (u, z) in $\mathcal{U} \times Z$, let $\Theta_{u,z} = \{\theta \in \Theta : W(u, \theta, z) \geq 0\}$. Because the argument is standard, the proof for the following lemma is relegated to the Supplementary Appendix.

Lemma 5. *Beliefs satisfy the martingale property: for A, a Borel subset of Θ, $E\left[\mu_{t+1}(A)|\mu_t\right] = \mu_t(A)$.*

Accordingly, the Martingale Convergence Theorem applies, and there is a $P(\Theta)$-valued random variable μ_∞ such that $\mu_t \to_{a.s.} \mu_\infty$. I describe possible limit-beliefs in the support of

[31] I note that since $\pi(\theta, z)$ does not satisfy the single-crossing property in $(\theta, -z)$, a direct proof is necessary.

μ_∞. Suppose that $\hat{\mu}_\infty$ is a limit-belief on a sample-path for which beliefs converge, and for this sample-path, let $\hat{z}_\infty$ be the set of limit points of the sequence of optimal commitment choices $\{z(\mu_t)\}_{t=1}^\infty$. Formally,

$$\hat{z}_\infty = \{z \in Z : \text{for all } \varepsilon > 0,\, T < \infty,\ |z(\hat{\mu}_t) - z| \leq \varepsilon \text{ for some } t \geq T\};$$

since the correspondence of maximizers is u.h.c., $\hat{z}_\infty$ exists.

Lemma 6. *For all $\theta, \theta' \in Supp(\hat{\mu}_\infty)$, $u(\theta, z) = u(\theta', z)$ for all z in $\hat{z}_\infty$.*

Proof. For $\hat{\mu}_\infty$ to be a limit belief, it must be that the distribution of posteriors, $Q(\hat{\mu}_\infty, z)$, puts probability 1 on $\hat{\mu}_\infty$ for all z in $\hat{z}_\infty$; otherwise, $\hat{\mu}_\infty$ almost-surely could not be the limit of $\hat{\mu}_t$ when commitment choices arbitrarily close to z are selected infinitely often. Suppose that for some z in $\hat{z}_\infty$, $u(\theta, z) < u(\theta', z)$. Then there exists $\lambda < 1$ and open sets $O_\varepsilon(\theta)$ and $O_\varepsilon(\theta')$ such that for all t in $O_\varepsilon(\theta)$ and t' in $O_\varepsilon(\theta')$, $\frac{F(u(t,z))}{F(u(t',z))} < \lambda$. By Bayes rule, it must be that

$$\begin{aligned} \frac{\hat{\mu}_\infty(O_\varepsilon(\theta))}{\hat{\mu}_\infty(O_\varepsilon(\theta'))} &= \frac{\int_{O_\varepsilon(\theta)} F(u(t,z))\, d\hat{\mu}_\infty(t)}{\int_{O_\varepsilon(\theta')} F(u(t,z))\, d\hat{\mu}_\infty(t)}. \\ &< \left(\frac{\hat{\mu}_\infty(O_\varepsilon(\theta))}{\hat{\mu}_\infty(O_\varepsilon(\theta'))}\right)\left(\frac{\max_{t \in O_\varepsilon(\theta)} F(u(t,z))}{\min_{t \in O_\varepsilon(\theta')} F(u(t,z))}\right) \\ &< \lambda\left(\frac{\hat{\mu}_\infty(O_\varepsilon(\theta))}{\hat{\mu}_\infty(O_\varepsilon(\theta'))}\right), \end{aligned}$$

which is a contradiction since $\frac{\hat{\mu}_\infty(O_\varepsilon(\theta))}{\hat{\mu}_\infty(O_\varepsilon(\theta'))} > 0$. Q.E.D.

Let ∂_θ denote the Dirac measure that places probability 1 on type θ.

Lemma 7. *For $\theta \in Supp(\hat{\mu}_\infty)$, and $z \in \hat{z}_\infty$, $u(\theta, z) \notin \{\underline{u}, \bar{u}\}$ implies that $\hat{\mu}_\infty = \partial_\theta$. If distinct θ and θ' are both in $Supp(\hat{u}_\infty)$, then $u(\theta, z) = u(\theta', z) = \underline{u}$.*

Proof. Since W is strictly monotone in its arguments, $u(\theta, z) \notin \{\underline{u}, \bar{u}\}$ implies that $u(\theta, z) \neq u(\theta', z)$ for all $\theta' \neq \theta$. Therefore, by Lemma 6, $\hat{\mu}_\infty = \partial_\theta$. If θ and $\theta' \neq \theta$ are both in $Supp(\hat{\mu}_\infty)$, then $u(\theta, z) \in \{\underline{u}, \bar{u}\}$; since it is not optimal for the Planner to commit to not undertaking the action ($u(\theta, z) = \bar{u}$), the only possibility is that $u(\theta, z) = \underline{u}$. Q.E.D.

Recall that $m(\mu, z) = \int_\theta \pi(\theta, z)\, d\mu$; let $m^*(\mu) = \max_{z \in Z} m(\mu, z)$. As is standard in experimentation contexts, the dynamically optimal commitment choice in response to some limit belief $\hat{\mu}_\infty$ must attain the maximal stage payoff $m^*(\hat{\mu}_\infty)$ since at the limiting belief, the Planner is no longer learning and all option value is exhausted. By Lemma 4 of Easley and Kiefer (1988), the following holds.

Lemma 8. *$z \in \hat{z}_\infty$ implies that $m(\hat{\mu}_\infty, z) = m^*(\hat{\mu}_\infty)$.*

Proof of Theorem 2 on p. 14

I begin by proving that learning is globally adequate if Z satisfies FCM. Suppose that $\hat{\Theta}$ is non-empty. By Lemma 4, $\hat{\Theta}$ is a closed interval of the form $\left[\tilde{\theta}, \theta_h\right]$. Denote $\bar{z}\left(\tilde{\theta}\right)$ as $\tilde{z}$, and observe that since W is monotonic in θ, $\tilde{z} \geq \bar{z}(\theta)$ for all $\theta > \tilde{\theta}$. Therefore, $\pi(\theta, \tilde{z}) = \pi^*(\theta)$ for all θ in $\hat{\Theta}$, and thus, $m(\mu, \tilde{z}) = m^*(\mu)$ for all μ with support $\hat{\Theta}$. Lemma 3 establishes that for $\theta < \tilde{\theta}$, $z^*(\theta) \geq \tilde{z}$, and $\theta \notin \hat{\Theta}$ implies that $\bar{z}(\theta) > z^*(\theta) \geq \tilde{z}$.

Lemma 9. *Supp* $(\mu_\infty) \subset \left\{\hat{\mu}_\infty \in P(\Theta) : \textit{Supp}(\hat{\mu}_\infty) \subseteq \hat{\Theta} \text{ or } \hat{\mu}_\infty = \partial_\theta \text{ for some } \theta \notin \hat{\Theta}\right\}$.

Proof. Suppose by way of contradiction that there exists a limit belief $\hat{\mu}_\infty$ in the support of μ_∞ that is not an element of the RHS. Lemma 7 implies that for all $z \in \hat{z}_\infty$, $u(\theta, z) = \underline{u}$ for all θ in *Supp* $(\hat{\mu}_\infty)$. Since *Supp* $(\hat{\mu}_\infty)$ includes $\theta < \tilde{\theta}$, it must be that for all $z \in \hat{z}_\infty$, $z > \tilde{z}$. Let $\check{\theta} = \inf(\textit{Supp}(\hat{\mu}_\infty))$, and $\check{z} = z^*\left(\check{\theta}\right)$. Observe that

$$\begin{aligned} m(\hat{\mu}_\infty, \check{z}) - m(\hat{\mu}_\infty, z) &= \int_\theta (\pi(\theta, \check{z}) - \pi(\theta, z))\, d\hat{\mu}_\infty \\ &= \int_{\Theta \backslash \Theta_{\underline{u}, \check{z}}} (\pi(\theta, \check{z}) - \pi(\theta, z))\, d\hat{\mu}_\infty. \end{aligned}$$

Observe that for $\theta \in \textit{Supp}(\hat{\mu}_\infty) \cap \Theta \backslash \Theta_{\underline{u}, \check{z}}$, $\check{z} \in [z^*(\theta), \bar{z}(\theta))$, and accordingly $\pi(\theta, \check{z}) > \pi(\theta, \bar{z}(\theta)) = \pi(\theta, z)$. Since $\check{\theta} \in \Theta \backslash \Theta_{\underline{u}, \check{z}}$, all sets of the form $B_\varepsilon = \left[\check{\theta}, \check{\theta} + \varepsilon\right)$ are of strictly positive $\hat{\mu}_\infty$-measure, and for sufficiently small ε, $B_\varepsilon \subset \Theta \backslash \Theta_{\underline{u}, \check{z}}$. Therefore, $\hat{\mu}_\infty\left(\Theta \backslash \Theta_{\underline{u}, \check{z}}\right) > 0$ leading to the conclusion that $m(\hat{\mu}_\infty, \check{z}) > m(\hat{\mu}_\infty, z)$ contradicting Lemma 8. Q.E.D.

That payoffs converge to the complete information benchmark follows from Lemma 9. If $\theta \in \hat{\Theta}$, then almost-surely, *Supp* $(\hat{\mu}_\infty) \subset \hat{\Theta}$. Therefore, $\hat{z}_\infty \geq \tilde{z}$, and so $V(\mu_t; \delta) \rightarrow_{a.s.} \pi^*(\theta)$. If $\theta \notin \hat{\Theta}$, then almost-surely, $\hat{\mu}_\infty = \partial_\theta$, and therefore $V(\mu_t; \delta) \rightarrow_{a.s} \pi^*(\theta)$. The argument for $\hat{\Theta} = \emptyset$ follows immediately.

Now suppose that Z fails FCM; thus Θ_P is non-empty. Let $\tilde{\theta} = \inf\left(\hat{\Theta}\right)$; Lemma 4 establishes that $\tilde{\theta} \in \hat{\Theta}$. For any θ, let $z_*(\theta) = \max\{z \in Z : z < \bar{z}(\theta)\}$. For any θ in $\hat{\Theta}$, Z must have a discontinuity at $\bar{z}(\theta)$, otherwise the fully informed Planner would set $z^*(\theta) < \bar{z}(\theta)$. Therefore, $\bar{z}(\theta) - z_*(\theta) > 0$ for θ in $\hat{\Theta}$. Denote $\tilde{z} = z^*\left(\tilde{\theta}\right)$. By continuity of $\pi(\theta, z)$ in θ, there exists $\varepsilon', \varepsilon > 0$ such that $\pi(\theta, \tilde{z}) - \pi\left(\theta, z_*\left(\tilde{\theta}\right)\right) > \varepsilon$ for all θ in $\left[\tilde{\theta}, \tilde{\theta} + \varepsilon'\right]$.

Lemma 10. *For each $\delta > 0$, there exists $\rho > 0$ such that $z(\mu) = \tilde{z}$ for all μ with $\mu\left(\left[\tilde{\theta}, \tilde{\theta} + \varepsilon'\right]\right) > 1 - \rho$.*

Proof. For any $z \leq z_*\left(\tilde{\theta}\right)$, the difference in expected static payoffs between commitment choices z and $\tilde{z}$ is

$$\begin{aligned} m(\mu, \tilde{z}) - m(\mu, z) &= \int_{\Theta_{\underline{u}, \tilde{z}} \backslash \Theta_{\underline{u}, z}} (\pi(t, \tilde{z}) - \pi(t, z))\, d\mu \\ &> \varepsilon\mu\left[\left(\tilde{\theta}, \tilde{\theta} + \varepsilon'\right)\right] - \left(1 - \mu\left[\left(\tilde{\theta}, \tilde{\theta} + \varepsilon'\right)\right]\right)(E[u | u \geq 0]). \end{aligned}$$

Each commitment choice leads to a probability distribution over future beliefs. Since the Planner can always fully commit, the following holds:

$$\int_{P(\Theta)} V(\tilde{\mu};\delta)\, dQ(\mu,\tilde{z}) \geq \frac{E[u]}{1-\delta}.$$

Also, since an alternative commitment z can at best lead to full-information payoffs in one period, and because $z^*(\theta) = \tilde{z}$ for $\theta \in \left[\tilde{\theta}, \tilde{\theta}+\varepsilon'\right]$,

$$\int_{P(\Theta)} V(\tilde{\mu};\delta)\, dQ(\mu,z) \leq \left(\frac{1}{1-\delta}\right)\left(E[u]\,\mu\left[\left(\tilde{\theta},\tilde{\theta}+\varepsilon'\right)\right] + E[u|u\geq 0]\left(1-\mu\left[\left(\tilde{\theta},\tilde{\theta}+\varepsilon'\right)\right]\right)\right).$$

Combining the above inequalities, the payoff difference between choosing commitment $\tilde{z}$ and z is

$$\begin{aligned} & m(\mu,\tilde{z}) + \delta\int_{P(\Theta)} V(\tilde{\mu};\delta)\, dQ(\mu,\tilde{z}) - \left(m(\mu,z) + \delta\int_{P(\Theta)} V(\tilde{\mu};\delta)\, dQ(\mu,z)\right) \\ > \;& \varepsilon\mu\left[\left(\tilde{\theta},\tilde{\theta}+\varepsilon'\right)\right] - \frac{\left(1-\mu\left[\left(\tilde{\theta},\tilde{\theta}+\varepsilon'\right)\right]\right)}{1-\delta}\left(E[u|u\geq 0] - \delta E[u]\right). \end{aligned}$$

Since ε, $E[u]$, and $E[u|u\geq 0]$ are constants, there exists ρ such that $\mu\left[\left(\tilde{\theta},\tilde{\theta}+\varepsilon'\right)\right] > 1-\rho$ ensures that the above inequality is positive. Thus, for any such belief μ, $z(\mu) = \tilde{z}$. Q.E.D.

Suppose that $\theta' \in \left[\tilde{\theta},\tilde{\theta}+\varepsilon'\right]$: by the Martingale Convergence Theorem, beliefs eventually converge. There are two possibilities for belief convergence:

(a) There exists T such that for all $t > T$, $\mu_t\left[\left(\tilde{\theta},\tilde{\theta}+\varepsilon'\right)\right] < 1-\rho$, but $z(\mu_t) \geq \tilde{z}$.

(b) There exists T such that $\mu_T\left[\left(\tilde{\theta},\tilde{\theta}+\varepsilon'\right)\right] > 1-\rho$. By Lemma 10, $z(\mu_T) = \tilde{z}$, and therefore, for all $t > T$, $\mu_t\left[\left(\tilde{\theta},\tilde{\theta}+\varepsilon'\right)\right] > 1-\rho$. Thus, for all $t > T$, $z(\mu_t) = \tilde{z}$.

In either case, with probability 1, there exists T such that for all $t \geq T$, $z(\mu_t) \geq \tilde{z}$. Let $H_{\theta'}^T$ denote the set of sample-paths in which $z(\mu_t) \geq \tilde{z}$ for all $t > T$, and let $H_{\theta'} = \cup_T H_{\theta'}^T$. By above, $\Pr(H_{\theta'}|\theta') = 1$. By definition, $\Theta_P = \left[\tilde{\theta},\theta_h\right]\backslash\hat{\Theta}$, and therefore, by monotonicity of W, Θ_P is a subset of $\Theta_{\underline{u},\tilde{z}}$. By Assumption 1, $\Pr(H_{\theta'}|\theta) > 0$ for all $\theta \in \Theta_P$, and therefore, with positive probability, the Planner chooses commitment levels of $\tilde{z} > z^*(\theta)$ for θ in Θ_P.

Suppose that $\theta \notin \Theta_P \cup \hat{\Theta}$: then $\theta < \tilde{\theta}$. An argument analogous to Lemma 9 establishes that learning is adequate for such a type.

Proof of Theorem 3 on p. 16

Suppose $Z = [-1,1]$. Because W is continuous in each argument, the Intermediate Value Theorem implies that for each θ, there exists z such that $u(\theta,z) = 0$. Thus, the Planner can align the Doer's preferences perfectly, and accordingly $\hat{\Theta} = \emptyset$. Therefore, by Theorem 2, globally adequate learning obtains.

Now suppose that $Z \neq [-1,1]$; since Z is closed, there exists $\varepsilon > 0$ and $z', z'' \in Z$ such that $(z'', z') \cap Z = \emptyset$. Consider $\Theta = \mathcal{U} = [-0.5, 0.5]$, and let $W(u,\theta,z) = u + \theta + \frac{z-z''}{z'-z''}$. Since $W(u,\theta,z) \geq 0$ for all (u,θ) when $z \geq z'$, any commitment level that exceeds z' constitutes full commitment for all types. Since $W(u,\theta,z'') = u+\theta$, the optimal commitment choice for type 0 is z'' but for types sufficiently close to -0.5, the optimal choice is z' that yields full commitment. Therefore, FCM is not satisfied.

Proof of Theorem 4 on p. 18

If Z satisfies FCM, then globally adequate learning follows from Theorem 2. Suppose that Z does not satisfy FCM. For each θ and $\theta' > \theta$, let $\mu_{\theta\theta'}$be the truncated measure of types on the interval $[\theta, \theta']$.[32] As in the proof of Theorem 2, let $\tilde{\theta} = \inf\left(\hat{\Theta}\right)$, $\tilde{z} = z^*\left(\tilde{\theta}\right)$, and recall that $\tilde{\theta} \in \hat{\Theta}$, and that $\bar{z}\left(\tilde{\theta}\right) - z_*\left(\tilde{\theta}\right) > 0$. Since FCM is not satisfied, $\Theta \backslash \hat{\Theta} \cap \left[\tilde{\theta}, \theta_h\right]$ is non-empty. By continuity of π, there exists $\theta^* > \tilde{\theta}$ such that for $\theta \in \left[\tilde{\theta}, \theta^*\right)$, $z^*(\theta) = \bar{z}(\theta) = \tilde{z}$, and for $\theta > \theta^*$, $z^*(\theta) \leq z_*\left(\tilde{\theta}\right)$. By continuity, there exists $\varepsilon > 0$ such that $[\theta^*, \theta^* + \varepsilon) \subset \Theta \backslash \hat{\Theta}$. Consider intervals of the form $B_{\theta\theta'} = [\theta, \theta']$ such that $\theta \leq \tilde{\theta}$, and $\theta' \geq \theta^*$, and $\mu_{\theta\theta'}\left(\left[\tilde{\theta}, \theta^*\right)\right) \geq 1 - \rho$. A straightforward adaptation of Lemma 10 yields that for each $\delta > 0$, there exists $\bar{\rho}$ such that for $\rho < \bar{\rho}$, $z(\mu) = \tilde{z}$. Fix δ and $\rho \in (0, \bar{\rho})$. Let $\bar{\theta} > \theta^*$ be such that $\mu_{\theta\bar{\theta}}\left(\left[\tilde{\theta}, \theta^*\right)\right) \geq 1 - \frac{\rho}{2}$, and let $\underline{\theta} < \tilde{\theta}$ be such that $\mu_{\underline{\theta}\theta^*}\left(\left[\tilde{\theta}, \theta^*\right)\right) \geq 1 - \frac{\rho}{2}$. The combination of both yields $\mu_{\underline{\theta}\bar{\theta}}\left(\left[\tilde{\theta}, \theta^*\right)\right) \geq \frac{2-\rho}{2+\rho} > 1 - \rho$. For any $\theta \in \left[\theta^*, \bar{\theta}\right]$, $\Pr(z_t \geq \tilde{z}) > 0$, and since $\mu_0\left(\left[\theta^*, \bar{\theta}\right] \cap \Theta \backslash \hat{\Theta}\right) > 0$, the result follows.

A.3 Section 4.1

I first describe basic properties of the savings problem that prove useful in later analysis. Let $U(c) = \frac{c^{1-\sigma}-1}{1-\sigma}$. Consider a recursive formulation of the problem in the absence of self-control problems:

$$V(u,y) = Max_{c\in[0,y]} uU(c) + \delta \int_{\mathcal{U}} V\left(u', R(y-c)\right) dF. \tag{3}$$

Restricting attention to a linear consumption path, let $c(u)\,y$ denote the optimal policy when the wealth is y and value of consumption is u. The stochastic Euler equation is

$$\frac{u}{(c(u)\,y)^{\sigma}} = \delta R E\left[\frac{\tilde{u}}{(c(\tilde{u})\,R(1-c(u))\,y)^{\sigma}}\right].$$

Using a guess-and-verify approach, one can see that

$$c(u) = \frac{Cu^{1/\sigma}}{Cu^{1/\sigma} + \delta^{1/\sigma} R^{(1-\sigma)/\sigma}}$$

where C is the unique solution to $E\left[\left(Cu^{1/\sigma} + \delta^{1/\sigma} R^{(1-\sigma)/\sigma}\right)^{\sigma}\right] = 1$ satisfies the above equa-

[32] I.e., for a Borel set A, $\mu_{\theta\theta'}(A) = \int_{\theta}^{\theta'} \mathbf{1}_{t\in A} d\mu_0$.

tion.[33] Also consider the solution to the optimization problem when the Planner is restricted to having no flexibility:

$$\begin{aligned} V^*(y) &= \max_{c\in[0,y]}\left[\int_{\underline{u}}^{\bar{u}} uU(c)\,dF + \delta V^*(R(y-c))\right] \quad (4)\\ &= \max_{c\in[0,y]}\left[U(c)\,dF + \delta V^*(R(y-c))\right], \end{aligned}$$

where the equality follows from $E[u]=1$. A guess and verify approach yields that the optimal policy is a *consumption rate* $\tilde{c} = 1-\delta^{1/\sigma}R^{(1-\sigma)/\sigma}$.

Lemma 11. *$\tilde{c}\in(c(\underline{u}),c(\bar{u}))$ for all y.*

Proof. If $c(\underline{u})>\tilde{c}$, then by calculation,

$$\frac{Cu^{1/\sigma}}{Cu^{1/\sigma}+\delta^{1/\sigma}R^{(1-\sigma)/\sigma}} > 1-\delta^{1/\sigma}R^{(1-\sigma)/\sigma}$$

yields

$$\left(C\underline{u}^{1/\sigma}+\delta^{1/\sigma}R^{(1-\sigma)/\sigma}\right)^{\sigma} > 1.$$

Since the term on the LHS is increasing in u, taking expectations contradicts the definition of C. A similar argument establishes that $c(\bar{u})>\tilde{c}$. Q.E.D.

Now consider the full-information benchmark in which the Planner has beliefs ∂_θ. For any $\bar{c}$, let $\tilde{u}(\theta,\bar{c},y)$ denote the bunching point:

$$\tilde{u}(\theta,\bar{c},y) = \begin{cases} \max\{u\in\mathcal{U}:\theta c(u,y)\le\bar{c}\} & \text{if } \theta c(\underline{u},y)\le\bar{c} \\ \underline{u} & \text{otherwise.}\end{cases}$$

The dynamic program that a fully informed Planner faces is

$$\bar{V}^\theta(y) = Max_{\bar{c}}\left(\begin{array}{c} \int_{u<\tilde{u}(\theta,\bar{c},y)}\left(uU(\theta yc(u))+\delta\bar{V}^\theta(Ry(1-\theta c(u)))\right)dF \\ +\int_{u\ge\tilde{u}(\theta,\bar{c},y)}\left(uU(\bar{c})+\delta\bar{V}^\theta(R(y-\bar{c}))\right)dF \end{array}\right).$$

Lemma 12. *For each Doer type θ, there exists an optimal bunching point $\tilde{u}(\theta)$ independent of y. Moreover, $\tilde{u}(\theta)$ is non-increasing in θ.*

Proof. Suppose that $\tilde{u}(\theta,\bar{c},y)\in(\underline{u},\bar{u})$ for some y: then the optimization problem can be rewritten as selecting some $\tilde{u}(\theta,y)$ and with $\bar{c}=\theta yc(\tilde{u}(\theta))$. Applying Leibniz's Rule, the first order condition w.r.t $\tilde{u}$ is

$$\left(\int_{u>\tilde{u}}\frac{u}{(\theta yc(\tilde{u}))^{\sigma}}dF-\delta R(1-F(\tilde{u}))\frac{d\bar{V}^\theta}{dy'}|_{y'=Ry(1-\theta c(\tilde{u}))}\right)\left(\theta y\frac{dc}{d\tilde{u}}\right)=0.$$

[33] By monotonicity, uniqueness is guaranteed, and $\delta R^{1-\sigma}<1$ ensures existence.

By the Envelope Theorem,

$$\frac{d\bar{V}^{\theta}(y)}{dy}=\int_{u<\tilde{u}}\frac{u}{(\theta yc(u))^{\sigma}}dF+\int_{u>\tilde{u}}\frac{u}{(\theta yc(\tilde{u}))^{\sigma}}dF.$$

Accordingly, $\tilde{u}(\theta,y)$ is determined by an equation independent of y:

$$\frac{E[u|u\geq\tilde{u}]}{(c(\tilde{u}))^{\sigma}}-\frac{\delta R^{1-\sigma}}{(1-\theta c(\tilde{u}))^{\sigma}}\left(\int_{u<\tilde{u}}\frac{u}{(c(u))^{\sigma}}dF+\int_{u>\tilde{u}}\frac{u}{(c(\tilde{u}))^{\sigma}}dF\right)=0. \tag{5}$$

Similarly, by above, if $\tilde{u}(\theta,\bar{c},y)=\underline{u}$ for some y, then the equality in Equation 5 should be replaced by $\leq$, which then implies that $\tilde{u}(\theta,\lambda y)=\underline{u}$ for all $\lambda>0$. If $\tilde{u}(\theta,\bar{c},y)=\bar{u}$, then the Planner is according the Doer with full flexibility, and therefore, it suffices to set $\bar{c}^{*}(\theta)=\theta c^{*}(\bar{u})$. Since the marginal gain from increasing $\tilde{u}$ is decreasing in θ, it follows by Topkis Theorem that $\tilde{u}(\theta)$ is non-increasing in θ. Q.E.D.

As in Section 3, denote the set of types for which the Planner wishes to fully commit as $\hat{\Theta}=\{\theta\in\Theta:\tilde{u}(\theta)=\underline{u}\}$. Since the RHS of Equation 5 is strictly increasing in θ, if $\theta\in\hat{\Theta}$, then for all $\theta'>\theta$, $\theta'\in\hat{\Theta}$. In equation 5 setting $\tilde{u}(\theta)=\underline{u}$ yields

$$\frac{1}{(c(\underline{u}))^{\sigma}}-\frac{\delta R^{1-\sigma}}{(1-\theta c(\underline{u}))^{\sigma}}\left(\frac{1}{(c(\tilde{u}))^{\sigma}}\right)=0,$$

which implies that $\theta c(\underline{u})=\tilde{c}$: the optimal commitment path is the solution to Equation 4. If $\bar{\theta}>\frac{\tilde{c}}{c(\underline{u})}$, then observe that for $\theta=\frac{\tilde{c}}{c(\underline{u})}$, Equation 5 holds with equality at $\tilde{u}=\underline{u}$. Thus, $\hat{\Theta}=\left[\frac{\tilde{c}}{c(\underline{u})},\bar{\theta}\right]$.

Proof of Theorem 5 on p. 21

By the Martingale Convergence Theorem, there exists a random variable μ_{∞} such that $\mu_t\rightarrow_{a.s.}\mu_{\infty}$. Suppose that complete learning does not obtain: consider $\hat{\mu}_{\infty}\in Supp(\mu_{\infty})$ in which there exists θ and $\theta'<\theta$ in the support of $\hat{\mu}_{\infty}$. For both types to be in the support, a generalization of Lemma 6 establishes that the two types must yield the same distribution over consumption paths given commitment choices in the limit. Therefore, along such a sample-path, the limiting upper-bound on consumption rate, $\bar{c}_{\infty}$, must be weakly less than $\theta c^{*}(\underline{u})$ for all $\theta\in Supp(\hat{\mu}_{\infty})$. By the reasoning above, $\bar{c}_{\infty}=(1-\delta)$, and therefore $Supp(\hat{\mu}_{\infty})\subseteq\hat{\Theta}$. Accordingly, the only possibility for incomplete learning is when $\theta\in\hat{\Theta}$; if $\theta\notin\hat{\Theta}$, then almost-surely, the Planner identifies the Doer's type.

A.4 Section 4.2

Consider an individual who faces a binary choice of substance consumption in each period ($x\in\{0,1\}$), and has an addictive state s in $\{0,1,...,S\}$ that summarizes past consumption. Use in a particular state leads to a transition to the next highest state (unless $s=S$) whereas no use leads to a transition to the next lowest state (unless $s=1$ or $s=0$ in which case the state remains unchanged). In each period, the DM, while in the *cold mode*, chooses an activity a in $\{R, A, E\}$. Activity E *exposes* the DM to a high level of substance-related cues, activity A *avoids* many of these cues, and activity R is forced *rehabilitation* that makes consumption infeasible. The activity choice and the prior addictive state triggers the risk of entering the *hot mode* in which the DM always consumes drugs. The probability with which this happens depends on the individual's characteristics and his ability to resist strong cravings. Denote the individual's characteristics by θ drawn from $\{\theta_G, \theta_B\}$ where θ_G is less than θ_B and denote the probability that the DM enters the hot mode by $p(s, a, \theta)$ (where p is increasing in each argument). For choices A and E, consumption of the drug yields a stage payoff of $w_s(x, a)$, which can be negative or positive, reflecting the potential benefits or costs of drug consumption evaluated from the DM's perspective when in the cold mode. I assume that $w_s(x, E)$ exceeds $w_s(x, A)$ for all s, reflecting that avoidance is intrinsically cost. Selecting the activity R yields a stage payoff of $-P$, the cost of rehabilitation. The DM trades off present versus future benefits using a discount factor $\delta < 1$.

A.5 Section 4.3

Proof of Theorem 6 on p. 24

Suppose that the cost of self-control is given by γ. The value from flexibility is

$$Max_{c^*}\left[\Pr(c \le c^*)(b - \gamma E[c|c \le c^*]) + \Pr(c > c^*)(E[c|c > c^*])\right].$$

Solving for the first-order condition under the distributional assumptions yields $c^* = \frac{b}{1+\gamma}$, and a value of $\frac{1}{2}\left(\frac{b^2}{1+\gamma} + 1\right)$. In contrast, the value from commitment is simply b, and therefore, the Planner prefers to be flexible and exercise self-control if and only

$$\frac{1}{2}\left(\frac{b^2}{1+\gamma} + 1\right) \ge b,$$

which reduces to $\gamma \le \gamma^* = \frac{(1-b)^2}{2b-1}$.

Suppose that $\bar{\gamma} \le \gamma^*$: then the set of types for which the Planner chooses to fully commit is empty. Accordingly, an analogue of Theorem 4 applies, yielding globally adequate learning. In contrast, when $\bar{\gamma} > \gamma^*$, then the Planner prefers to fully commit for $\gamma > \gamma^*$, but remain flexible for $\gamma \le \gamma^*$. Distinguishability is violated because once the Planner chooses to fully commit, he cannot observe the efficacy of self-control, and so analogous to Theorem 4, learning is inadequate.

A.6 Section 5

Proof of Proposition 2 on p. 26

The precise condition that is necessary and sufficient is that $\frac{\mu_0\lambda_0}{\mu_0\lambda_0+(1-\mu_0)(1-\lambda_0)} > \frac{2b-1-\underline{\theta}b}{2b\left(\bar{\theta}-\underline{\theta}\right)}$;

notice that the RHS is strictly in $(0,1)$ since $b \in \left(\frac{1}{2-\underline{\theta}}, \frac{1}{2-\bar{\theta}}\right)$, and therefore, if μ_0 and λ_0 are sufficiently high, the condition is satisfied.

To trace out how this condition implies undercommitment: it can be shown through calculation that of the four possible cases $\left\{\left(\bar{\theta},0\right), \left(\bar{\theta}, \left(\bar{\theta}-\underline{\theta}\right)b\right), \left(\underline{\theta},0\right), \left(\underline{\theta}, \left(\bar{\theta}-\underline{\theta}\right)b\right)\right\}$, only in the case where $(\underline{\theta},0)$ is realized would a fully informed Planner (who knows (θ,ξ)) choose to commit. Suppose that this is the realized state of the world. Along any history h^i in which the Planner has been flexible so far, the ratio of posteriors satisfies:

$$\frac{\Pr\left(\bar{\theta}, \left(\bar{\theta}-\underline{\theta}\right)b|h^i\right)}{\Pr\left(\left(\underline{\theta},0\right)|h^i\right)} = \frac{\mu_0\lambda_0}{\left(1-\mu_0\right)\left(1-\lambda_0\right)}, \tag{6}$$

because the state-dependent action probabilities are identical. Consider a sample-path in which the Planner has been flexible at all stages up to this point (I later return to verify that this is optimal). By the Strong Law of Large Numbers, almost-surely, the limiting frequency with which $a=1$ is $\underline{\theta}b$, and therefore, the Planner's posterior ascribe zero probability to the states $\left\{\left(\bar{\theta},0\right), \left(\underline{\theta}, \left(\bar{\theta}-\underline{\theta}\right)b\right)\right\}$. Let $\tilde{\mu}$ be the Planner's posterior belief that $(\theta,\xi) = \left(\bar{\theta}, \left(\bar{\theta}-\underline{\theta}\right)b\right)$, and observe that such a Planner prefers flexibility to commitment if:

$$b - \frac{1}{2} - \tilde{\mu}\left(\bar{\theta}-\underline{\theta}\right)b < \tilde{\mu}\underline{\theta}b\left(b - \left(\frac{\left(\bar{\theta}-\underline{\theta}\right)b + \bar{\theta}b}{2}\right)\right) + \left(1-\tilde{\mu}\right)\underline{\theta}b\left(b - \frac{\underline{\theta}b}{2}\right)$$

which simplifies into $\tilde{\mu} > \frac{2b-1-\underline{\theta}b}{2b\left(\bar{\theta}-\underline{\theta}\right)}$. Observe that since $\tilde{\mu} = \frac{\mu_0\lambda_0}{\mu_0\lambda_0+(1-\mu_0)(1-\lambda_0)}$, the Planner prefers flexibility to commitment at the limit of such a sample-path. It is straightforward to prove that the Planner would also remain flexible at every time period: in every finite period, the Planner ascribes the ratio of posteriors as in Equation 6 and positive probability to the two other states in which he strictly prefers to remain flexible. Therefore, a partially naive Planner always undercommits relative to the full information benchmark whenever $(\underline{\theta},0)$ is the realized state.

Now suppose that the Planner is informed that $\xi=0$; this coincides with Section 2 and so the results of Theorem 1 apply. If the Planner is instead informed that $\underline{\theta}=0$, then along any sample-path in which he remains flexible, his beliefs converge to putting probability 1 on $\xi=0$, and therefore, again, such a Planner chooses to commit.

References

Acemoglu, D., V. Chernozhukov, and M. Yildiz (2009): "Fragility of Asymptotic Agreement under Bayesian Learning," Mimeo.

Aghion, P., P. Bolton, C. Harris, and B. Jullien (1991): "Optimal Learning by Experimentation," *Review of Economic Studies*, 58, 621–654.

Amador, M., I. Werning, and G.-M. Angeletos (2006): "Commitment Vs. Flexibility," *Econometrica*, 74, 365–396.

Ambrus, A. and K. Rozen (2009): "Rationalizing Choice with Multi-Self Models," Cowles Foundation Discussion Papers.

Ameriks, J., A. Caplin, and J. Leahy (2003): "Wealth Accumulation and the Propensity to Plan," *Quarterly Journal of Economics*, 118, 1007–1047.

Ameriks, J., A. Caplin, J. Leahy, and T. Tyler (2007): "Measuring Self Control Problems," *American Economic Review*, 97, 966–972.

Barro, R. (1999): "Ramsey Meets Laibson in The Neoclassical Growth Model*," *Quarterly Journal of Economics*, 114, 1125–1152.

Battaglini, R. Benabou, and J. Tirole (2005): "Self-control in peer groups," *Journal of Economic Theory*, 123, 105–134.

Baumeister, R. F., T. F. Heatherton, and D. M. Tice (1994): *Losing Control: How and Why People Fail at Self-regulation*, Academic Press.

Benabou, R. and M. Pycia (2002): "Dynamic inconsistency and self-control: a planner-doer interpretation," *Economics Letters*, 77, 419–424.

Benabou, R. and J. Tirole (2002): "Self-Confidence and Personal Motivation," *Quarterly Journal of Economics*, 117, 871–915.

——— (2004): "Willpower and Personal Rules," *Journal of Political Economy*, 112, 848–886.

Benhabib, J. and A. Bisin (2005): "Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumption–saving decisions," *Games and Economic Behavior*, 52, 460–492.

Bernheim, B. D. and A. Rangel (2004): "Addiction and Cue-Triggered Decision Processes," *American Economic Review*, 94, 1558–1590.

Billingsley, P. (1995): *Probability and Measure*, Wiley Series in Probabilitiy and Mathematical Statistics, New York: Wiley, third ed.

Bodner, R. and D. Prelec (2003): "Self-Signaling and Diagnostic Utility in Everyday Decision Making," in *The Psychology of Economic Decisions*, ed. by I. Brocas and J. D. Carrillo, Oxford: Oxford University Press.

Brocas, I. and J. D. Carrillo (2008): "The Brain as a Hierarchical Organization," *American Economic Review*, 98, 1312–1346.

Brunnermeier, M. K. and J. A. Parker (2005): "Optimal Expectations," *The American Economic Review*, 95, 1092–1118.

Cadena, B. and B. Keys (2008): "Can Self-Control Explain Avoiding Free Money? Evidence from Interest-Free Student Loans," Mimeo.

Carrillo, J. D. and M. Dewatripont (2008): "Promises, Promises," *The Economic Journal*, 118, 1453–1473.

Carrillo, J. D. and T. Mariotti (2000): "Strategic Ignorance as a Self-Disciplining Device," *The Review of Economic Studies*, 67, 529–544.

Charness, G. and U. Gneezy (2009): "Incentives to Exercise," *Econometrica*, 77, 909–931.

Chatterjee, K. and R. Krishna (2009): "A "Dual Self Representation for Stochastic Temptation," *American Economic Journal: Microeconomics*, 1, 148–167.

Compte, O. and A. Postlewaite (2004): "Confidence-Enhanced Performance," *American Economic Review*, 94, 1536–1557.

Dal Bo, E. and M. Tervio (2008): "Self-Esteem, Moral Capital, and Wrongdoing," Mimeo.

Dekel, E., D. Fudenberg, and D. K. Levine (1999): "Payoff Information and Self-Confirming Equilibrium," *Journal of Economic Theory*, 89, 165–185.

——— (2004): "Learning to Play Bayesian Games," *Games and Economic Behavior*, 46, 282–303.

Dekel, E. and B. L. Lipman (2007): "Self–Control and Random Strotz Representations," Mimeo.

DellaVigna, S. (2009): "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47, 315–372.

DellaVigna, S. and U. Malmendier (2006): "Paying Not to Go to the Gym," *The American Economic Review*, 96, 694–719.

Easley, D. and N. M. Kiefer (1988): "Controlling a Stochastic Process with Unknown Parameters," *Econometrica*, 56, 1045–1064.

Esponda, I. (2008): "Behavioral Equilibrium in Economies with Adverse Selection," *American Economic Review*, 98, 1269–91.

Fudenberg, D. (2006): "Advancing Beyond Advances in Behavioral Economics," *Journal of Economic Literature*, 44, 694–711.

Fudenberg, D. and D. K. Levine (1993a): "Self-Confirming Equilibrium," *Econometrica*, 61, 523–545.

——— (1993b): "Steady State Learning and Nash Equilibrium," *Econometrica*, 61, 547–573.

——— (2006): "A Dual Self Model of Impulse Control," *American Economic Review*, 96, 1449–1476.

——— (2009a): "Learning and Equilibrium," *Annual Review of Economics*, 1, 385–420.

——— (2009b): "Self Control, Risk Aversion, and the Allais Paradox," Mimeo.

Giné, X., D. Karlan, and J. Zinman (2009): "Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation," Mimeo.

Gul, F. and W. Pesendorfer (2001): "Temptation and Self-Control," *Econometrica*, 69, 1403–1435.

——— (2005): "The Revealed Preferences Theory of Changing Tastes," *Review of Economic Studies*, 72, 429–448.

Harris, C. and D. Laibson (2003): "Hyberbolic Discounting and Consumption," in *Advances in Economics and Econometrics: Theory and Applications: Eighth World Congress*, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, Econometric Society, 258–297.

Huffman, D. and M. Barenstein (2005): "A Monthly Struggle for Self-Control? Hyperbolic Discounting, Mental Accounting, and the Fall in Consumption Between Paydays," Mimeo.

Kahneman, D., P. P. Wakker, and R. Sarin (1997): "Back to Bentham? Explorations of Experienced Utility," *The Quarterly Journal of Economics*, 112, 375–405.

Kaper, J., E. Wagena, M. Willemsen, and C. Van Schayck (2005): "Reimbursement for smoking cessation treatment may double the abstinence rate: results of a randomized trial." *Addiction*, 100, 1012.

——— (2006): "A randomized controlled trial to assess the effects of reimbursing the costs of smoking cessation therapy on sustained abstinence." *Addiction*, 101, 1656.

Kőszegi, B. (2006): "Ego utility, overconfidence, and task choice," *Journal of the European Economic Association*, 4, 673–707.

Kivetz, R. and I. Simonson (2002): "Self-Control for the Righteous: Toward a Theory of Precommitment to Indulgence," *Journal of Consumer Research*, 29, 199–217.

Kreps, D. (1979): "A Representation Theorem for "Preference for Flexibility"," *Econometrica*, 47, 565–578.

Krusell, P. and A. A. Smith, Jr (2003): "Consumption-Savings Decisions with Quasi-Geometric Discounting," *Econometrica*, 71, 365–375.

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112, 443–477.

Laibson, D., A. Repetto, and J. Tobacman (1998): "Self-Control and Saving for Retirement," *Brookings Papers on Economic Activity*, 1, 91–172.

Lander, J. (1990): "Fallacies and phobias about addiction and pain," *Addiction*, 85, 803–809.

Loewenstein, G. and T. O'Donoghue (2007): "The Heat of the Moment: Modeling Interactions Between Affect and Deliberation," Mimeo.

Noor, J. (2007): "Commitment and Self-Control," *Journal of Economic Theory*, 135, 1–34.

O'Brien, M. S. and J. C. Anthony (2005): "Risk of becoming cocaine dependent: epidemiological estimates for the United States, 2000–2001," *Neuropsychopharmacology*, 30, 1006–1018.

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

Orphanides, A. and D. Zervos (1995): "Rational Addiction with Learning and Regret," *Journal of Political Economy*, 103, 739–758.

Pollak, R. A. (1968): "Consistent planning," *The Review of Economic Studies*, 35, 201–208.

Prelec, D. and G. Loewenstein (1998): "The Red and the Black: Mental Accounting of Savings and Debt," *Marketing Science*, 17, 4–28.

Rothschild, M. (1974): "A Two-Armed Bandit Theory of Market Pricing," *Journal of Economic Theory*, 9, 185–202.

Rubinstein, A. (2005): "Discussion of 'Behavioral Economics'," in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, ed. by R. Blundell, W. Newey, and T. Persson, Econometric Society, 246–254.

Rubinstein, A. and A. Wolinsky (1994): "Rationalizable conjectural equilibrium: between Nash and rationalizability," *Games and Economic Behavior*, 6, 299–311.

Shui, H. and L. M. Ausubel (2004): "Time Inconsistency in the Credit Market," Mimeo.

Strotz, R. H. (1955): "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies*, 23, 165–180.

Thaler, R. H. and H. M. Shefrin (1981): "An Economic Theory of Self-Control," *Journal of Political Economy*, 89, 392–406.

Viscusi, W. K. (1990): "Do Smokers Underestimate Risks?" *Journal of Political Economy*, 98, 1253–1269.

Wagner, F. A. and J. C. Anthony (2002): "From First Drug Use to Drug Dependence: Developmental Periods of Risk for Dependence upon Marijuana, Cocaine, and Alcohol," *Neuropsychopharmacology*, 26, 479–488.

Wang, R. (2007): "The Optimal Consumption and the Quitting of Harmful Addictive Goods," *The B.E. Journal of Economic Analysis and Policy (Contributions)*, 7, 1–36.

Wertenbroch, K. (2003): "Self-Rationing: Self-Control in Consumer Choice," in *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*, ed. by R. Baumeister, G. Loewenstein, and D. Read, Russell Sage Foundation, 419–516.

B Supplementary Appendix

Proof of Proposition 1 on p. 8

I first establish that V exists, satisfies Equation 1, is continuous and non-decreasing in μ. Define the function $V^t(\mu)$ recursively as follows:

$$\begin{aligned} V^0(\mu) &= 0 \\ V^t(\mu) &= \max \left\{ \begin{array}{c} \delta V^{t-1}(\mu) \\ b - \frac{1}{2} + \delta V^{t-1}(\mu) \\ (1-\mu)\left(\underline{\theta} b \left(b - \frac{\underline{\theta} b}{2} + \delta V^{t-1}(\mu^+)\right) + \delta\left(1 - \underline{\theta} b\right) V^{t-1}(\mu^-)\right) + \\ \mu\left(\bar{\theta} b \left(b - \frac{\bar{\theta} b}{2} + \delta V^{t-1}(\mu^+)\right) + \delta\left(1 - \bar{\theta} b\right) V^{t-1}(\mu^-)\right) \end{array} \right\} \end{aligned}$$

where μ^+ and μ^- are the posteriors conditional on observing $(z_t, a_t) = (0,1)$ and $(z_t, a_t) = (0,0)$ respectively, given a prior μ. To simplify notation, let

$$W^t(\mu) = \begin{array}{c} (1-\mu)\left(\underline{\theta} b \left(b - \frac{\underline{\theta} b}{2} + \delta V^{t-1}(\mu^+)\right) + \delta\left(1 - \underline{\theta} b\right) V^{t-1}(\mu^-)\right) + \\ \mu\left(\bar{\theta} b \left(b - \frac{\bar{\theta} b}{2} + \delta V^{t-1}(\mu^+)\right) + \delta\left(1 - \bar{\theta} b\right) V^{t-1}(\mu^-)\right) \end{array}$$

I verify that V^t satisfies continuity and non-increasing. I then define V as the point-wise limit of V^t but show that convergence is uniform. This verifies the existence of V, and that V inherits continuity and monotonicity from V^t.

Lemma 13. *For each $t \geq 0$, $V^t(\mu)$ is continuous and non-decreasing in μ.*

Proof. Continuity clearly holds for $t = 0$. By induction, if continuity holds for $t \leq \tau$, then it must hold for $\tau + 1$ since $V^{\tau+1}$ is the upper-envelope of sums of continuous functions. Observe that $V^0(\mu)$ is non-decreasing in μ. By induction, if $V^t(\mu)$ is nondecreasing for all $t \leq \tau$, observe that $V^{\tau+1}(\mu)$ is the upper-envelope of functions that are non-decreasing in μ, and therefore is also non-decreasing in μ. Q.E.D.

Lemma 14. *For all $\mu \in [0,1]$, $V^t(\mu)$ is non-decreasing in t.*

Proof. Observe that $V^t(\mu) \neq \delta V^{t-1}(\mu)$ for any μ, t. Note that $V^1(\mu) > 0 = V^0(\mu)$ for all $\mu \in [0,1]$. Let $V^t(\mu) \geq V^{t-1}(\mu)$ for all $\mu \in [0,1]$, and $t \leq \tau$. Then $b - \frac{1}{2} + \delta V^\tau(\mu) \geq b - \frac{1}{2} + \delta V^{\tau-1}(\mu)$, and $W^{\tau+1}(\mu) \geq W^\tau(\mu)$. Therefore, $V^{\tau+1}(\mu) \geq V^\tau(\mu)$. Q.E.D.

Given that $V^t(\mu)$ is non-decreasing in t, and $V^t(\mu)$ is uniformly bounded above by $\frac{b}{1-\delta}$, by the Bolzano-Weierstrass Theorem, there exists a point-wise limit $V(\mu) = \lim_{t\to\infty} V^t(\mu)$. In fact, since $V^t(\mu) \leq V(\mu) \leq V^t(\mu) + \delta^t \frac{b}{1-\delta}$, for every μ, $\left|V(\mu) - V^t(\mu)\right| \leq \delta^t \frac{b}{1-\delta}$. Consequently, the convergence is uniform leading to the following corollary.

Corollary 1. *$V(\mu)$ satisfies Equation 1, and is continuous and non-decreasing in μ.*

Let $W(\mu) = \lim_{t\to\infty} W^t(\mu)$. Observe that $V(0) = \frac{b-\frac{1}{2}}{1-\delta} > \underline{\theta} b\left(b - \frac{\underline{\theta} b}{2}\right) + \delta V(0) = W(0)$, and $\frac{b-\frac{1}{2}}{1-\delta} < \frac{\bar{\theta} b\left(b-\frac{\bar{\theta} b}{2}\right)}{1-\delta} = V(1) = W(1)$. Also observe that if $V(\mu) > W(\mu)$, then for all $\tilde{\mu} < \mu$, $V(\tilde{\mu}) = V(\mu) > W(\mu) \geq W(\tilde{\mu})$, since it can be shown that W is non-decreasing in μ. Therefore, the set, $\{\mu \in [0,1] : V(\mu) > W(\mu)\}$ is an interval. By continuity of V and W, $\mu^* = \sup\{\mu \in [0,1] : V(\mu) > W(\mu)\} = \min\{\mu \in [0,1] : V(\mu) = W(\mu)\}$.

Proof of Lemma 5 on p. 29

Observe that

$$
\begin{aligned}
& E\left[\mu_{t+1}(A)\,|\mu_t\right] \\
= & \int_\Theta \int_{\mathcal{U}} \left(\mathbf{1}_{U_{\theta,z}} \int_Y B(1,y,z)(A)\, d\tilde{F}(y|u,1) + \mathbf{1}_{\mathcal{U}\setminus U_{\theta,z}} \int_Y B(0,y,z)(A)\, d\tilde{F}(y|u,0) \right) dF d\mu_t
\end{aligned}
$$

where $B(a,y,z)(A)$ is the posterior belief that $\theta \in A \subseteq \Theta$ given observed action a, signal y, and commitment choice z. Denote $U_{\theta,z}$ by $\tilde{U}^1_{\theta,z}$, and $\mathcal{U}\setminus U_{\theta,z}$ by $\tilde{U}^0_{\theta,z}$, $\Theta_{u,z}$ by $\tilde{\Theta}^1_{u,z}$, $\Theta\setminus\Theta_{u,z}$ by $\tilde{\Theta}^0_{u,z}$. Observe that

$$
\begin{aligned}
B(a,y,z)(A) &= \frac{\int_A \int_{\mathcal{U}} \mathbf{1}_{\tilde{U}^a_{\theta,z}}\, d\tilde{F}(y|u,a)\, dF d\mu_t}{\int_\Theta \int_{\mathcal{U}} \mathbf{1}_{\tilde{U}^a_{\theta,z}}\, d\tilde{F}(y|u,a)\, dF d\mu_t} \\
&= \frac{\int_{\mathcal{U}} \mu_t\left(A \cap \tilde{\Theta}^a_{u,z}\right) d\tilde{F}(y|u,a)\, dF}{\int_{\mathcal{U}} \mu_t\left(\tilde{\Theta}^a_{u,z}\right) d\tilde{F}(y|u,a)\, dF}.
\end{aligned}
$$

Accordingly,

$$
\begin{aligned}
& \int_\Theta \int_{\mathcal{U}} \mathbf{1}_{\tilde{U}^a_{\theta,z}} \left(\int_Y B(a,y,z)(A)\, d\tilde{F}(y|u,a)\, dy \right) dF d\mu_t \\
= & \int_{\mathcal{U}} \int_\Theta \mathbf{1}_{\tilde{\Theta}^a_{u,z}} \left(\int_Y B(a,y,z)(A)\, d\tilde{F}(y|u,a)\, dy \right) d\mu_t dF \\
= & \int_{\mathcal{U}} \mu_t\left(\tilde{\Theta}^a_{u,z}\right) \left(\int_Y \frac{\int_{u'\in\mathcal{U}} \mu_t\left(A \cap \tilde{\Theta}^a_{u',z}\right) d\tilde{F}(y|u',a)\, dF}{\int_{u'\in\mathcal{U}} \mu_t\left(\tilde{\Theta}^a_{u',z}\right) d\tilde{F}(y|u',a)\, dF} d\tilde{F}(y|u,a)\, dy \right) dF \\
= & \int_Y \left(\frac{\int_{u'\in\mathcal{U}} \mu_t\left(A \cap \tilde{\Theta}^a_{u',z}\right) d\tilde{F}(y|u',a)\, dF}{\int_{u'\in\mathcal{U}} \mu_t\left(\tilde{\Theta}^a_{u',z}\right) d\tilde{F}(y|u',a)\, dF} \right) \left(\int_{\mathcal{U}} \mu_t\left(\tilde{\Theta}^a_{u,z}\right) d\tilde{F}(y|u,a)\, dF \right) dy \\
= & \int_Y \int_{\mathcal{U}} \mu_t\left(A \cap \tilde{\Theta}^a_{u,z}\right) d\tilde{F}(y|u,a)\, dF dy \\
= & \int_{\mathcal{U}} \mu_t\left(A \cap \tilde{\Theta}^a_{u,z}\right) dF
\end{aligned}
$$

Hence, $E\left[\mu_{t+1}(A)\,|\mu_t\right] = \int_{\mathcal{U}} \mu_t\left(A \cap \Theta_{u,z}\right) dF + \int_{\mathcal{U}} \mu_t\left(A \setminus \Theta_{u,z}\right) dF = \mu_t(A)$.