INCENTIVES FOR PROCRASTINATORS*

TED O'DONOGHUE AND MATTHEW RABIN

We examine how principals should design incentives to induce time-inconsistent procrastinating agents to complete tasks efficiently. Delay is costly to the principal, but the agent faces stochastic costs of completing the task, and efficiency requires waiting when costs are high. If the principal knows the task-cost distribution, she can always achieve first-best efficiency. If the agent has private information, the principal can induce first-best efficiency for time-consistent agents, but often cannot for procrastinators. We show that second-best optimal incentives for procrastinators typically involve an increasing punishment for delay as time passes.

I. INTRODUCTION

While the standard economics model assumes that any desire to delay an unpleasant task must be time-consistent, many people have a time-*inconsistent* preference for *procrastination.* Today we feel we should write a referee report tomorrow, but tomorrow we tend to delay again. A small set of economists and psychologists has over the years proposed formal models of time-inconsistent preferences and self-control problems, where people have a tendency to pursue their immediate well-being in a way that their "long-run selves" do not appreciate. O'Donoghue and Rabin [1999] build from this prior research by showing that a person who is time-inconsistent and unaware of the time inconsistency will procrastinate in completing an unpleasant task.[1]

In this paper we examine the implications of time-inconsistent procrastination for the design of *temporal incentive schemes,* which reward agents based on when they complete tasks. Tempo-

* We thank Steven Blatt, Olivier Blanchard, Erik Eyster, two anonymous referees, and seminar participants at the University of California at Berkeley, University of Chicago, and University of Michigan for helpful comments, and Doug Almond, Steven Blatt, and Erik Eyster for valuable research assistance. For financial support, we thank the National Science Foundation (Award 9709485), O'Donoghue thanks the Alfred P. Sloan Foundation; and Rabin thanks the Russell Sage, MacArthur, and Alfred P. Sloan Foundations. Part of this research was completed while both authors were visitors at the Math Center at Northwestern University, and we are grateful for their hospitality and financial support. This draft was completed while Rabin was a visiting fellow at the Center for Advanced Study in the Behavioral Sciences, supported by NSF Grant No. SBR-960123.

1. This finding replicates and extends a similar example of procrastination by Akerlof [1991], who motivates a mathematically similar model of choice behavior by emphasizing how the costs of doing a task are more salient when they are immediate than when they are delayed. For other papers discussing procrastination and time-inconsistent preferences, see Prelec [1989], Fischer [1997], and O'Donoghue and Rabin [1998a, 1998b].

The Quarterly Journal of Economics, August 1999

ral incentive schemes are a central aspect of organizational design and many types of contracts. People face punishments for delay; sometimes these punishments are explicit in the form of decreased compensation, but more often they are implicit in the form of admonitions from supervisors and decay in reputation. Such incentives are needed when an agent finds a task unpleasant and does not intrinsically value its timely completion.

In most of our analysis we examine how principals should design incentives to avoid inefficient delay. We first show that principals must make incentives for procrastinators "harsher" than those for nonprocrastinators: whereas for time-consistent agents the optimal incentive scheme exactly reflects the principal's true cost of delay, for procrastinators the principal must punish delay more severely than the true cost of delay in order to counteract procrastination. But our main conclusion is that, even when the true cost of delay is constant, optimal incentive schemes for procrastinators often involve increasing punishment for delay as time passes.

In Section II we introduce a simple model of time-inconsistent preferences originally proposed by Phelps and Pollak [1968] in the context of intergenerational altruism and later used by Laibson [1994] to model time inconsistency within an individual: a person always values her well-being *now* more than her well-being at any future moment, and values her well-being at all future moments equally.

In Section III we introduce our model of temporal incentive schemes. A risk-neutral principal hires a risk-neutral agent to complete some task. Because the principal faces a *delay cost,* she prefers that the task be done sooner rather than later. But the agent faces a stochastic *task cost,* and it may be best for him to delay when the task-cost realization is particularly high. Efficient behavior minimizes the sum of expected delay costs and expected task costs. Moral hazard arises because the principal cannot observe task-cost realizations, and therefore she must compensate the agent based solely on when the agent completes the task. Unfortunately, she does not know whether an observed delay is an efficient response to a high task-cost realization or inefficient procrastination. We explore whether temporal incentive schemes can induce efficient behavior, and, if not, which temporal incentive schemes are second-best optimal.

Time-consistent agents do not procrastinate, and therefore the optimal incentive scheme is straightforward: if the incremental punishment for delay exactly mirrors the principal's delay costs, then the agent will internalize those delay costs, balance them against his task costs, and behave efficiently. Importantly, the optimal incentive scheme is independent of the probability distribution of the agent's task costs.

Time-*inconsistent* agents procrastinate, and as a result incentive schemes must deter inefficient procrastination, yet still encourage efficient delay when the task-cost realization is high. We show that as long as the principal knows the distribution of task costs, she can fully counteract the agent's tendency to procrastinate with incentives that punish delay by more than its actual cost. Regardless of the agent's propensity to procrastinate, therefore, a fully efficient incentive scheme can be implemented. However, this (first-best) optimal incentive scheme for time-inconsistent agents very much depends on the distribution of task costs. Specifically, higher task costs (on average) make the agent more prone to procrastinate, in which case the principal must impose a more severe punishment for delay to counteract procrastination.

In Sections IV and V we assume that the agent has private information about the distribution of task costs, making the principal uncertain about the agent's propensity to procrastinate. We then investigate the nature of incentive schemes when the principal's incremental cost of delay is stationary. For time-consistent agents the optimal full-information incentive scheme is independent of the distribution of task costs, and therefore first-best efficiency is feasible. Specifically, the principal should impose a stationary punishment for delay exactly equal to the true stationary delay cost. For time-inconsistent agents, in contrast, the optimal full-information punishment for delay depends on the agent's propensity to procrastinate, and therefore the principal's uncertainty creates a problem: punishment for delay that is harsh enough to prevent excessive procrastination by severe procrastinators may be so harsh that moderate procrastinators complete the task when it would be more efficient to wait. As a result, first-best optimality typically will not be feasible when the agent is privately informed.

In Section IV we consider the case where task-cost distributions differ only in their means, with the same probability distribution around this mean. We show that second-best optimal incentive schemes for procrastinators typically will be nonstationary, taking the following form: incremental delay will be punished moderately early on, but more severely after some "deadline." Such a scheme initially allows those with little propensity to procrastinate to wait until it is efficient to do the task, while the deadline assures that severe procrastinators do not delay too long. An intuition for why the principal finds this scheme optimal is that it implicitly reflects her Bayesian updating: the longer an agent delays completing a task, the more likely the agent is a severe procrastinator, and therefore the more attractive the principal finds it to punish incremental delay severely.

In Section V we relax the assumption that the distribution of task costs around the mean is the same for all task-cost distributions. We show that, for the natural case where agents with lower average task costs also have lower variance in those costs, the "deadline result" from Section IV holds.

An important issue in modeling time-inconsistent preferences has been downplayed in the discussion above: how aware are people that in the future they might behave in ways contrary to their current preferences? In our context, do people predict their tendencies to procrastinate? In Section II we discuss two extreme assumptions that have appeared in the literature on time-inconsistent preferences: *sophisticated* people are fully aware of their future self-control problems, and *naive* people are completely *un*aware of their future self-control problems.[2] Because we feel that day-to-day procrastination is characterized by a large degree of naivete—and because naive behavior is far more tractable than sophisticated behavior—our formal model assumes naivete. We discuss in Section VI how sophistication might affect our results.

The assumption that people are naive about their self-control problems gives rise to an important additional issue: since naive people overestimate their payoff from an incentive scheme (because they do not realize that their procrastination will lower their wages), principals aware of this procrastination might hire people merely to bilk them of money rather than to efficiently complete a task. We discuss in Section III how reputational

2. Strotz [1956] and Pollak [1968] carefully lay out these two assumptions (and develop the labels), but do not much consider the implications of assuming one versus the other. More recent papers have assumed either one or the other, without attempting to justify the choice with behavioral evidence. For instance, Akerlof [1991] assumes naive beliefs (in a slightly different model), as do O'Donoghue and Rabin [1998a, 1998b], while Laibson [1994, 1995, 1997] assumes sophisticated beliefs. O'Donoghue and Rabin [1999] consider both, and explicitly contrast the two, but likewise do not provide behavioral evidence for either.

pressures and other factors might lead a principal to want to induce efficiency, as assumed in the discussion above, and most of our analysis in the paper assumes that the principal prefers to induce efficient behavior. Even so, in Section III we examine the alternative assumption that the principal wishes to bilk the agent. We show that she can always do so with an incentive scheme that punishes delay sufficiently leniently so as to lull the agent into severe and costly procrastination.

We conclude in Section VII with a discussion of some caveats and possible extentions to the model of this paper, and a discussion of some other implications of procrastination for organizational design.

II. Present-Biased Preferences and Procrastination

O'Donoghue and Rabin [1999] coined the term "present-biased preferences" for the class of time-inconsistent preferences where a person puts greater and greater weight on his well-being at an earlier moment over a later moment as the earlier moment gets closer. Casual observation, introspection, and psychological research all indicate such time inconsistency. Variants of such preferences (frequently going by the name of "hyperbolic discounting") have been studied by many researchers.[3] To illustrate, consider a choice between spending seven hours to complete an unpleasant task on April 1 and spending eight hours to complete the same unpleasant task on April 15, assuming your opportunity cost of time is the same on both dates. The task could be completing your taxes—on April 15 you would have to take the extra time to go to the post office to mail your returns, whereas on April 1 you could simply mail it without hassle on your way to work the next day.

If asked to commit on February 1 to one or the other, most people would prefer to do less work in April, and would therefore choose seven hours on April 1. If they must choose on April 1, however, many people would be inclined to put off the task two weeks rather than do it right away. When April 1 arrives, people

3. See Chung and Herrnstein [1967], Ainslie [1974, 1975, 1987, 1991, 1992], Ainslie and Herrnstein [1981], Thaler [1991], Funder and Block [1989], Hoch and Loewenstein [1991], Ainslie and Haslam [1992a, 1992b], Loewenstein and Prelec [1992], and Kirby and Herrnstein [1995]. For early economics papers focusing on time-inconsistent discounting, see Strotz [1956], Phelps and Pollak [1968], Pollak [1968], and Goldman [1979, 1980].

have a preference for immediate gratification—not doing the unpleasant task today—with which their long-run selves disagree.

Phelps and Pollak [1968] put forward an elegant model of intertemporal preferences in the context of intergenerational altruism, which Laibson [1994] later used to capture this time-inconsistent taste for immediate gratification.[4] Let u_t be the instantaneous utility a person gets in period t. Then her intertemporal preferences at time t, U^t, can be represented by the following utility function:

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

The parameter δ represents long-run, time-consistent impatience, and for $\beta = 1$ these preferences are simply (the discrete version of) exponential discounting. But for $\beta < 1$, these preferences parsimoniously capture a time-inconsistent taste for immediate gratification. Since we shall focus in this paper on relatively short horizons, we assume that $\delta = 1$, so there is no time-consistent discounting. Hence, the intertemporal utility function is

$$\text{For all } t,\ U^t(u_t,u_{t+1}, \ldots , u_T) = u_t + \beta \sum_{\tau=t+1}^{T} u_\tau .$$

Consider again the above example. Assume that your instantaneous disutility from doing work is simply the number of hours of work, so that $u_t(7) = -7$ and $u_t(8) = -8$ for all t. Suppose that $\beta = .8$: you are willing to forgo a given gain in utility *in the future* for a gain in utility *now* that is only 80 percent as large. Consider your decision on February 1. Because on February 1 you discount both dates by β, you will choose to work seven hours on April 1 rather than eight hours on April 15. Contrast this with what your decision would be on April 1. You can experience a utility of -7 by working today, or experience a discounted utility of $.8(-8) = -6.4$ by delaying the work for two weeks. You will, therefore, delay work. Hence, for the exact same decision, your choice on April 1 is different than your choice on February 1. Irrespective of its specific prediction, exponential discounting would predict that your choice would be the same whether you decide on February 1 or April 1.

4. This model has since been used by Laibson [1995, 1997], O'Donoghue and Rabin [1998a, 1998b, 1999], Fischer [1997], and others.

To examine dynamic choice given time-inconsistent preferences, researchers have converged on a simple modeling strategy: a single individual is modeled as many separate "selves," one for each period. Each period's self chooses her current behavior to maximize her current preferences, where the person's future selves will control her future behavior. In such a framework, an important issue arises: what are a person's beliefs about how her future selves will behave? Two extreme assumptions have appeared in the literature. *Sophisticated* people are fully aware of their future self-control problems and therefore know exactly how their future selves will behave. *Naive* people are fully *un*aware of their future self-control problems and therefore believe their future selves will behave exactly as they currently would like them to behave.

There seem to be elements of both sophistication and naivete in people. Some degree of sophistication is implied by the fact that people often pay to commit themselves to smaller choice sets (e.g., joining fat farms or Christmas clubs, or buying small rather than large packages of enticing goods). A naive person would never worry that her tomorrow's self might choose an option that she does not like today, and she therefore would find committing herself unattractive. On the other hand, people do seem to overestimate the degree to which they will abide by their plans for the future. For example, people who repeatedly do not have the "willpower" to forgo tempting foods or quit smoking predict that tomorrow they will.[5]

O'Donoghue and Rabin [1999] examine the implications of assuming sophistication versus naivete. One of the conclusions is that sophistication often leads to complicated behavior. In the context of this paper, small changes in incentive schemes can lead to dramatic changes in behavior, and incentive schemes that yield stationary behavior for both time-consistent agents and naive

5. Economists examining self-control issues seem to be inclined to assume sophistication, well beyond what we feel behavioral evidence supports. In part, this inclination derives from a desire to depart from familiar economic assumptions one step at a time—naivete is two steps away by simultaneously assuming time-inconsistent preferences and "irrational expectations" about those preferences. Indeed, the analysis in this paper to some extent reinforces this worry: many of the issues to which naivete gives rise are unfamiliar and problematic for economic analysis. Yet, in this and related research, we have discovered a pattern: in many models, naivete is *far* more tractable than sophistication in terms of the practical logistics of formal models. We fear, therefore, that a conservative weddedness to the sophistication assumption may not only be tenuous behaviorally, but also may create unnecessary technical roadblocks to the incorporation of self-control and time-inconsistency issues into economics.

time-inconsistent agents can yield highly nonstationary behavior for sophisticated time-inconsistent agents. This makes the search for optimal incentive schemes a much more difficult exercise. In part to avoid such difficulties, we shall focus mostly in this paper on naive beliefs. While extreme, we do not think our focus on naive beliefs is without behavioral foundation. Much day-to-day procrastination seems to be characterized by a large degree of naivete: we procrastinate today thinking we will complete some task tomorrow, but tomorrow we decide to delay again. Importantly, even when we are aware of a general tendency to procrastinate, we seem capable of underestimating this tendency on a case-by-case basis. Even so, we discuss briefly in Section VI how sophistication might affect our results.

III. A Model of Temporal Incentive Schemes

Suppose that a principal hires an agent to complete some task. The two parties sign a contract specifying how the principal will compensate the agent, where wages can depend only on information available to both parties. In contrast to the typical principal-agent model, we assume that there is no uncertainty about whether an action has been taken, nor about the level of effort by the agent, so that there is no moral hazard of the traditional sort. Rather, we focus on the problem of *when* the agent completes the task if there is day-to-day uncertainty over the cost to the agent of completing the task. For example, on any given day the agent may be sick, may have a particularly exciting episode of *21 Jump Street* to watch, may not have ready access to equipment needed for the task, or may simply have more pressing projects to complete. Efficiency may require that the agent wait on days with a high task cost, and do it on days with a low task cost. Moral hazard can arise if the principal cannot observe the task-cost realizations. We explore the role of *temporal incentive schemes*—contracts where wages are contingent on when the agent completes the task—in such an environment.[6]

The trade-off typically studied in principal-agent models is between incentives and insurance. Temporal incentives can impose risk on the agent, since he will get low wages if he faces

6. Although incentive schemes have been studied extensively in organizational and mechanism-design literatures, to our knowledge, this literature has not examined temporal incentive schemes.

unusually high task costs early on. Because we wish to focus solely on the procrastination issue, we will assume that the agent is risk-neutral, so insurance is not an issue.

Suppose that the task can be completed in any period $t \in \{1,2, \ldots, T\}$, where T can be finite or infinite. The principal prefers to have the task done sooner because she faces a cost of delay. The principal's exogenously determined gross payoff schedule is $\mathrm{X} = (X_1, X_2, \ldots, X_T)$, where she gets payoff X_t if the agent completes the task in period t. The marginal delay cost is captured by $x_t^\Delta \equiv X_t - X_{t+1} > 0$ for all t. We often assume a *stationary delay cost* x^Δ, where $T = \infty$ and $x_t^\Delta = x^\Delta$ for all t.[7]

Although the principal prefers to have the task done sooner rather than later, there can be benefits to delay if it yields a lower task cost to the agent: lowering the agent's expected task cost allows the principal to pay a smaller expected wage (i.e., it will be easier to satisfy the agent's participation constraint). In period t the *task cost* to the agent, c_t, is drawn from a stationary distribution C with support $[\underline{c},\bar{c}]$, $\underline{c} \geq 0$, and cumulative distribution function $F(c_t)$. In this section we assume that $F(\underline{c}) = 0$ for simplicity, but the results all hold if we relax this assumption.[8] The task cost is meant to capture any immediate disutility to the agent arising from completing the task, including opportunity cost.

In period t the agent observes the task-cost realization c_t before choosing whether or not to perform the task. The agent's behavior can be described by a strategy $\mathrm{s} \equiv (s_1,s_2, \ldots, s_T)$, which is a vector of cutoff costs such that the agent completes the task in period t if and only if $c_t \leq s_t \in [\underline{c},\bar{c}]$.[9] Before introducing temporal incentive schemes, we characterize first-best efficient behavior,

7. In the case of $T = \infty$, we assume that the principal's payoff if the agent never completes the task is less than X_t for all t. We can think of this as the principal receiving X_1 in period 1, and paying x_τ^Δ in each period τ that the task is delayed. This assumption implies that the principal cannot avoid losses relative to X_1 by inducing the agent never to complete the task.

8. This assumption implies that "completing the task when $c_t \leq \underline{c}$" is equivalent to "waiting"—not doing the task for sure. Without this assumption, we have to define an action to represent waiting (as we do in Sections IV and V).

9. By defining the strategy this way, and not allowing each cutoff s_t to be a function of the history of task costs $(c_1,c_2, \ldots, c_{t-1})$, we are restricting the set of possible strategies that the agent could employ. This simplification is unrestrictive for TCs and naifs. In each period, both TCs and naifs choose the continuation strategy that maximizes their continuation payoffs, and continuation payoffs are independent of past task costs. For sophisticates, in contrast, this simplification can be restrictive, as we discuss in a footnote in Section VI.

which we denote by $\gamma^* \equiv (\gamma_1^*,\gamma_2^*, \ldots ,\gamma_T^*)$. Throughout, we denote a generic strategy by s and a specific strategy (e.g., the efficient strategy or an "equilibrium" strategy) by γ.

In this environment there are two real costs: the delay costs incurred by the principal and the task costs incurred by the agent. Efficient behavior $\gamma^* \equiv (\gamma_1^*,\gamma_2^*, \ldots ,\gamma_T^*)$ will therefore minimize the sum of expected task costs and expected delay costs.[10] It is convenient to express γ^* recursively; that is, the efficient cutoff in period t must exactly equal the expected total costs of waiting in period t conditional on following strategy γ^* in the future. To express the expected total costs from waiting in period t, we first define a hazard function $h(\tau|t, \mathrm{s})$, which represents the probability that the agent will not complete the task before period $\tau > t$ if the agent waits in period t and follows strategy s thereafter. Then

$$h(\tau|t, \mathrm{s}) = \begin{cases} 1 & \text{if } \tau = t + 1 \\ \displaystyle\prod_{i=t+1}^{\tau-1} (1 - F(s_i)) & \text{if } \tau > t + 1. \end{cases}$$

Let $\chi^t(\mathrm{s})$ be the *expected delay cost* incurred by the principal if the agent waits in period t and follows strategy s thereafter (so the expected gross payoff to the principal if the agent waits in period t is $X_t - \chi^t(\mathrm{s})$). Then

$$\chi^t(\mathrm{s}) = \sum_{\tau=t+1}^{T} h(\tau|t,\mathrm{s})x_{\tau-1}^{\Delta}.$$

Let $\zeta^t(\mathrm{s})$ be the *expected task cost* incurred by the agent from waiting in period t and following the strategy s thereafter. Then,

$$\zeta^t(\mathrm{s}) = \sum_{\tau=t+1}^{T} h(\tau|t,\mathrm{s})F(s_\tau)E(c|c \leq s_\tau),$$

where $E(c|c \leq s) \equiv 1/F(s) \int_{\underline{c}}^{s} c\, dF(c)$.

In period t the agent should complete the task if the known task cost c_t is less than the total expected costs from waiting,

10. Since wages are merely transfer payments between the two parties, γ^* maximizes the sum of the principal's payoff and the agent's payoff. Moreover, all points on the Pareto frontier involve the agent's following strategy γ^*. We assume throughout the paper that γ^* exists and is (as we define it below) "unique" in the sense that all efficient strategies yield equivalent behavior. These conditions hold in all cases we have considered, and we suspect they hold without loss of generality, but we have not proved any such results.

which implies that γ^* must satisfy for each $t < T$:

$$\gamma_t^* = \begin{cases} \chi^t(\gamma^*) + \zeta^t(\gamma^*) & \text{if } \underline{c} \leq \chi^t(\gamma^*) + \zeta^t(\gamma^*) \leq \bar{c} \\ \underline{c} & \text{if } \chi^t(\gamma^*) + \zeta^t(\gamma^*) < \underline{c} \\ \bar{c} & \text{if } \chi^t(\gamma^*) + \zeta^t(\gamma^*) > \bar{c}. \end{cases}$$

The principal would like to require the agent to complete the task in period t if and only if $c_t \leq \gamma_t^*$; but since the principal cannot observe c_t, the contract cannot specify a wage contingent on the task cost, but only on when the agent completes the task. Denote a *temporal incentive scheme* by $W = (W_1, W_2, \ldots, W_T)$, where the agent receives wage W_t if he completes the task in period t. The *incremental wage* w_t^Δ is defined by $w_t^\Delta \equiv W_t - W_{t+1}$.[11] Since the agent is risk-neutral, only the incremental wage w_t^Δ affects behavior. The level of wages matters only for the question of whether the agent accepts the contract.

The agent has present-biased preferences, as described in Section II, with $\delta = 1$. We consider two types of agents. TCs have $\beta = 1$, so they have standard time-consistent preferences. (*TC* stands for time-consistent.) *Naifs* have $\beta < 1$, but they are naive and therefore believe they will behave like TCs beginning next period. We examine TC behavior both as a benchmark against which to compare naifs' behavior and because it represents naifs' perceived future behavior. In addition to being time-inconsistent, naifs are also more impatient than TCs. It will become clear that the main results are driven by the time inconsistency, and *not* by the relative impatience.

To use the preferences described in Section II, we must convert wages and task costs into instantaneous utilities. Our crucial assumption is that task costs are incurred immediately whereas wages are received sometime in the future. Consequently, a naive agent gives the current task cost more weight in his decision than future wages, causing a tendency to procrastinate.[12] Formally, we assume that if the agent completes the task

11. For the case of $T = \infty$, we make an assumption comparable to the one we made for the principal: if the agent never completes the task, his payoff is below W_t for all t. An interpretation is that from time to time the agent must "settle up" incurred delay penalties.

12. We define "procrastination" to be instances where a person delays when from a long-run perspective (or in fact from any prior perspective) he would prefer to complete the task.

in period $\hat{t}$, then his instantaneous utilities are $u_{\hat{t}} = -c_{\hat{t}}$, $u_{\hat{t}+1} = W_{\hat{t}}$ and $u_t = 0$ for all $t \notin \{\hat{t}, \hat{t} + 1\}$.[13]

To examine how the agent will behave given an incentive scheme W, we use the concept of *perception-perfect strategies* introduced by O'Donoghue and Rabin [1999]. Rather than give the general definition, we describe the implications for the two types of agents in the model here. For TCs, a perception-perfect strategy is the standard, simple decision-theoretic prediction: at all times TCs maximize their expected utility given their current information, so they complete the task now if the utility from doing so is higher than the expected utility from waiting. Naifs similarly compare their utility from completing the task now with their *perceived* expected utility from waiting; but because naifs think they will behave like TCs in the future, their perceived utility is systematically wrong (and, in particular, overoptimistic). That is, naifs misperceive their future behavior and consequently their future utility from waiting.

As it was for the efficient strategy γ^*, it is convenient to express perception-perfect strategies recursively. The agent's cutoff in period t must exactly equal his perceived total expected cost from waiting in period t. From the agent's perspective, there are two costs associated with waiting in period t: the expected task cost he will incur in the future and any lost wages. If the agent perceives that he will follow strategy s in the future, then the expected task cost from waiting is $\zeta^t(\mathrm{s})$. Let $p^t(\mathrm{s})$ be the *expected wage cost* from waiting in period t when the agent perceives he will follow strategy s in the future (so the expected wage if the agent waits in period t is $W_t - p^t(\mathrm{s})$). Then

$$p^t(\mathrm{s}) = \sum_{\tau=t+1}^{T} h(\tau|t,\mathrm{s}) w^{\Delta}_{\tau-1}.$$

Let $\hat{\gamma} \equiv (\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_T)$ denote a perception-perfect strategy for TCs. A TC does not discount future costs, and knows exactly how he will behave in the future. Hence, $\hat{\gamma} \equiv (\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_T)$ must

13. Even if wages are received immediately, they will *effectively* be delayed if the agent cannot enjoy their benefits immediately; wages will be given exaggerated weight in decisions only if they affect immediate consumption. We discuss the use of immediate rewards such as breaks or parties in Section VII. Our assumption that the wage is received in period $\hat{t} + 1$ is merely for simplicity; because we have assumed no time-consistent discounting, the agent does not care when in the future he receives the wage.

satisfy for each $t < T$,

$$\hat{\gamma}_t = \begin{cases} p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma}) & \text{if } \underline{c} \leq p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma}) \leq \bar{c} \\ \underline{c} & \text{if } p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma}) < \underline{c} \\ \bar{c} & \text{if } p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma}) > \bar{c}. \end{cases}$$

Let $\gamma \equiv (\gamma_1, \gamma_2, \ldots, \gamma_T)$ denote a perception-perfect strategy for naifs. Naifs discount future costs by factor β. Moreover, they believe that they will behave like TCs in the future and follow strategy $\hat{\gamma}$, so their *perceived* expected task cost from waiting is $\zeta^t(\hat{\gamma})$, and their *perceived* expected wage cost from waiting is $p^t(\hat{\gamma})$. Hence, $\gamma \equiv (\gamma_1, \gamma_2, \ldots, \gamma_T)$ must satisfy for each $t < T$,[14]

$$\gamma_t = \begin{cases} \beta[p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma})] & \text{if } \underline{c} \leq \beta[p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma})] \leq \bar{c} \\ \underline{c} & \text{if } \beta[p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma})] < \underline{c} \\ \bar{c} & \text{if } \beta[p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma})] > \bar{c}. \end{cases}$$

The final components of the model are the structure of ex ante negotiation and the participation constraint. As our focus is efficiency, which is unaffected by the absolute level of wages, the participation constraint will not play a prominent role in our analysis. However, with naive time-inconsistent agents, some important issues arise that need to be addressed.

For simplicity, we assume that the principal proposes a contract which the agent can either accept or reject. If the agent accepts, he will be compensated according to the incentive scheme. If the agent rejects, then there is no contract, and he gets utility $\bar{U} = 0$. Which contracts will the agent accept? For TCs the answer is clear. Given an incentive scheme W, TCs correctly predict their future behavior (i.e., $\hat{\gamma}$) and therefore accept any contract that pays an ex ante expected wage larger than the ex ante expected task cost. For naifs, several issues arise. Since naifs have time-inconsistent preferences, it matters whether the contract is signed in period 1 or prior to period 1. We assume ex ante negotiations occur prior to period 1, so the agent's "long-run utility" is relevant when signing the contract. In our model, we can interpret this as ex ante negotiations occurring in period 0, where the agent's preferences are described by U^0.

14. $\hat{\gamma}$ and γ as defined in the text are unique, but there could be other strategies that would yield identical observed behavior (and would be perception-perfect strategies under a more general definition). Throughout, we refer to $\hat{\gamma}$ and γ as *the* perception-perfect strategies. For $T = \infty$, under some W $\hat{\gamma}$ or γ or both may not exist; but for all W we consider they do exist.

More importantly, naifs incorrectly perceive future behavior, and consequently will generally be overoptimistic about their utility from signing a contract. Should the participation constraint for naifs be based on the utility naifs perceive at the time they sign, or on the average utility they actually get from the contract? In other words, should we use utility from an ex ante view ("perceived utility") or an ex post view ("experienced utility")? We primarily take the point of view that the participation constraint for naifs should be based on experienced utility of the agent. But we first consider the perceived-utility perspective. In that case, naifs could be exploited as a "money pump": The principal could hire a naif to do some task and get the naif to *pay her* a large sum of money (i.e., earn a large negative wage) to do the task. Consider the following example.

Example 1

Consider a task that is "useless" to the principal (i.e., $X_t = 0$ for all t), and suppose that the distribution of task costs for the agent has support $[\underline{c},\bar{c}]$, with mean E_C, and that $\underline{c}/\beta > \bar{c}$. Consider the incentive scheme $W = (E_C, E_C - w, \ldots, E_C - (T-1)w)$ with $\bar{c} \leq w + E_C < \underline{c}/\beta$. Since $\bar{c} \leq E_C + w$, TCs will complete the task for sure in period 1. Hence, naifs' perceived utility from the contract is $W_1 - E_C = 0 = \overline{U}$, so naifs would be willing to sign the contract. However, since $\underline{c} > \beta(E_C + w)$, in fact naifs procrastinate until period T. Hence, the principal could use this useless task to bilk an arbitrarily large amount of money from the agent, since $\lim_{T\to\infty} W_T = -\infty$.

The general intuition reflected in Example 1 is that when the agent mispredicts her own future behavior, the principal can bilk the agent by offering an incentive scheme that punishes delay sufficiently leniently so as to lull the agent into severe and costly procrastination. For a degenerate enough task-cost distribution such as that in Example 1, it is possible for the principal to offer a contract that the agent will accept under the premise that he will complete the task immediately, when in fact he will for sure procrastinate until the very end. By proposing a contract with T arbitrarily large, the principal can guarantee an arbitrarily large negative wage. For general task-cost distributions, creating such a stark contrast between beliefs and behavior will not be possible. Nonetheless, the following lemma establishes that the principal can bilk arbitrarily large sums of money from naifs as long as the task cost is bounded away from zero.

LEMMA 1. Suppose that there exists a task such that $\underline{c} > 0$. Then for any $\overline{W} < 0$ and any $\beta < 1$, there exists a contract W that naifs perceive to yield expected utility $U^p \geq 0$, but that actually pays less than $\overline{W}$.

All proofs are in the Appendix. The intuition for Lemma 1 is similar to the reasoning in Example 1. For any cost distribution the principal can create an incentive scheme such that despite believing they will complete the activity with positive probability in all periods, naifs in fact procrastinate until period T. To do this, the principal must make the incremental punishment for delay large enough that the naif believes he will complete the task relatively soon, yet small enough that the naif will actually delay until period T. For T sufficiently large, naifs believe that they will complete the task long before period T, and as a result they will accept a contract with a large negative wage in period T.

Lemma 1, taken literally, says that principals can become arbitrarily rich by hiring naive agents not for any useful function, but rather to exploit their overoptimism. While it *is* plausible that firms take some advantage of such bilking opportunities, we think studying efficiency-oriented temporal incentive schemes is probably more important than studying bilking-oriented schemes. There are several reasons for this perspective. First of all, in some situations a principal might use efficiency-oriented contracts even when bilking is possible. There are clearly settings in which a principal would find it optimal to sacrifice some efficiency in order to bilk an agent—e.g., when the principal can hire only one agent and this agent can complete only one task. But in other settings—e.g., if the supply of agents is large, or if a given agent is not time-constrained—it may instead be optimal for the principal to offer one contract to get a task efficiently completed and other contracts to bilk people.

In addition, there are several reasons that the principal may be unable to bilk people. For instance, reputational pressures may induce firms to offer incentive contracts that are ex post acceptable to agents, which would imply that firms wish to induce efficient behavior. A firm (i.e., the principal) will likely have many tasks that it needs completed. To maintain a pool of willing agents, the firm might need to develop a reputation for making agents on average ex post pleased with the outcomes, which means that the expected *experienced* utility for the agent must be at least $\overline{U}$. Such a "reputation constraint" for principals would imply that, effectively, the participation constraint for naifs is based on experienced utility. If working at a firm involves completing, say, one hundred tasks over the course of a year, and the most important "participation constraint" by employees is not whether they wish to take a job but whether they wish to change jobs, then the experienced-utility perspective is clearly more appropriate.

Another reason to downplay the principal's bilking opportunities is somewhat more subtle, and somewhat slippery to model formally, but seems psychologically realistic. A pattern for many psychological biases is the coexistence of day-to-day errors with a "meta-awareness" of these errors. In terms of procrastination, people seem to be "meta-sophisticated" about their tendency to procrastinate, and yet they exhibit day-to-day naivete. If so, then considering ex post efficiency may not be a bad approximation: agents may sign contracts completely aware of their general tendency to procrastinate, and yet on a day-to-day and case-by-case basis, they may be overoptimistic about avoiding future procrastination. This conceptualization also accords well with our assumption that the principal is aware of the procrastination problem, because it allows that the principal is no more aware than the agent himself, only that they are both meta-aware.

Finally, the often invoked (if rarely formalized) presumption that efficient institutions and production schemes tend to survive over time—even if people are not aware of why they work—may suggest a focus on efficiency contracts rather than bilking contracts. Principals may use deadline contracts, and employees may tend to accept such contracts, even if neither party knows why they work.

In any event, for the rest of this paper we assume the participation constraint for naifs is based on experienced utility, and therefore focus on efficient contracts. We next ask whether temporal incentive schemes can induce efficient behavior when the principal has complete information about the agent, i.e., when the principal knows both the agent's propensity to procrastinate, β, and his distribution of task costs.

Since TCs are time-consistent, TC behavior minimizes the sum of the expected task cost and the expected wage cost. Since efficient behavior minimizes the sum of the expected task cost and the expected delay cost, TCs behave efficiently if the expected wage cost is identical to the expected delay cost; that is, if the incentive scheme internalizes the principal's payoff schedule.

With a stationary delay cost, this means a stationary incentive scheme that reflects the true delay cost. We formalize this intuition in Proposition 1.

PROPOSITION 1. TCs behave efficiently under any incentive scheme W satisfying $w_t^{\Delta} = x_t^{\Delta}$ for all t; and if X has a stationary delay cost x^{Δ}, TCs behave efficiently under any stationary incentive scheme with $w^{\Delta} = x^{\Delta}$.

Proposition 1 implies that TCs behave efficiently if the incentive scheme internalizes the principal's delay costs. In contrast, for naifs such an incentive scheme would induce inefficiently low cutoffs, because naifs tend to procrastinate. When the principal has complete information about the agent, however, she can in fact induce efficient behavior for naifs with an incentive scheme that exactly counteracts the tendency to procrastinate. With a stationary delay cost, this means a stationary incentive scheme reflecting a delay cost larger than the true delay cost. We formalize this intuition in Proposition 2.

PROPOSITION 2. For every X, C, and $\beta < 1$,

(i) There exists an incentive scheme W such that naifs behave efficiently, and

(ii) If $\gamma_t^* < \bar{c}$ for all t, then any such W satisfies $w_t^{\Delta} > x_t^{\Delta}$ for all $t \leq T$; and if X has a stationary delay cost x^{Δ}, there exists a stationary incentive scheme W with $w^{\Delta} > x^{\Delta}$ such that naifs behave efficiently.

The following example demonstrates the results in Propositions 1 and 2, and also illustrates that a "steeper" incentive scheme for naifs relative to TCs implies that the initial wage for naifs must be larger in order to satisfy the participation constraint. Since both types are induced to behave efficiently, they face the same expected task cost.

Example 2

Suppose that $T = \infty$, $x^{\Delta} = 1/32$, and C is distributed uniformly on $[\underline{c}, \underline{c} + 1]$, so $F(c) = c - \underline{c}$ for $c \in [\underline{c}, \underline{c} + 1]$.

Efficiency: Clearly, the efficient cutoff cost will be stationary. Let $\gamma^* \equiv (\gamma^*, \gamma^*, \ldots)$ denote the efficient strategy. For each t, $\chi^t(\gamma^*) = x^{\Delta}/F(\gamma^*) = 1/32/(\gamma^* - \underline{c})$ and $\zeta^t(\gamma^*) = E(c \mid c \leq \gamma^*) = (\underline{c} + \gamma^*)/2$. Hence, $\gamma^* = 1/32/(\gamma^* - \underline{c}) + (\underline{c} + \gamma^*)/2$, implying that $\gamma^* = \underline{c} + 1/4$.

TCs: Proposition 1 establishes that a stationary incentive scheme with incremental wage $w^{\Delta} = x^{\Delta} = 1/32$ will induce efficient behavior. The principal will offer an incentive contract such that the agent behaves efficiently and the expected wage equals the expected task cost. Hence, the incentive scheme must satisfy $W_1 = (1 - F(\gamma^*))p^1(\gamma^*) + \zeta^0(\gamma^*)$. Since $p^1(\gamma^*) = w^{\Delta}/(\gamma^* - \underline{c}) = 4w^{\Delta}$ and $\zeta^0(\gamma^*) = (\underline{c} + \gamma^*)/2 = \underline{c} + 1/8$, $W_1 = 3w^{\Delta} + \underline{c} + 1/8 = \underline{c} + 7/32$. Hence, the first-best contract for TCs is $W = (\underline{c} + 7/32,\ \underline{c} + 6/32,\ \underline{c} + 5/32,\ \underline{c} + 4/32, \ldots)$.

Naifs: Suppose that $\beta = 1/2$. Proposition 2 establishes that a stationary incentive scheme can induce efficient behavior for naifs, so that there is an incremental wage w^{Δ} that will induce efficient behavior. Given w^{Δ}, $\hat{\gamma}$ will clearly be stationary (i.e., cutoff $\hat{\gamma}$ in all periods) and satisfies $\hat{\gamma} = w^{\Delta}/(\hat{\gamma} - \underline{c}) + (\underline{c} + \hat{\gamma})/2$, or $\hat{\gamma} = \underline{c} + \sqrt{2w^{\Delta}}$ (as long as $w^{\Delta} \leq 1/2$ so $\hat{\gamma} \leq \underline{c} + 1$). In general, γ satisfies $\gamma_t = \beta\hat{\gamma}_t$ for all t, so $\gamma_t = \gamma^*$ for all t if $\gamma^* = \beta(\underline{c} + \sqrt{2w^{\Delta}})$ or $w^{\Delta} = 1/2\,(\underline{c} + 1/2)^2$ (and $w^{\Delta} \leq 1/2$ as long as $\underline{c} \leq 1/2$). As for TCs, the incentive scheme must satisfy $W_1 = (1 - F(\gamma^*))p^1(\gamma^*) + \zeta^0(\gamma^*) = 3w^{\Delta} + \underline{c} + 1/8$. So for any $\underline{c} \leq 1/2$, the first-best contract for naifs is described by $W_1 = 3w^{\Delta} + \underline{c} + 1/8$ and $w^{\Delta} = 1/2\,(\underline{c} + 1/2)^2$.

The contracts used to induce efficiency in Example 2 vary according to $\underline{c}$, which determines how high the average task cost is. Figure I illustrates the incentive schemes that will be chosen for both naifs and TCs, for two different values of $\underline{c}$, a low cost of $\underline{c} = 0$ and a high cost of $\underline{c} = 1/2$.[15] For each task-cost distribution, the optimal incentive scheme is steeper for naifs than for TCs; therefore, naifs must have a larger intercept to satisfy the participation constraint. Another feature of Figure I is crucial for the intuition of Section IV: for TCs, changing the task-cost distribution changes the intercept (i.e., the participation constraint) but *not* the slope. In contrast, for naifs, changing the task-cost distribution changes both the intercept *and* the slope. Intuitively, higher average task costs imply a greater propensity to procrastinate, and steeper incentives are required to overcome procrastination. This difference, that changing the average task cost affects the optimal incentives for naifs but not for TCs, implies a qualitative difference in how principals deal with uncertainty over average task costs for the two types of agents.

We conclude this section with a brief example illustrating

15. For $\underline{c} = 0$, the first-best contract is $W = (7/32, 6/32, 5/32, \ldots)$ for TCs and $W = (4/8, 3/8, 2/8, \ldots)$ for naifs; for $\underline{c} = 1/2$, the first-best contract is $W = (23/32, 22/32, 21/32, \ldots)$ for TCs and $W = (17/8, 13/8, 9/8, \ldots)$ for naifs.

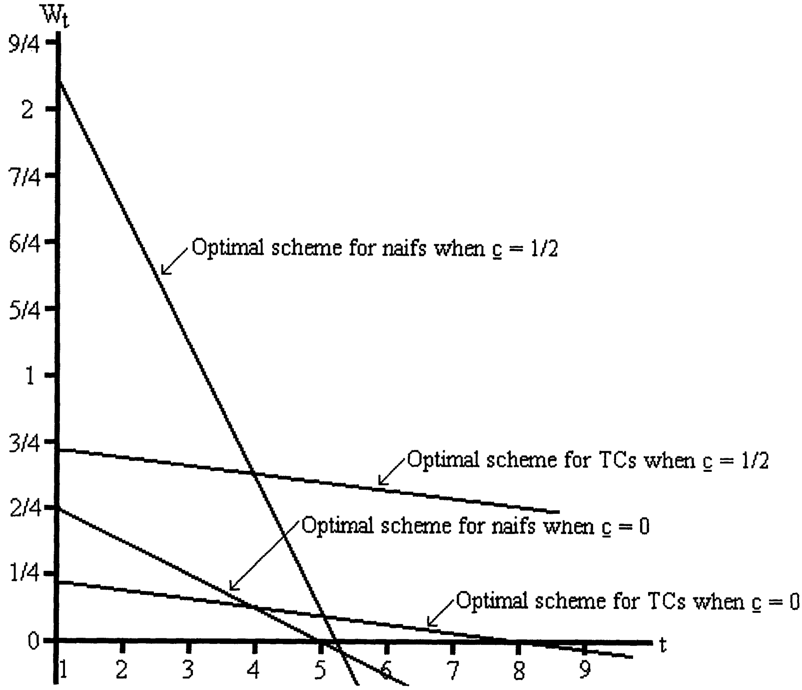

FIGURE I
Optimal Incentive Schemes for Example 2

another implication of Proposition 2: even if a principal faces no delay costs before some absolute deadline, she may have to impose incremental punishments for delay even before the deadline to combat inefficient procrastination.

Example 3

Suppose that the principal faces a pure deadline: $X_t = K > 0$ for all $t \leq D$, $X_t = 0$ for all $t > D$. For TCs, the optimal incentive scheme will clearly punish the agent by K if and only if he delays past period D (i.e., $w_t^{\Delta} = 0$ for all $t < D$ and $w_t^{\Delta} = K$ for $t = D$). For naifs, however, the optimal incentive scheme may punish delay even before the deadline (i.e., $w_t^{\Delta} > 0$ for $t < D$). That is, the principal must "falsely" punish the agent for delay if she wants to induce efficient behavior.

In periods where there is no delay cost, efficiency still may call for completing the task if the task-cost realization is particularly small. Because naifs tend to procrastinate, to induce efficiency the principal must punish delay even though she does not care at all directly about delay. Even if a professor feels that it is only important for a student to understand the material by exam time, she may still want to grade problem sets throughout the semester. Although such a policy punishes the few students who *would* successfully learn the material with little effort at the last moment, it benefits the many students who would put off learning the material until it becomes so late that they cannot adequately do so.

IV. Heterogeneous Propensities to Procrastinate

In this section we assume that the principal is uncertain about the agent's propensity to procrastinate. There are two ways in which agents can differ in their propensities to procrastinate. First, agents can differ in their *inherent* propensities to procrastinate, β. Second, agents can differ in their *induced* propensities to procrastinate: for a fixed β, different task-cost distributions induce different propensities for procrastination. While our model would be similar under either variant of procrastination, we focus solely on the latter for two reasons. First, we consider it a result of direct interest that the scale of the disparity between efficient waiting and inefficient procrastination can depend on the environment, rather than solely on the agent's inherent procrastinatory tendencies. Second, we suspect that for long-employed agents, uncertainty over the environment can persist while uncertainty over the inherent propensity to procrastinate may not. Consider a single employee who is given many tasks over time to complete. Eventually his supervisor may figure out his inherent propensity to procrastinate, β. But if the agent must perform a long series of idiosyncratic tasks of uncertain difficulty, then case-by-case uncertainty over his propensity to procrastinate may remain in the long run.

We suppose the task-cost distribution C is unknown to both parties in period 0 when they sign the contract, and is revealed to the agent but not to the principal sometime before period 1. The specific C realized determines an agent's "type." We assume in this section that there are two types of agents who differ only in their mean task costs, with the exact same distribution of task costs around the mean. Agents who face a higher mean task cost are more prone to procrastinate. Perhaps the best interpretation of this situation is that the principal (and, ex ante, the agent) is unsure of how hard the agent will find the task, but that she has a rough sense of the day-to-day variance in the opportunity cost for the agent to do the task. In Section V we consider the case where not only may the means of the task costs differ, but also the distributions around the means.

Our results in Section III imply that for a time-consistent agent, private information about the task-cost distribution does not cause a problem: TCs behave efficiently no matter the distribution of task costs, as long as the incentive scheme internalizes the principal's preferences. As illustrated by Example 2 and Figure I, the slope of the optimal incentive scheme for TCs does not depend on the task-cost distribution C. For naifs, on the other hand, a problem arises because their optimal incentive scheme depends on the distribution of task costs. As illustrated by Example 2 and Figure I, agents with higher average task costs (who are more prone to procrastinate) require "steeper" incentive schemes. Hence, when the agent's propensity to procrastinate is unknown, efficient incentives must be steep enough to prevent high-cost types from procrastinating, yet shallow enough to induce low-cost types to wait when waiting is efficient.

In this section and the next, we consider the case in which the principal's incremental cost of delay is stationary: $X_t - X_{t+1} = x^{\Delta}$ for all t, with $T = \infty$. Our focus on stationary environments allows us to highlight an interesting nonstationarity result. Proposition 1, presented earlier, established that when delay costs are stationary the optimal incentive scheme for TCs is stationary, and Proposition 2 established that the optimal "full-information" incentive scheme for naifs is also stationary. We show, however, that if the agent has private information about the distribution of his task costs, then the second-best optimal incentive scheme for naifs is generally *not* stationary. Rather, it is a "deadline scheme": An agent is initially punished only mildly for delay, but there is a date after which punishment for delay becomes more severe.

In this section and the two that follow, we use the term "deadlines" to describe incentive schemes with discrete jumps in how severely the agent is punished for delay. Two comments are in order. First, although such discrete jumps are second-best optimal in our simple model, a more general model would not generate such clear deadlines. Rather, the more general qualitative result is that second-best optimal incentive schemes should impose increasingly severe punishments over time. Second, real-world deadline contracts often take a different, simpler form: the agent is punished each time he misses a deadline (with no punishments between deadlines). We discuss the relationship between our model and such simple deadlines in the concluding section.

For tractability we consider a highly simplified model with two types of agent, where each type faces two possible task costs. Let $i \in \{L,H\}$ denote an agent's type, and let π denote the ex ante probability that the agent is type L (so $1 - \pi$ is the ex ante probability that the agent is type H). The cost distribution C_i for each $i \in \{L,H\}$ is

$$c = \begin{cases} \underline{c}_i \equiv c_i - k & \text{with probability } \tfrac{1}{2} \\ \bar{c}_i \equiv c_i + k & \text{with probability } \tfrac{1}{2}, \end{cases}$$

where $c_H > c_L$ and $k > 0$. Hence, the H agent has a higher average task cost than the L agent, and therefore the H agent is more prone to procrastinate. In this 2×2 model, the agent could be a high-cost type or a low-cost type, and each type can have a high-cost realization or a low-cost realization. To clarify our discussion, we use the following terminology to describe the three possible plans that the agent might employ in any given period:[16]

Terminology	**Action**
do it for sure (d):	complete task if $c = \underline{c}_i$ or $c = \bar{c}_i$
be selective (s):	complete task only if $c = \underline{c}_i$
wait for sure (w):	do not complete task.

We redefine strategies in terms of these three plans. A strategy is therefore $\mathrm{s} \equiv (s_1, s_2, \ldots)$ such that $s_t \in \{d,s,w\}$ for all t. Since X and W are the same for both types, we can define $h(\tau|t,\mathrm{s})$, $\chi^t(\mathrm{s})$, and $p^t(\mathrm{s})$ exactly as in Section III, where we redefine F to be $F(d) = 1$, $F(s) = \tfrac{1}{2}$, and $F(w) = 0$. Since the two types face different task-cost distributions, however, the expected task cost is type-dependent. Let $\zeta_i^t(\mathrm{s})$ be identical to $\zeta^t(\mathrm{s})$ except that $E_i(c|c \leq s_t)$ replaces $E(c|c \leq s_t)$, where we define $E_i(c|c \leq s_t)$ as

$$E_i(c|c \leq s_t) \equiv \begin{cases} c_i & \text{if } s_t = d \\ c_i - k & \text{if } s_t = s \\ 0 & \text{if } s_t = w. \end{cases}$$

16. The fourth possible plan, complete the task only if $c = \bar{c}_i$, would obviously be neither optimal nor chosen.

Given X, let $\gamma^{*i} \equiv (\gamma_1^{*i},\gamma_2^{*i},\ldots)$ be the efficient strategy for type $i \in \{L,H\}$. Similarly, given W, let $\hat{\gamma}^i \equiv (\hat{\gamma}_1^i,\hat{\gamma}_2^i,\ldots)$ and $\gamma^i \equiv (\gamma_1^i,\gamma_2^i,\ldots)$ be the perception-perfect strategies of type $i \in \{L,H\}$ for, respectively, TCs and naifs. Then, for all t, γ^{*i}, $\hat{\gamma}^i$, and γ^i satisfy[17]

$$\gamma_t^{*i} = \begin{cases} d & \text{if } c_i + k \le \chi^t(\gamma^{*i}) + \zeta_i^t(\gamma^{*i}) \\ s & \text{if } c_i + k \ge \chi^t(\gamma^{*i}) + \zeta_i^t(\gamma^{*i}) \end{cases}$$

$$\hat{\gamma}_t^i = \begin{cases} d & \text{if } c_i + k \le p^t(\hat{\gamma}^i) + \zeta_i^t(\hat{\gamma}^i) \\ s & \text{if } c_i + k \ge p^t(\hat{\gamma}^i) + \zeta_i^t(\hat{\gamma}^i) \end{cases}$$

$$\gamma_t^i = \begin{cases} d & \text{if } c_i + k \le \beta[p^t(\hat{\gamma}^i) + \zeta_i^t(\hat{\gamma}^i)] \\ s & \text{if } c_i - k \le \beta[p^t(\hat{\gamma}^i) + \zeta_i^t(\hat{\gamma}^i)] \le c_i + k \\ w & \text{if } c_i - k \ge \beta[p^t(\hat{\gamma}^i) + \zeta_i^t(\hat{\gamma}^i)]. \end{cases}$$

The assumption that task-cost distributions differ in only their means is important for two reasons. First, efficient behavior depends on the distribution of task costs but *not* their mean, so efficient behavior will be the same for both types (i.e., $\gamma^{*H} = \gamma^{*L} \equiv \gamma^*$). Second, TC behavior also depends on the distribution of task costs but *not* the mean, so TC behavior will be the same for both types (i.e., $\hat{\gamma}^H = \hat{\gamma}^L \equiv \hat{\gamma}$). This second result has an important implication: naifs perceive that they will behave like TCs in the future, so L's and H's perceive the same continuation strategies. Hence, for any incentive scheme, we have $p^t(\hat{\gamma}) + \zeta_H^t(\hat{\gamma}) - c_H = p^t(\hat{\gamma}) + \zeta_L^t(\hat{\gamma}) - c_L$ for all t. That is, both types of agent have the exact same perceptions of how delaying will affect their net gain or loss in expected cost.[18]

When there is a stationary delay cost x^Δ, efficient behavior calls for *either* doing it for sure in all periods *or* being selective in all periods (i.e., $\gamma^* = (d,d,\ldots)$ or $\gamma^* = (s,s,\ldots)$). Analysis of the first case is trivial, since efficiency can be achieved simply with a

17. As written, these definitions are ambiguous because they do not specify what the agent does when indifferent between two actions. In the analysis below, we follow the incentive-design literature by assuming that when indifferent the agent behaves as the principal would like him to behave.

18. This second result also explains why we do not consider direct mechanisms—contracts where the agent reveals his type to the principal, allowing the incentive scheme to be type-dependent. For naifs, it is not possible to differentiate the types: for any incentive scheme, high-cost types and low-cost types perceive the same future behavior, and hence there cannot be a "separating equilibrium." Moreover, direct mechanisms seem unreasonable in this environment. Our analysis may be most applicable to situations where an agent is hired to complete many tasks over time, or when many agents in an organization are given the same incentive schemes. In such environments, renegotiation of each incentive scheme to take account of case-specific information seems unrealistic.

very steep incentive scheme in the first period. We focus instead on the case where first-best efficiency calls for being selective in all periods, which holds as long as $k > x^{\Delta}$.

Before examining how a principal should deal with multiple types, we first characterize which incentive schemes induce efficiency for a specific type. Lemma 2 characterizes the behavior of naifs under stationary incentive schemes. For small incremental wages, naifs wait for sure in all periods; for moderate incremental wages, naifs are selective in all periods; and for large incremental wages, naifs do it for sure in all periods.

LEMMA 2. Suppose that there is a stationary incentive scheme with incremental wage w^{Δ}. Then for agent $i \in \{L,H\}$:

$$\gamma^i = (w,w,\ldots) \text{ if and only if } w^{\Delta} < \underline{w}_i;$$

$$\gamma^i = (s,s,\ldots) \text{ if and only if } \underline{w}_i \leq w^{\Delta} \leq \overline{w}_i;$$

$$\gamma^i = (d,d,\ldots) \text{ if and only if } w^{\Delta} > \overline{w}_i,$$

where $\underline{w}_i$ and $\overline{w}_i$ are given by

$$\overline{w}_i = \frac{1-\beta}{\beta} c_i + \frac{1}{\beta} k \text{ and } \underline{w}_i = \begin{cases} \dfrac{1-\beta}{\beta} c_i - \dfrac{1}{\beta} k & \text{if } c_i \geq \dfrac{1+\beta}{1-\beta} k \\ \dfrac{1-\beta}{2\beta} c_i - \dfrac{1-\beta}{2\beta} k & \text{if } c_i \leq \dfrac{1+\beta}{1-\beta} k. \end{cases}$$

The values $\underline{w}_i$ and $\overline{w}_i$ in Lemma 2 represent the minimum and maximum *stationary* incremental wages that can induce efficient behavior by a naif of type i.[19] Given the discrete nature of the model, it is not surprising that multiple incentive schemes can induce efficient behavior.

Of course, some *nonstationary* incentive schemes induce efficient behavior in this environment as well. However, it turns out that any incentive scheme that induces efficient behavior must "on average" reflect the incentives of some stationary incentive scheme that induces efficiency. To formalize this claim,

19. The equation for $\underline{w}_i$ depends on the agent's perceptions, which is why the result involves two cases. For $c_i < (1 + \beta)k/(1 - \beta)$, the agent perceives that he will follow $\hat{\gamma}^i = (s, s, \ldots)$ in the future when $w^{\Delta} = \underline{w}_i$; and for $c_i > (1 + \beta)k/(1 - \beta)$, the agent perceives that he will follow $\hat{\gamma}^i = (d,d,\ldots)$ in the future when $w^{\Delta} = \underline{w}_i$.

we first note that a constant incremental wage w^{Δ} implies that the ex ante expected wage cost of a selective agent is $\Sigma_{t=1}^{\infty}(1/2)^t w^{\Delta}$, which reduces to simply w^{Δ}. The following lemma establishes that for any incentive scheme that induces efficient behavior for type i, the ex ante expected wage cost must be in the range $[\underline{w}_i, \overline{w}_i]$.

LEMMA 3. If an incentive scheme W induces $\gamma_t^i = s$ for all t, then $\underline{w}_i \leq \Sigma_{t=1}^{\infty}(1/2)^t\, w_t^{\Delta} \leq \overline{w}_i$.

Lemmas 2 and 3 characterize the types of incentive schemes that can induce efficiency for a specific type. They also permit a simple characterization of when the principal can induce efficiency for *both* types: it follows directly from Lemmas 2 and 3 that either there exists a stationary incentive scheme that can induce efficiency for both types (when $\underline{w}_H \leq \overline{w}_L$), or no incentive scheme can induce efficiency for both types (when $\underline{w}_H > \overline{w}_L$). Which case holds depends on the magnitude of the difference between the two types in the mean task costs.

PROPOSITION 3. $\underline{w}_H \leq \overline{w}_L$ if and only if $c_H - c_L \leq 2k/(1-\beta)$. Hence, if $c_H - c_L \leq 2k/(1-\beta)$, there exists a stationary incentive scheme under which both types behave efficiently (by being selective every period). If $c_H - c_L > 2k/(1-\beta)$, then no incentive scheme can induce both types to behave efficiently (by being selective every period).

Lemma 2 establishes that each type will behave efficiently for a range of stationary incentive schemes. If the average costs of the two types are close, these ranges will overlap, so a stationary incentive scheme can induce efficient behavior for both types.[20] Otherwise, no stationary incentive scheme can induce efficient behavior for both types. Lemma 3 then establishes that any nonstationary incentive scheme that induces efficiency for a given type must "on average" reflect the incentives of some stationary incentive scheme that induces efficiency for that type. Hence, if no stationary incentive scheme can induce efficiency for both types, then no incentive scheme of any form can induce efficiency for both types.

Given that the principal cannot induce first-best efficiency for

20. This result is clearly an artifact of our assumption that the distribution of types is discrete. For full-support distributions of types and "interior" efficient behavior, the principal will be unable to induce full efficiency. The analog to Proposition 3 in such a continuous model would be a convergence result that we get closer to efficiency as the distribution of types becomes less dispersed.

both types when $c_H - c_L > 2k/(1 - \beta)$, we search for second-best optimal incentive schemes. A particular type of nonstationary incentive scheme will figure prominently in our results.

DEFINITION 1. A deadline scheme with deadline $D \geq 1$ is an incentive scheme such that there exists w^* such that $w_t^{\Delta} < w^*$ for all $t < D$ and $w_t^{\Delta} > w^*$ for all $t \geq D$.

DEFINITION 2. A $(w_A^{\Delta}, w_B^{\Delta})$-deadline scheme is a deadline scheme in which $w_t^{\Delta} = w_A^{\Delta}$ for all $t < D$ and $w_t^{\Delta} = w_B^{\Delta}$ for all $t \geq D$ for some $w_A^{\Delta} < w_B^{\Delta}$.

As we use the term here, a deadline scheme is an incentive scheme where the incremental wage before some period D is everywhere smaller than the incremental wage after period D. A $(w_A^{\Delta}, w_B^{\Delta})$-deadline scheme is a two-part linear incentive scheme where w_A^{Δ} and w_B^{Δ} are the respective slopes. We use the label "deadline schemes" because agents are punished relatively lightly for delay up to some date D (the "deadline"), after which further delay leads to more severe punishment.

To intuit the advantages of deadline schemes, consider the use of stationary incentive schemes when $c_H - c_L > 2k/(1 - \beta)$. Since efficiency calls for both L's and H's to be selective in all periods, clearly the best stationary incentive scheme should induce either L's or H's to be selective each period. But if L's are selective every period, then H's never complete the task (which occurs when $\underline{w}_L \leq w^{\Delta} \leq \overline{w}_L < \underline{w}_H$); and if H's are selective every period, the L's complete the task immediately (which occurs when $\overline{w}_L < \underline{w}_H \leq w^{\Delta} \leq \overline{w}_H$). Hence, the best possible stationary incentive scheme has the limitation that either H's procrastinate forever or L's inefficiently complete the task too early.

Deadline schemes have the potential to mitigate these problems. The initial small incremental wage gives L's some efficiency value of waiting, while the eventual large incremental wage prevents H's from procrastinating forever. As a preliminary step in establishing the second-best optimality of deadline schemes, Lemma 4 below implies that under any incentive scheme, for any given period either H's wait for sure or L's do it for sure.

LEMMA 4. Suppose that $c_H - c_L > 2k/(1 - \beta)$. Then for any t, $\gamma_t^H \neq w$ implies that $\gamma_t^L = d$.

Lemma 4 reflects the intuition discussed earlier that under any incentive scheme, $\hat{\gamma}^H = \hat{\gamma}^L \equiv \hat{\gamma}$ and therefore $p^t(\hat{\gamma}) + \zeta_H^t(\hat{\gamma}) -$

$c_H = p^t(\hat{\gamma}) + \zeta_L^t(\hat{\gamma}) - c_L$ for all t. This property guarantees that if we can induce both types to be selective in *any* period, then we can do so in *every* period by adjusting the incremental wages appropriately. The second-best optimality of deadline schemes follows from Lemma 4. Given that for any given period either H's wait for sure or L's do it for sure, initially the principal will induce L's to be selective and tolerate that H's inefficiently wait. But if L's are selective while H's wait, then as time passes without completion of the task it becomes more likely that the agent is type H. When the likelihood that the agent is an H becomes large enough, it is optimal to increase the punishment for delay so as to induce H's to start being selective. That point is the deadline. To summarize,

PROPOSITION 4. Suppose that $c_H - c_L > 2k/(1 - \beta)$, so that no incentive scheme can induce efficiency for both types. Then,

(i) There exists $D^* \geq 1$ such that the $(\overline{w}_L,\overline{w}_H)$-deadline scheme with deadline D^* is second-best optimal. This incentive scheme will induce $\gamma_t^L = s$ and $\gamma_t^H = w$ for all $t < D^*$ and $\gamma_t^L = d$ and $\gamma_t^H = s$ for all $t \geq D^*$. D^* is either an integer satisfying $\alpha - 1 < D^* < \alpha + 1$ (when $\alpha > 1$) or $D^* = 1$ (when $\alpha \leq 1$), where

$$\alpha \equiv 1 + \frac{\ln(\pi/(1-\pi)) + \ln((k - x^{\Delta})/x^{\Delta}) + \ln(\ln 2)}{\ln 2}.$$

(ii) If $c_L \geq (1 + \beta)/(1 - \beta)k$, then all second-best optimal incentive schemes are deadline schemes.

Part (i) of Proposition 4 establishes that a $(\overline{w}_L,\overline{w}_H)$-deadline scheme is always second-best optimal. While there can be other incentive schemes that are second-best optimal, they all induce exactly the same behavior except for knife-edge parameter values. While part (ii) of Proposition 4 establishes that sometimes all second-best optimal incentive schemes are deadline schemes, when $c_L < (1 + \beta)k/(1 - \beta)$, second-best behavior can be induced with some nondeadline schemes. In this case, the principal can mimic the outcome of the simple deadline scheme specified in Proposition 4 part (i) with alternative (more complicated) schemes.[21]

21. A result analogous to Lemma 3 could be formalized, however: the "average" incremental wage before date D^* must be less than the "average" incremental wage after date D^*.

The optimal incentive scheme in Proposition 4 implies the following observed pattern of behavior: a number of people complete the task immediately (half the L's), and another large group completes the task in period D, just before the more severe punishment kicks in (the remaining L's and half the H's). In between, we would observe smaller (and decreasing) numbers of agents doing it. In other words, people less prone to procrastinate complete the task at the first convenient time, or just before the deadline if no convenient time arises. People more prone to procrastinate wait until the deadline (and often beyond) before completing the task.

Proposition 4 also implies some comparative statics that reflect the intuition of why these schemes are attractive. The optimal deadline D^* is independent of c_L, c_H, and β except insofar as they determine whether efficiency is possible. That is, if $c_H - c_L < 2k/(1 - \beta)$, then a stationary incentive scheme is optimal. Otherwise, a deadline scheme is optimal, and which deadline is optimal is independent of c_L, c_H, and β. The deadline D^* instead depends only on the relative likelihood of L's versus H's (i.e., π), and on the relative benefits of discouraging delay versus encouraging selective performance (i.e., on k versus x^{Δ}). As π approaches 1, D^* approaches ∞, and as π approaches 0, D^* approaches 1. If the population is predominantly L's, it is optimal to give them more opportunities to get a low-cost realization; and if the population is predominantly H's, it is optimal to have no delay before inducing them to be selective. The term $((k - x^{\Delta})/x^{\Delta})$ reflects the relative benefits of discouraging delay versus encouraging selective performance. The larger is k relative to the delay cost x^{Δ} (while maintaining the condition $c_H - c_L \geq 2\,k/(1 - \beta)$), the larger is D^*. In other words, as task-cost considerations become more important than delay-cost considerations, the principal can give L's a longer time to find a small cost and still get H's to eventually find a small cost.[22]

This section has shown in a simple model how deadline schemes are second-best optimal for present-biased agents in a way they would not be for time-consistent agents. Generalizing our model to allow full-support task-cost distributions rather

22. We feel an important subsidiary contribution of Proposition 4 is to introduce into economic theory the previously neglected number ln (ln (2)).

than discrete task-cost distributions or more than two types of agent would make it unlikely that the optimal incentive schemes for naifs would be simple two-part linear schemes. Optimal incentive schemes will, however, be concave, with the punishment for delay becoming increasingly harsh over time. Because these generalizations yield no qualitatively different results or insights beyond indicating the nongenerality of our two-part linear scheme, we have omitted them from the paper. In the next section, however, we turn to a qualitatively different generalization.

V. Different Task-Cost Distributions

In this section we generalize the model of Section IV to allow different types of the agent to have task-cost distributions with different variances. Consider a model identical to that in Section IV except that now the task-cost distribution is

$$c = \begin{cases} c_i - k_i & \text{with probability } 1/2 \\ c_i + k_i & \text{with probability } 1/2, \end{cases}$$

for type $i \in \{L,H\}$, where $c_L < c_H$. We assume $k_L < c_L$ and $k_H < c_H$, guaranteeing that the task cost is always positive. While we feel it is natural to assume that $k_H \geq k_L$, meaning a higher average task cost is associated with increased day-to-day variance, we also consider the case where $k_L > k_H$.[23] For the majority of this section we also assume $x^{\Delta} < \min\{k_L,k_H\}$, so that, as in Section IV, it is efficient for both types to be selective in all periods. We briefly comment at the end of this section on the cases $k_H > x^{\Delta} \geq k_L$ and $k_L > x^{\Delta} \geq k_H$.

This section develops two main findings regarding the robustness of the deadline results in Section IV. First, when $k_H \geq k_L$, Propositions 6 and 7 establish that whenever a stationary incentive scheme cannot induce efficiency for both types, deadline schemes are second-best optimal. Second, when $k_L > k_H$, our deadline result need not hold: in Example 4 below, a nondeadline scheme is superior to all deadline schemes.

23. The condition $k_H > k_L$ would hold, for instance, if we interpret c_H and c_L as the number of hours required to complete the task and assume that the stochastic opportunity costs for each hour of the day are identically and independently distributed with a mean of one.

As a preliminary step, we first establish the conditions under which the principal can induce efficiency for both types.[24] Recall that Lemmas 2 and 3 characterize the types of incentive schemes that can induce type i to be selective in all periods. These lemmas continue to hold replacing k with k_i in the formulas for $\underline{w}_i$ and $\overline{w}_i$. In other words, for type i there exist $\underline{w}_i$ and $\overline{w}_i$ such that i's are selective in all periods for any stationary incremental wage $w^\Delta \in [\underline{w}_i,\overline{w}_i]$. Moreover, *any* incentive scheme that induces i's to be selective in all periods must "on average" reflect the incentives of some stationary incentive scheme that induces i's to be selective in all periods. Hence, Proposition 5 below is analogous to Proposition 3: either a stationary incentive scheme can induce efficiency for both types (when $\underline{w}_H \leq \overline{w}_L$), or no incentive scheme can induce efficiency for both types (when $\underline{w}_H > \overline{w}_L$).[25]

PROPOSITION 5. First-best efficiency can be induced if and only if it can be induced with a stationary incentive scheme.

For the case when $\underline{w}_H > \overline{w}_L$, no incentive scheme can induce full efficiency. Will deadline schemes be second-best optimal? When $k_H \geq k_L$, Lemma 5 establishes that as long as $c_H \geq (1 + \beta)k_H/(1 - \beta)$ a result identical to Lemma 4 holds: *for any individual period t,* if L's are selective (or wait) in period t then H's must wait in period t.

LEMMA 5. Suppose that $k_H \geq k_L$ but $\underline{w}_H > \overline{w}_L$, so that no incentive scheme can induce efficiency for both types. If $c_H \geq (1 + \beta)k_H/(1 - \beta)$, then for any t, $\gamma_t^H \neq w$ implies that $\gamma_t^L = d$.

From Lemma 5 we can establish Proposition 6, which is equivalent to Proposition 4 except that k_L replaces k in the equation for α.[26]

PROPOSITION 6. Suppose that $k_H \geq k_L$ but $\underline{w}_H > \overline{w}_L$, so that no incentive scheme can induce efficiency for both types. If $c_H \geq$

24. Since much of the analysis here is exactly analogous to that in Section IV, we do not provide as much detail.

25. There is, however, no simple condition on underlying parameters for when the principal can induce efficiency.

26. Since H's are waiting until after the deadline, k_H is irrelevant to determining the deadline date.

$(1 + \beta)k_H/(1 - \beta)$, then

(i) There exists $D^* \geq 1$ such that the $(\overline{w}_L, \overline{w}_H)$-deadline scheme with deadline D^* is second-best optimal. This incentive scheme will induce $\gamma_t^L = s$ and $\gamma_t^H = w$ for all $t < D^*$ and $\gamma_t^L = d$ and $\gamma_t^H = s$ for all $t \geq D^*$. D^* is either an integer satisfying $\alpha - 1 < D^* < \alpha + 1$ (when $\alpha > 1$) or $D^* = 1$ (when $\alpha \leq 1$) where

$$\alpha \equiv 1 + \frac{\ln(\pi/(1 - \pi)) + \ln((k_L - x^{\Delta})/x^{\Delta}) + \ln(\ln 2)}{\ln 2}.$$

(ii) If $c_L \geq (1 + \beta)k_L/(1 - \beta)$, then all second-best optimal incentive schemes are deadline schemes.

When $k_H \geq k_L$ but $c_H < (1 + \beta)k_H/(1 - \beta)$, there can be periods in which L's and H's are both selective, so Lemma 5 does not hold. However, there is a slightly weaker result, Lemma 6: if L's are selective (or wait) *in all periods,* then H's must wait *in all periods.*

LEMMA 6. Suppose that $k_H \geq k_L$ but $\underline{w}_H > \overline{w}_L$, so that no incentive scheme can induce efficiency for both types. If $c_H < (1 + \beta)k_H/(1 - \beta)$, then if $\gamma_t^L \neq d$ for all t, $\gamma_t^H = w$ for all t.

Although Lemma 6 is weaker than Lemmas 4 and 5, it is sufficient to establish the second-best optimality of deadlines: Proposition 7 shows that there must be some "deadline" D at which L's complete the task for sure. However, behavior under second-best optimal deadline schemes can be slightly different (and "better") in the conditions under which Proposition 7 applies than in Propositions 4 and 6. While L's behave the same as before—they are selective before the deadline and do it at the deadline—H's might behave differently. If k_H is large enough, then an impending deadline can induce H's to start being selective before the deadline. In Proposition 7, $\hat{d}$ represents the number of periods just prior to the deadline in which H's are selective.

PROPOSITION 7. Suppose that $k_H \geq k_L$ but $\underline{w}_H > \overline{w}_L$, so that no incentive scheme can induce efficiency for both types. If $c_H < (1 + \beta)k_H/(1 - \beta)$, then

(i) There exists $D^{**} \geq 1$ such that the $(\overline{w}_L, \overline{w}_H)$-deadline scheme with deadline D^{**} is second-best optimal. This incentive scheme will induce $\gamma_t^L = s$ for $t < D^{**}$, $\gamma_t^L = d$ for $t \geq D^{**}$, $\gamma_t^H = w$ for $t < D^{**} - \hat{d}$, and $\gamma_t^H = s$ for $t \geq D^{**} -$

$\hat{d}$, where

$$\hat{d} \equiv \min \left\{ n \in \{0,1,2,\ldots\} \,\middle|\, \frac{1-\beta}{\beta} c_H - \frac{1}{\beta} k_H > k_H - \sum_{j=0}^{n} \left(\frac{1}{2}\right)^j (k_H - \overline{w}_L) \right\},$$

and $D^{**} = \max\{D^*, \hat{d} + 1\}$, where D^* is defined in Proposition 6.[27]

(ii) If $c_L \geq (1 + \beta)k_L/(1 - \beta)$, then all second-best optimal incentive schemes are deadline schemes.

Hence, when $k_H \geq k_L$, our main result holds: if $\underline{w}_H > \overline{w}_L$, so that no incentive scheme can induce efficient behavior for both types, there is always a second-best optimal $(\overline{w}_L, \overline{w}_H)$-deadline scheme. In the less likely case $k_L > k_H$, however, deadline schemes may not be optimal. The following example illustrates this possibility.

Example 4

Suppose that $\beta = \frac{1}{2}$, $c_L = 8$, $k_L = 6$, $c_H = 22$, $k_H = \frac{1}{2}$, $x^\Delta \in (\frac{1}{4}, \frac{1}{2})$, and $\pi = \frac{1}{2}$.

Since $k_L > x^\Delta$ and $k_H > x^\Delta$, so efficiency calls for both types to be selective in all periods. It can be shown that $\underline{w}_L = 1$, $\overline{w}_L = 20$, $\underline{w}_H = 21$, and $\overline{w}_H = 23$. Hence, $\underline{w}_L < \overline{w}_L < \underline{w}_H < \overline{w}_H$, so Proposition 5 implies that the principal cannot induce efficient behavior from both types. Consider the following nondeadline incentive scheme: $w_1^\Delta = 24$, and $w_t^\Delta = 1$ for $t \in \{2,3,4,\ldots\}$. Under this scheme, $\gamma^L = (s,s,\ldots)$ and $\gamma^H = (d,w,w,w,\ldots)$, so L's behave efficiently, and H's (inefficiently) complete the task immediately. One can show that this incentive scheme is better than any deadline scheme when $x^\Delta \in (\frac{1}{4}, \frac{1}{2})$.

Example 4 relies on the fact that when $k_H < k_L$ there are conditions where H's are *more* likely to complete the task than L's.

27. The optimal deadline $D^{**} = \max\{D^*, \hat{d} + 1\}$ because $\hat{d}$ does not affect the benefits to the principal of changing the deadline unless the deadline is less than $\hat{d} + 1$. Hence, the optimal deadline D^{**} will be equal to D^* when $D^* > \hat{d}$. For $D^* \leq \hat{d}$, the optimal deadline is $D^{**} = \hat{d} + 1$: a deadline of $\hat{d} + 1$ implies that H's are always selective and L's are selective in the first $\hat{d}$ periods, so any shorter deadline is clearly inferior because H's are unaffected but L's are selective for fewer periods.

Indeed, when k_H is sufficiently smaller than k_L, a reverse-deadline scheme can induce H's to complete the task immediately while still inducing L's to be selective in all periods. Intuitively, a small k_H implies that H's have very little to gain from waiting in period 1 when their task-cost realization is high, while at the same time a large k_L implies that L's have a lot to gain from waiting in period 1 when their task-cost realization is high. Moreover, since any stationary predeadline wage that would induce L's to be selective would induce H's to wait, inducing immediate completion by H's might be more attractive than having H's wait until a deadline. Indeed, the parameters in Example 4 are such that by far the most important efficiency concern is enabling L's to find a low task-cost realization, and therefore the reverse-deadline scheme outperforms deadline schemes.

Finally, consider what happens when it is more efficient for only one type to be selective. If $k_H > x^\Delta \geq k_L$, then efficiency requires that L's complete the task and H's be selective in all periods. In this case, first-best efficiency is always achievable: because $k_H > k_L$ implies that $\overline{w}_H > \overline{w}_L$, true that $\underline{w}_L < \overline{w}_H$, efficiency will be induced with any stationary incremental wage $w^\Delta \in [\max\{\overline{w}_L, \underline{w}_H\}, \overline{w}_H]$. The intuition is straightforward: in all earlier examples where the first best was not attainable, any attempt to get H's to be selective always made L's inefficiently complete the task for sure. If instead it is efficient for L's to complete the task for sure, then efficiency is clearly attainable.

If $k_L > x^\Delta \geq k_H$, then it is efficient for L's to be selective and for H's to complete the task in all periods. This case is less straightforward. Since $k_L > k_H$, it is possible to have $\overline{w}_L \geq \overline{w}_H$, in which case efficiency can be achieved with any stationary incremental wage $w^\Delta \in [\overline{w}_H, \overline{w}_L]$. If $\overline{w}_L < \overline{w}_H$, a stationary incentive scheme cannot induce efficiency. However, it may be possible to induce efficiency with a nonstationary incentive scheme: the "reverse-deadline" scheme in Example 4 above induces efficiency when $x^\Delta \in (\frac{1}{2}, 6)$.[28] For some parameter values, however, no incentive scheme can induce efficiency. We have found examples where stationary incentive schemes that induce both types to always be selective are second-best optimal; and we have found examples where

28. That nonstationary schemes can induce efficiency in this situation where stationary schemes cannot is not surprising, since it is efficient for the two types of agent to behave differently.

deadline schemes under which H's complete the task for sure at the deadline are second-best optimal.

VI. Sophisticated Agents

Our analysis in this paper assumes that agents are naive about their time inconsistency. How might our analysis change if we assumed that agents were sophisticated? While a complete analysis of sophisticated agents gives rise to a number of complications, and is beyond the scope of this paper, we argue that the intuition for the second-best optimality of deadline incentive schemes is not merely an artifact of the naivete assumption.

In the complete-information model of Section III, sophisticates, unlike naifs, are not susceptible to bilking; but in other respects, the behavior of sophisticates is qualitatively the same as the behavior of naifs. When the principal knows the distribution of task costs, there always exists an incentive scheme under which sophisticates behave efficiently. Moreover, in order to counteract procrastination, this incentive scheme must punish delay by more than the true cost of delay. Finally, the larger the expected task cost, the larger the propensity to procrastinate, and therefore the steeper must be the incentive scheme to counteract procrastination. Although sophistication always mitigates the tendency to procrastinate (often by a lot), it does not eliminate the problem.[29] Hence, the principal faces the same (qualitative) procrastination problem.

Consider a model identical to the asymmetric-information model in Section IV, except now suppose that the agent is sophisticated. Efficient behavior is independent of whether the agent is a TC, naif, or sophisticate. Suppose again that $k > x^{\Delta}$, so it is efficient for the agent to be selective in all periods.

Let $\overline{\gamma}^i \equiv (\overline{\gamma}_1^i, \overline{\gamma}_2^i, \ldots, \overline{\gamma}_T^i)$ be a perception-perfect strategy for sophisticates. Like naifs, in each period sophisticates compare the known cost of completing the task now with their *perceived* expected total cost from waiting. Also like naifs, sophisticates discount future costs by factor β. But unlike naifs, sophisticates correctly predict their future behavior, so their perceived total cost from waiting is $p^t(\overline{\gamma}^i) + \zeta_i^t(\overline{\gamma}^i)$. Hence, $\overline{\gamma}^i \equiv (\overline{\gamma}_1^i, \overline{\gamma}_2^i, \ldots, \overline{\gamma}_T^i)$ satisfies

29. This claim is formalized and proved in O'Donoghue and Rabin [1999].

for each $t < T$[30]

$$\overline{\gamma}_t^i = \begin{cases} d & \text{if } c_i + k \leq \beta[p^t(\overline{\gamma}^i) + \zeta_i^t(\overline{\gamma}^i)] \\ s & \text{if } c_i - k \leq \beta[p^t(\overline{\gamma}^i) + \zeta_i^t(\overline{\gamma}^i)] \leq c_i + k \\ w & \text{if } c_i - k \geq \beta\,[p^t(\overline{\gamma}^i) + \zeta_i^t(\overline{\gamma}^i)]. \end{cases}$$

When can the principal induce efficient behavior for both types? The answer corresponds exactly to the result for naifs captured in Proposition 3.

PROPOSITION 8. Suppose that the agent is sophisticated. If $c_H - c_L \leq 2k/(1-\beta)$, there exists a stationary incentive scheme under which both types behave efficiently (by being selective every period); if $c_H - c_L > 2k/(1-\beta)$, then no incentive scheme can induce both types to behave efficiently.

The intuition for Proposition 8 is identical to that for Proposition 3. For each type of sophisticate, a range of stationary incremental wages can induce efficient behavior. Furthermore, any nonstationary incentive scheme that induces efficient behavior must "on average" reflect the incentives of some stationary incentive scheme that induces efficient behavior. Hence, as for naifs, either a stationary incentive scheme can induce efficiency for both types (when the difference in average costs for the two types is small), or no incentive scheme can induce efficiency for both types (when the difference is large).

Unfortunately, it is difficult to explicitly solve for second-best optimal incentive schemes for sophisticated agents because their behavior can be quite complicated, even under stationary incentive schemes. It is therefore difficult to address whether deadline incentive schemes will be second-best optimal for sophisticates when first-best efficiency is unachievable. But we suspect that the basic intuition behind the second-best optimality of deadline incentive schemes for naifs also holds for sophisticates. Consider the following example of how sophisticates behave under stationary incentive schemes.

30. As for TCs and naifs, this definition does not specify what the agent should do if he is indifferent between two actions, and we again assume that when indifferent the agent behaves as the principal would like him to behave.

Example 5

Suppose that $c_i = 30$ and $k = 5$, and consider a sophisticate with $\beta = 1/2$ facing a stationary incentive scheme with incremental wage w^{Δ}. Efficient behavior $(s,s, \ldots)$ is a perception-perfect strategy for sophisticates if and only if $w^{\Delta} \in [12.5,22.5]$. For $w^{\Delta} < 12.5$, the punishment for delay is not sufficiently severe to induce efficient behavior. In contrast to naifs, however, sophisticates will not procrastinate forever in this case because they correctly predict that such behavior would yield an infinite wage cost. Rather, they pursue a cyclical strategy; i.e., there exists $n \in \{2,3,4, \ldots\}$ such that $\overline{\gamma}_{t+n} = \overline{\gamma}_t$ for all t. For instance, the strategy $(s,s,w,s,s,w, \ldots)$ is a perception-perfect strategy for $w^{\Delta} = 10$, and the strategy $(s,s,w,s,w, s,s,w,s,w, \ldots)$ is a perception-perfect strategy for $w^{\Delta} = 9$.[31]

Example 5 helps motivate the potential second-best optimality of deadline incentive schemes for sophisticates. The crucial intuition behind the second-best optimality of deadline schemes for naifs is that *for any incentive scheme* H's are always less likely to complete the task than L's—indeed, for naifs, H's wait whenever L's are selective. Under any second-best optimal incentive scheme, therefore, the probability that a delaying agent is type H must eventually become so large that after some date it is optimal for the incentive scheme to induce H's to be selective from that date forward; this date is the "deadline." Example 5 suggests that this intuition may hold for sophisticates as well. We can in fact extrapolate from Example 5 to conclude that a stationary incentive scheme designed to induce sophisticated L's to be selective always will in all periods induce sophisticated H's either to be selective or to wait. Hence, at least for *stationary* incentive

31. For the case where the incremental wage w^{Δ} is too small to induce efficient behavior, we suspect (but have not proved) that for each w^{Δ} there is a unique perception-perfect strategy satisfying $\overline{\gamma}_t \neq d$ for all t, and that the smaller is the incremental wage w^{Δ} the less often the agent is selective. When c_i is large relative to k, there can also exist additional perception-perfect strategies where $\overline{\gamma}_t = d$ for some t (e.g., the strategy $(w,w,d,w,w,d, \ldots)$ might be perception-perfect). The existence of nonstationary perception-perfect strategies implies that our focus in this paper on history-independent strategies, while innocuous for naifs, may be important for sophisticates. Indeed, it can be the case that the efficient strategy $(s,s, \ldots)$ is not perception-perfect under our simple definition, whereas under a more general definition efficient behavior could be supported by the "threat" to revert to continuation strategy $(w,w,d,w,w,d, \ldots)$ if the person ever delays in the face of a low task-cost realization. Such threat-based strategies seem quite removed from reality, and would not be possible even for sophisticates in a generic, finite-horizon model.

schemes, sophisticated H's would be less likely to complete the task than sophisticated L's.

Because the analysis becomes quite complicated, however, we have not addressed the behavior of sophisticates under nonstationary (and nondeadline) incentive schemes. Since our preliminary calculations suggest that sophisticates can behave in very strange ways, we would not be surprised if there are examples where nonlinear, nondeadline incentive schemes are superior to the best deadline incentive schemes. If it turns out that such problems strongly undermine our earlier intuitions, then our results in this paper would apply only to the extent that people are naive about their self-control problems. Even so, since we believe that of the two extreme assumptions about people's awareness of their self-control problems, naivete is the more realistic, and since we suspect that *small* amounts of sophistication would not undermine our qualitative results, we believe the results of the naif model are relevant.

VII. Discussion and Conclusion

Various aspects of incentives for procrastinators are not incorporated into our model. An important issue that we have ignored is the delivery date of rewards. We assume that the reward the principal offers the agent for completing the task is *not salient* to the agent, in the sense that the agent does not experience the pleasure of the reward right away. The nonsalience of rewards seems realistic in most contexts—especially if the reward is money. But the principal might offer the agent, in addition to money, some sort of immediate nonmonetary reward, such as breaks or parties, once a task is completed. The same preference for immediate gratification that tempts the agent to put off incurring the task cost will tempt him to grab this "salient" reward. Because a naive agent will mispredict his reaction to such immediate rewards, it might be possible for the principal to use salient rewards to extract more surplus from agents. However, it is likely that agents do not value such rewards from a long-run perspective as much as nonsalient rewards such as income. We are therefore skeptical of the inefficient use of salient rewards for the same reason that we make the ex post break-even assumption throughout the paper: in the long run, agents will find a job unsatisfying if they are getting mostly short-run rewards that they do not value highly in the long run.

But salient rewards can potentially be efficiency-enhancing as well, because they can be used to align incentives for heterogeneous agents. Suppose—unlike the model in Section IV—that agents differ in their *innate* preference for immediate gratification, as measured by the parameter β. As in our model, punishments for delay harsh enough to induce efficient behavior for those with large self-control problems (i.e., small β) may be so harsh that those with small self-control problems (i.e., large β) complete the task when waiting is efficient. If there is some salient reward, however, then those agents with the largest self-control problems will react most to this incentive. Hence, the use of salient rewards may provide a second-best mechanism for aligning incentives for heterogeneous agents, and could in some cases be more efficient than the deadline schemes analyzed in our model.

Another important issue is that although we have treated reward schemes throughout this paper as monetary incentives, agents in organizations are rarely given explicit monetary incentives for early completion of specific tasks. More often, an employee's basic incentive scheme is that he is either fired or not fired, promoted or not promoted, depending loosely on his performance. Even so, whatever the "unit of account" by which a firm keeps track of an employee's performance, our model predicts that the firm will wish to somehow generate more and more severe marginal incentives for the agent to complete a task as completion is delayed further and further.[32]

Clearly, there are many reasons for deadlines other than combating procrastination. A major one, intuitively, is coordination among agents: in an organization it is often useful to know a date by which a project is intended to be complete. A second potential reason for deadlines is their simplicity: it may be easier to monitor whether somebody has met or missed a deadline than to monitor exactly when a project was completed.

But we think that these alternative reasons for deadlines complement rather than contradict the message of this paper. Many real-world deadline contracts seem to take a different (and simpler) form than the deadline schemes predicted by our model: whereas our model predicts an increasing *marginal* punishment

32. Of course, monetary incentives are not often used in many of the contexts considered by formal principal-agent models. Insofar as risk aversion plays no role in our results, perhaps our model suffers less from inapplicability to nonmonetary incentives than standard principal-agent models.

for delay, real-world deadline schemes often involve discrete punishments for missing deadlines with no marginal punishments between deadlines. Imposing such "lumpy" deadlines in environments where the actual marginal cost of delay is relatively constant over time might in fact be a necessary evil because of organizational and transactions-costs explanations. Moreover, even if the ideal incentive contract according to our model is smooth and concave, concavity implies that, among simple schemes, simple deadlines may be *better* than simple linear schemes.

We conclude by noting a subtext to this paper which we suspect might generate some of the paper's interest to many readers: the model may not only shed light on how a "principal" copes with an "agent's" procrastination, but also on how an individual copes with her own procrastination. In other words, we can interpret the "principal" as our current self and the "agent" as our future self.[33] Many people who procrastinate only moderately do so not because of intrinsic self-control, but because they have developed schemes to overcome procrastination. Some such schemes may use external commitment devices: people commit to giving a seminar in the hopes that this will force them to finish a paper. Other such schemes are internal: people try to fool themselves into believing in false deadlines; they exaggerate to themselves ahead of time how crucial it is that they meet their deadlines; and they impose on themselves internalized sanctions for missed deadlines.[34] Conceptualizing such self-incentives may be subtle, but we hope the analysis of this paper might be useful in this regard.

APPENDIX: PROOFS

Proof of Lemma 1. If $\underline{c}/\beta > \bar{c}$, then the result follows from Example 1.

33. Implicit in this interpretation is an issue we have discussed previously: a person may be "meta-sophisticated" and aware of her general propensity to procrastinate, but naive about day-to-day procrastination. We can think of the "meta-sophisticated" person as setting self-incentives to overcome future day-to-day naivete.

34. Such "internal deadlines" are a common theme in popular advice on how to remedy procrastination. Previous research on time-inconsistent preferences also discusses rules for self-control (outside the context of procrastination). Ainslie [1992] explores this issue; Thaler [1985] conjectures that internal attempts at self-control may help explain mental accounting rules of thumb that constrain our flexibility in how we spend our money; and Laibson [1994] develops a formal model along these lines, where a (sophisticated) time-inconsistent person develops rigid rules to counteract moral hazard by her everyday self.

Suppose that $\underline{c}/\beta \leq \bar{c}$, and consider W with $w_T^\Delta = \infty$ (so the agent must complete the task in or before period T), $w_{T-1}^\Delta = \underline{c}/\beta - E_C$, and $w_t^\Delta = F(\underline{c}/\beta)[\underline{c}/\beta - E(c|c < \underline{c}/\beta)]$ for all $t \in \{1,2, \ldots, T-2\}$. It is straightforward to show that $\hat{\gamma}_t = p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma}) = \underline{c}/\beta$ for all $t < T$, so $\hat{\gamma} = (\underline{c}/\beta, \underline{c}/\beta, \ldots, \underline{c}/\beta, \bar{c})$. For naifs, we have $\gamma_t = \beta\hat{\gamma}_t$ for all $t < T$, and therefore $\gamma = (\underline{c}, \underline{c}, \ldots, \underline{c}, \bar{c})$. Hence, under W naifs complete the task in period T with probability 1, while TCs complete the task with positive probability in all periods.

When naifs sign the contract, they believe they will behave like TCs. Hence, naifs perceive $U^p \geq 0$ if and only if $W_1 \geq (1 - F(\hat{\gamma}_1))p^1(\hat{\gamma}) + \zeta^0(\hat{\gamma})$. Consider the contract where this holds with equality. Since $\zeta^0(\hat{\gamma}) = F(\hat{\gamma}_1)E(c|\ c < \hat{\gamma}_1) + (1 - F(\hat{\gamma}_1))\zeta^1(\hat{\gamma})$ and $\hat{\gamma}_1 = \underline{c}/\beta$, we have $W_1 = (1 - F(\hat{\gamma}_1))(p^1(\hat{\gamma}) + \zeta^1(\hat{\gamma})) + F(\hat{\gamma}_1)E(c|c < \hat{\gamma}_1) = (1 - F(\underline{c}/\beta)\underline{c}/\beta + F(\underline{c}/\beta)E\,(c|c < \underline{c}/\beta)$, and therefore $W_T = -(T-1)F(\underline{c}/\beta)[\underline{c}/\beta - E(c\ |c < \underline{c}/\beta)] + E_C$. Naifs do the task in period T for wage W_T, and the result follows from $\lim_{T\to\infty} W_T = -\infty$. □

Proof of Proposition 1. Efficient behavior satisfies for each t, $\gamma_t^* = \chi^t(\gamma^*) + \zeta^t(\gamma^*)$. TC behavior satisfies for each t, $\hat{\gamma}_t = p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma})$. If $w_t^\Delta = x_t^\Delta$ for each t, then $p^t(\mathrm{s}) = \chi^t(\mathrm{s})$ for each t and s. The result follows. □

Proof of Proposition 2. (i) Consider $T < \infty$. The proof is straightforward: use backwards-induction logic, noting that γ_τ is independent of w_t^Δ for all $t < \tau$. Set w_T^Δ very large so $\gamma_T = \bar{c}$. Then set w_{T-1}^Δ appropriately so $\gamma_{T-1} = \gamma_{T-1}^*$, then set w_{T-2}^Δ appropriately so $\gamma_{T-2} = \gamma_{T-2}^*$, and so on.

Consider $T = \infty$. Given the definition of γ, $\gamma = \gamma^*$ if and only if $p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma}) = \gamma_t^*/\beta$ for all t. Given the definition of $\hat{\gamma}$, $p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma}) = \gamma_t^*/\beta$ for all t implies that $\hat{\gamma}_t = \min\{\gamma_t^*/\beta, \bar{c}\}$ for all t.

Suppose that $\gamma_{t+1}^*/\beta \geq \bar{c}$ so $\hat{\gamma}_{t+1} = \bar{c}$ and therefore $p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma}) = w_t^\Delta + E_C$. Then $p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma}) = \gamma_t^*/\beta$ if and only if $w_t^\Delta = (1/\beta)[\gamma_t^* - \beta E_C]$.

Suppose that $\gamma_{t+1}^*/\beta < \bar{c}$ so $\hat{\gamma}_{t+1} = \gamma_{t+1}^*/\beta$ and therefore

$$p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma}) = w_t^\Delta + \int_c^{\hat{\gamma}_{t+1}} \underline{c}\, dF(c)$$

$$+ (1 - F(\hat{\gamma}_{t+1}))[p^{t+1}(\hat{\gamma}) + \zeta^{t+1}(\hat{\gamma})]$$

$$= w_t^\Delta + \int_c^{\gamma_{t+1}^*/\beta} \underline{c} dF(c) = \left(1 - F\left(\frac{1}{\beta}\gamma_{t+1}^*\right)\right)\frac{1}{\beta}\gamma_{t+1}^*.$$

Then $p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma}) = \gamma_t^*/\beta$ if and only if $w_t^\Delta = (1/\beta)[\gamma_t^* - \beta \int_{\underline{c}}^{\gamma_{t+1}^*/\beta} cdF(c) - (1 - F(\gamma_{t+1}^*/\beta))\gamma_{t+1}^*]$. Hence, $\gamma = \gamma^*$ if and only if W satisfies for all t,

$$(\text{A.1})\quad w_t^\Delta = \begin{cases} (1/\beta)[\gamma_t^* - \beta E_C] & \text{if } \gamma_{t+1}^*/\beta \geq \bar{c} \\ (1/\beta)\Big[\gamma_t^* - \beta \int_{\underline{c}}^{\gamma_{t+1}^*/\beta} cdF\,(c) & \text{if } \gamma_{t+1}^*/\beta \leq \bar{c}. \\ \quad - (1 - F(\gamma_{t+1}^*/\beta))\gamma_{t+1}^*\Big] & \end{cases}$$

(ii) If $\gamma_t^* < \bar{c}$ for all t, then $\gamma_t^* = \chi^t(\gamma^*) + \zeta^t(\gamma^*)$ for all t. Given this, we show that w_t^Δ from equation (A.1) is strictly greater than x_t^Δ for all $t < T$, where $\gamma_T^* = \bar{c}$ for the $T < \infty$ case. For all $t < T$, we have

$$\gamma_t^* = \chi^t(\gamma^*) + \zeta^t(\gamma^*) = x_t^\Delta + \int_{\underline{c}}^{\gamma_{t+1}^*} cdF(c) + (1 - F(\gamma_{t+1}^*))(\chi^{t+1}(\gamma^*) + \zeta^{t+1}(\gamma^*))$$

or

$$x_t^\Delta = \gamma_t^* - \int_{\underline{c}}^{\gamma_{t+1}^*} cdF(c) - (1 - F(\gamma_{t+1}^*))\gamma_{t+1}^*.$$

Suppose $\gamma_{t+1}^*/\beta \leq \bar{c}$, so equation (A.1) implies that $w_t^\Delta > \gamma_t^* - \beta \int_{\underline{c}}^{\gamma_{t+1}^*/\beta} cdF(c) - (1 - F(\gamma_{t+1}^*/\beta))\, \gamma_{t+1}^*$. Then $w_t^\Delta > x_t^\Delta$ if $\int_{\underline{c}}^{\gamma_{t+1}^*} cdF(c) + (1 - F(\gamma_{t+1}^*))\, \gamma_{t+1}^* \geq \beta \int_{\underline{c}}^{\gamma_{t+1}^*/\beta} cdF(c) + (1 - F(\gamma_{t+1}^*/\beta))\, \gamma_{t+1}^*$, which we can rewrite as

$$[F(\gamma_{t+1}^*/\beta) - F(\gamma_{t+1}^*)]\gamma_{t+1}^* + (1 - \beta)\int_{\underline{c}}^{\gamma_{t+1}^*} c\, dF(c) - \beta\left[\int_{\underline{c}}^{\gamma_{t+1}^*/\beta} cdF(c) - \int_{\underline{c}}^{\gamma_{t+1}^*} cdF(c)\right] \geq 0.$$

Using $\beta[\int_{\underline{c}}^{\gamma_{t+1}^*/\beta} cdF(c) - \int_{\underline{c}}^{\gamma_{t+1}^*} cdF(c)] = \beta \int_{\gamma_{t+1}^*}^{\gamma_{t+1}^*/\beta} cdF(c) < [F(\gamma_{t+1}^*/\beta) - F(\gamma_{t+1}^*)]\gamma_{t+1}^*$, the inequality holds.

Suppose that $\gamma_{t+1}^*/\beta > \bar{c}$, so equation (A.1) implies that $w_t^\Delta > \gamma_t^* - \beta E_C$. Then $w_t^\Delta > x_t^\Delta$ if $\int_{\underline{c}}^{\gamma_{t+1}^*} cdF(c) + (1 - F(\gamma_{t+1}^*))\gamma_{t+1}^* > \beta E_C$. Since $\gamma_{t+1}^*/\beta > \bar{c}$ implies that $\beta E_C < (\gamma_{t+1}^*/\bar{c})\, E_C$, this inequality holds if $\int_{\underline{c}}^{\gamma_{t+1}^*} cdF(c) + (1 - F(\gamma_{t+1}^*))\, \gamma_{t+1}^* \geq (\gamma_{t+1}^*/\bar{c})\, E_C$, which we can rewrite as $(1 - F(\gamma_{t+1}^*))\, \gamma_{t+1}^* + ((\bar{c} - \gamma_{t+1}^*)/\bar{c}) \int_{\underline{c}}^{\gamma_{t+1}^*} c\, dF(c) - (\gamma_{t+1}^*/\bar{c})\, [\int_{\underline{c}}^{\bar{c}} c\, dF(c) - \int_{\underline{c}}^{\gamma_{t+1}^*} c\, dF(c)] \geq 0$. Using $(\gamma_{t+1}^*/\bar{c})\, [\int_{\underline{c}}^{\bar{c}} c\, dF(c) - \int_{\underline{c}}^{\gamma_{t+1}^*} c\, dF(c)] = (\gamma_{t+1}^*/\bar{c}) \int_{\gamma_{t+1}^*}^{\bar{c}} c\, dF(c) < (1 - F(\gamma_{t+1}^*))\, \gamma_{t+1}^*$, the inequality holds.

Finally, we must show that $w^{\Delta}_{t-1} = w^{\Delta}_{t}$ for all t when X has a stationary delay cost x^{Δ}. A stationary delay cost implies that $\gamma^{*}_{t} = \gamma^{*}_{t+1}$ for all t. Using the equation for w^{Δ}_{t} above, the result follows.□

Proof of Lemma 2. First, suppose that $w^{\Delta} < k$. $w^{\Delta} < k$ implies that $\hat{\gamma}^{i} = (s,s,\ldots)$, in which case $\zeta^{t}_{i}(\hat{\gamma}^{i}) = c_i - k$ and $p^{t}(\hat{\gamma}^{i}) = 2w^{\Delta}$ for all t. Then for any t, $\gamma^{i}_{t} = d$ if and only if $c_i + k < \beta(c_i - k + 2w^{\Delta})$; but $w^{\Delta} < k$ implies that $c_i + k > \beta(c_i - k + 2w^{\Delta})$, so $w^{\Delta} < k$ implies $\gamma^{i}_{t} \neq d$. For any t, $\gamma^{i}_{t} = w$ if and only if $c_i - k > \beta(c_i - k + 2w^{\Delta})$ or $w^{\Delta} < (1-\beta)c_i/(2\beta) - (1-\beta)k/(2\beta)$, and $(1-\beta)c_i/(2\beta) - (1-\beta)k/(2\beta) < k$ if and only if $c_i < (1+\beta)k/(1-\beta)$. Hence, if $c_i \geq (1+\beta)k/(1-\beta)$, then $\gamma^{i} = (w, w, \ldots)$ for any $w^{\Delta} < k$, and if $c_i < (1+\beta)k/(1-\beta)$, then $\gamma^{i} = (s, s, \ldots)$ for any $(1-\beta)c_i/(2\beta) - (1-\beta)k/(2\beta) \leq w^{\Delta} < k$ and $\gamma^{i} = (w,w,\ldots)$ for any $w^{\Delta} < (1-\beta)c_i/(2\beta) - (1-\beta)k/(2\beta)$.

Second, suppose that $w^{\Delta} \geq k$. $w^{\Delta} \geq k$ implies $\hat{\gamma}^{i} = (d,d,\ldots)$, in which case $\zeta^{t}_{i}(\hat{\gamma}^{i}) = c_i$ and $p^{t}(\hat{\gamma}^{i}) = w^{\Delta}$ for all t. Then for any t, $\gamma^{i}_{t} = d$ if and only if $c_i + k < \beta(c_i + w^{\Delta})$ or $w^{\Delta} > (1-\beta)c_i/\beta + k/\beta$. For any t, $\gamma^{i}_{t} = w$ if and only if $c_i - k > \beta(c_i + w^{\Delta})$ or $w^{\Delta} < (1-\beta)c_i/\beta - k/\beta$, and $(1-\beta)c_i/\beta - k/\beta > k$ if and only if $c_i > (1+\beta)k/(1-\beta)$. Hence, if $c_i > (1+\beta)k/(1-\beta)$, then $\gamma^{i} = (w, w, \ldots)$ for any $k \leq w^{\Delta} < (1-\beta)c_i/\beta - k/\beta$, $\gamma^{i} = (s,s,\ldots)$ for any $(1-\beta)c_i/\beta - k/\beta \leq w^{\Delta} \leq (1-\beta)c_i/\beta + k/\beta$, and $\gamma^{i} = (d,d,\ldots)$ for any $w^{\Delta} > (1-\beta)/\beta c_i + 1/\beta k$. If $c_i \leq (1+\beta)k/(1-\beta)$, then $\gamma^{i} = (s,s,\ldots)$ for any $k \leq w^{\Delta} \leq (1-\beta)c_i/\beta + k/\beta$, and $\gamma^{i} = (d,d,\ldots)$ for any $w^{\Delta} > (1-\beta)c_i/\beta + k/\beta$.

Combining the cases $w^{\Delta} < k$ and $w^{\Delta} \geq k$, the result follows. □

Proof of Lemma 3. To prove that $\gamma^{i}_{t} = s$ for all t implies that $\Sigma^{\infty}_{t=1}(\tfrac{1}{2})^{t}\, w^{\Delta}_{t} \geq \underline{w}_{i}$, we use a revealed preference argument to put restrictions on $p^{t}(\hat{\gamma}) + \zeta^{t}(\hat{\gamma})$. For all t, $\hat{\gamma}$ minimizes $p^{t}(\hat{\gamma}) + \zeta^{t}(\hat{\gamma})$. Since TCs have the option of being selective in all periods, we can conclude that for all t, $p^{t}(\hat{\gamma}) + \zeta^{t}(\hat{\gamma}) \leq \Sigma^{\infty}_{\tau=t}(\tfrac{1}{2})^{\tau-t}\, w^{\Delta}_{\tau} + c_i - k$; and since TCs have the option of doing it for sure in all periods, we can conclude that $p^{t}(\hat{\gamma}) + \zeta^{t}(\hat{\gamma}) \leq w^{\Delta}_{t} + c_i$. Now, $\gamma^{i}_{t} = s$ only if $c_i - k \leq \beta[p^{t}(\hat{\gamma}) + \zeta^{t}(\hat{\gamma})]$. Using $p^{t}(\hat{\gamma}) + \zeta^{t}(\hat{\gamma}) \leq w^{\Delta}_{t} + c_i$, we must have $w^{\Delta}_{t} \geq (1-\beta)c_i/\beta - k/\beta$ for all t, which implies that $\Sigma^{\infty}_{t=1}(\tfrac{1}{2})^{t}\, w^{\Delta}_{t} \geq (1-\beta)c_i/\beta - k/\beta$. Using $p^{t}(\hat{\gamma}) + \zeta^{t}(\hat{\gamma}) \leq \Sigma^{\infty}_{\tau=t}(\tfrac{1}{2})^{\tau-t}\, w^{\Delta}_{\tau} + c_i - k$, we must have $\Sigma^{\infty}_{\tau=t}(\tfrac{1}{2})^{\tau-t}\, w^{\Delta}_{\tau} \geq (1-\beta)(c_i - k)/\beta$ for all t, which implies that $\Sigma^{\infty}_{t=1}(\tfrac{1}{2})^{t}\, w^{\Delta}_{t} \geq (1-\beta)(c_i - k)/(2\beta)$. Hence, for either $\underline{w}_{\mathrm{i}} = (1-\beta)c_i/\beta - k/\beta$ or $\underline{w}_{i} = (1-\beta)(c_i - k)/(2\beta)$, $\Sigma^{\infty}_{t=1}(\tfrac{1}{2})^{t}\, w^{\Delta}_{t} \geq \underline{w}_{i}$.

To prove that $\gamma^{i}_{t} = s$ for all t implies that $\Sigma^{\infty}_{t=1}(\tfrac{1}{2})^{t}\, w^{\Delta}_{t} \leq \overline{w}_{i}$, we prove that for all t there exists $\mu(t) \in \{t, t+1, \ldots\}$ such that

$\Sigma_{\tau=t}^{\mu(t)}(1/2)^{\tau-t}\, w_\tau^\Delta < \Sigma_{\tau=t}^{\mu(t)}(1/2)^{\tau-t}\, \overline{w}_i$, from which the result follows. Define $\mu(t) = \min\{\tau \geq t \mid \hat{\gamma}_{\tau+1}^i = d\}$ if this exists, otherwise let $\mu(t) = \infty$. If $\mu(t) = t$, then $p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma}) = w_t^\Delta + c_i$, in which case $c_i + k \geq \beta[p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma})]$ only if $w_t^\Delta \leq (1-\beta)c_i/\beta + k/\beta = \overline{w}_i$. If $\mu(t) > t$, then $p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma}) = \Sigma_{\tau=t}^{\mu(t)}(1/2)^{\tau-t}\, w_\tau^\Delta + c_i - \Sigma_{\tau=t+1}^{\mu(t)}(1/2)^{\tau-t}\, k$, in which case $c_i + k \geq \beta[p^t(\hat{\gamma}) + \zeta^t(\hat{\gamma})]$ only if $\Sigma_{\tau=t}^{\mu(t)}(1/2)^{\tau-t}\, w_\tau^\Delta \leq \overline{w}_i + \Sigma_{\tau=t+1}^{\mu(t)}(1/2)^{\tau-t}\, k < \Sigma_{\tau=t}^{\mu(t)}(1/2)^{\tau-t}\, \overline{w}_i$, where the last inequality follows from $k < \overline{w}_i$.

Proof of Proposition 3. That a stationary incentive scheme can induce efficiency if and only if $\underline{w}_H \leq \overline{w}_L$ follows directly from Lemma 2 and the fact that $\underline{w}_L < \overline{w}_H$. Specifically, any stationary incentive scheme with $w^\Delta \in [\max\{\underline{w}_L, \underline{w}_H\}, \min\{\overline{w}_L, \overline{w}_H\}]$ will induce efficiency. That no incentive scheme can induce efficiency if $\underline{w}_H > \overline{w}_L$ follows directly from Lemma 3. It remains to prove that $\underline{w}_H \leq \overline{w}_L$ if and only if $c_H - c_L \leq 2k/(1-\beta)$. For all c_L and k, $\overline{w}_L = (1-\beta)c_L/\beta + k/\beta > k$. Moreover, $c_H - c_L \leq 2k/(1-\beta)$ if and only if $(1-\beta)c_H/\beta - k/\beta \leq \overline{w}_L$. There are now two cases to consider for $\underline{w}_H$. First, $c_H \geq (1+\beta)k/(1-\beta)$ implies $\underline{w}_H = (1-\beta)c_H/\beta - k/\beta$, in which case $\underline{w}_H \leq \overline{w}_L$ if and only if $c_H - c_L \leq 2k/(1-\beta)$. Second, $c_H < (1+\beta)k/(1-\beta)$ actually requires $c_H - c_L < 2k/(1-\beta)$, so the result follows if $c_H < (1+\beta)k/(1-\beta)$ implies that $\underline{w}_H \leq \overline{w}_L$. And if $c_H < (1+\beta)k/(1-\beta)$, then $\underline{w}_H = (1-\beta)(c_H - k)/(2\beta) < k < \overline{w}_L$. □

Proof of Lemma 4. As discussed in the text, under any incentive scheme we must have $\hat{\gamma}^L = \hat{\gamma}^H \equiv \hat{\gamma}$, and therefore for all t, $p^t(\hat{\gamma}) + \zeta_H^t(\hat{\gamma}) - c_H = p^t(\hat{\gamma}) + \zeta_L^t(\hat{\gamma}) - c_L$. H's do not wait in period t *only if* $c_H - k \leq \beta\,(p^t(\hat{\gamma}^H) + \zeta_H^t(\hat{\gamma}^H))$ or $(1-\beta)c_H/\beta - k/\beta \leq p^t(\hat{\gamma}) + \zeta_H^t(\hat{\gamma}) - c_H$. L's complete the task for sure in period t *if* $c_L + k < \beta(p^t(\hat{\gamma}^L) + \zeta_L^t(\hat{\gamma}^L))$ or $(1-\beta)c_L/\beta + k/\beta < p^t(\hat{\gamma}) + \zeta_L^t(\hat{\gamma}) - c_L$. $c_H - c_L > 2k/(1-\beta)$ implies that $(1-\beta)c_H/\beta - k/\beta > (1-\beta)c_L/\beta + k/\beta$, and the result follows. □

Proof of Proposition 4. (i) First, we argue that Lemma 4 implies that we cannot do better than the following outcome: there is some period $D \geq 1$ such that L's are selective and H's wait for sure in periods $t < D$, and L's complete the task for sure and H's are selective in periods $t \geq D$. Lemma 4 says that if H's are selective in period t then L's must complete the task for sure in period t. If D is the first period in which H's are selective, clearly we want H's to be selective (i.e., to behave efficiently) in all periods $t \geq D$ because the probability of L's reaching period $t > D$ is zero. By the definition of D, H's wait for sure in any period $t < D$. Clearly, we would like L's to be selective in any period $t < D$.

Second, we show that for any D this outcome can be achieved with the $(\overline{w}_L,\overline{w}_H)$-deadline scheme with deadline D. Given $\overline{w}_H > \overline{w}_L > k$, clearly $\hat{\gamma}_t^L = \hat{\gamma}_t^H = d$ for all t. Hence, for any t, $\gamma_t^i = s$ if and only if $c_i - k \leq \beta(w_t^\Delta + c_i) \leq c_i + k$ or $(1-\beta)c_i/\beta - k/\beta \leq w_t^\Delta \leq (1-\beta)c_i/\beta + k/\beta = \overline{w}_i$. And $c_H - c_L > 2k/(1-\beta)$ implies that $\underline{w}_H = (1-\beta)c_H/\beta - k/\beta > \overline{w}_L$. Therefore, $w_t^\Delta = \overline{w}_L$ for $t < D$ implies that $\gamma_t^L = s$ and $\gamma_t^H = w$ for all $t < D$, and $w_t^\Delta = \overline{w}_H$ for all $t \geq D$ implies that $\gamma_t^L = d$ and $\gamma_t^H = s$ for all $t \geq D$.

Finally, we solve for the optimal deadline. Given D, the expected costs are as follows. Task cost for L's $= \Sigma_{n=1}^{D-1}(½)^n (c_L - k) + [1 - \Sigma_{n=1}^{D-1}(½)^n]\ c_L = c_L - [1-(½)^{D-1}]k$. Delay cost for L's $= \Sigma_{n=1}^{D-1}(½)^n(n-1)x^\Delta + [1-\Sigma_{n=1}^{D-1}(½)^n]\ (D-1)x^\Delta = [1-(½)^{D-1}]x^\Delta$. Task cost for H's $= c_H - k$ (because H's are selective in all periods $t \geq D$). Delay cost for H's $= \Sigma_{n=D}^{\infty}(½)^{n-D+1} (n-1)x^\Delta = Dx^\Delta$.
Hence, the expected total costs are

$$\pi(c_L - [1-(½)^{D-1}]k + [1-(½)^{D-1}]x^\Delta) + (1-\pi)(c_H - k + Dx^\Delta) \equiv Z(D).$$

The only component of the incentive scheme that affects Z is the deadline D, so all second-best incentive schemes will have the same deadline. Z is continuous, twice-differentiable, and $d^2Z/dD^2 > 0$. We have

$$\text{argmin}_D Z(D) \equiv \alpha = 1 + \frac{\ln(\pi/(1-\pi)) + \ln((k-x^\Delta)/x^\Delta) + \ln(\ln 2)}{\ln 2}.$$

However, the optimal deadline D^* must be an integer. Since Z is continuous and $d^2Z/dD^2 > 0$, if $\alpha > 1$ the optimal deadline is either the largest integer less than α or the smallest integer greater than α, and if $\alpha \leq 1$ the optimal deadline is clearly $D^* = 1$.

(ii) If $c_L \geq (1+\beta)k/(1-\beta)$, then $\gamma_t^L = s$ only if $w_t^\Delta \geq \underline{w}_L \geq k$. Hence, any incentive scheme that induces the second-best optimal outcome (i.e., $\gamma_t^L = s$ and $\gamma_t^H = w$ for all $t < D^*$ and $\gamma_t^L = d$ and $\gamma_t^H = s$ for all $t \geq D^*$) will imply that $\hat{\gamma}_t^L = \hat{\gamma}_t^H = d$. Then for any $t \geq D^*$ $\gamma_t^H = s$ only if $w_t^\Delta \geq \underline{w}_H$ and for any $t < D^*$ $\gamma_t^L = s$ only if $w_t^\Delta \leq \overline{w}_L$. By Proposition 3, $c_H - c_L > 2k/(1-\beta)$ implies that $\underline{w}_H > \overline{w}_L$, and the result follows. □

Proof of Proposition 5. Essentially identical to proof of Proposition 3, and so omitted.

Proof of Lemma 5. Define $X_t^i(\hat{\gamma}^i)$ as any perceived reductions in expected future task costs below c_i net of additional lost incremental wages. In other words, the period-t continuation payoff from waiting for type i is $\beta[w_t^\Delta + c_i - X_t^i(\hat{\gamma})^i)]$. Clearly, $X_t^i(\hat{\gamma}^i) \geq 0$, and if we define $\mu^i(t) \equiv \min\{\tau > t | \hat{\gamma}_\tau^i = d\}$, then

$$X_t^i(\hat{\gamma}^i) = \begin{cases} 0 & \text{for } \mu^i(t) = t+1 \\ \displaystyle\sum_{n=t+1}^{\mu^i(t)-1} \left(\frac{1}{2}\right)^{n-t} k_i - \sum_{n=t+1}^{\mu^i(t)-1} \left(\frac{1}{2}\right)^{n-t} w_n^\Delta & \text{for } \mu^i(t) \in \{t+2, t+3, \ldots\}. \end{cases}$$

Using a revealed preference argument, we can prove that $k_H \geq k_L$ implies that $X_t^H(\hat{\gamma}^H) \geq X_t^L(\hat{\gamma}^L)$: $k_H \geq k_L$ clearly implies that $X_t^H(\hat{\gamma}^L) \geq X_t^L(\hat{\gamma}^L)$, and $\hat{\gamma}^H$ represents how TCs would behave and therefore maximizes X_t^H. Hence, we have $X_t^H(\hat{\gamma}^H) \geq X_t^H(\hat{\gamma}^L) \geq X_t^L(\hat{\gamma}^L)$.

For any t, $\gamma_t^L \neq d$ only if $c_L + k_L \geq \beta[w_t^\Delta + c_L - X_t^L(\hat{\gamma}^L)]$ or $w_t^\Delta \leq (1-\beta)c_L/\beta + k_L/\beta + X_t^L(\hat{\gamma}^L) = \overline{w}_L + X_t^L(\hat{\gamma}^L)$. Similarly, for any t $\gamma_t^H = w$ if $c_H - k_H > \beta[w_t^\Delta + c_H - X_t^H(\hat{\gamma}^H)]$ or $w_t^\Delta < (1-\beta)c_H/\beta - k_H/\beta + X_t^H(\hat{\gamma}^H)$. And $c_H \geq (1+\beta)k_H/(1-\beta)$ implies that $\underline{w}_H = (1-\beta)c_H/\beta - k_H/\beta$, so the inequality becomes $w_t^\Delta < \underline{w}_H + X_t^H(\hat{\gamma}^H)$. Using $\underline{w}_H > \overline{w}_L$ and $X_t^H(\hat{\gamma}^H) \geq X_t^L(\hat{\gamma}^L)$, the result follows. □

Proof of Proposition 6. Essentially identical to proof of Proposition 4, and so omitted. (Proposition 6 follows from Lemma 5 in exactly the same way that Proposition 4 follows from Lemma 4.)

Proof of Lemma 6. We first prove that for any t, $\gamma_t^L \neq d$ implies that $\hat{\gamma}_t^H = s$. Define $X_t^i(\hat{\gamma}^i)$ as in the proof of Lemma 5, and again $k_H \geq k_L$ implies that $X_t^H(\hat{\gamma}^H) \geq X_t^L(\hat{\gamma}^L)$ for all t. For any t, $\gamma_t^L \neq d$ only if $c_L + k_L > \beta[w_t^\Delta + c_L - X_t^L(\hat{\gamma}^L)]$ or $w_t^\Delta < \overline{w}_L + X_t^L(\hat{\gamma}^L)$. For any t, $\hat{\gamma}_t^H = s$ if $c_H + k_H > w_t^\Delta + c_H - X_t^H(\hat{\gamma}^H)$ or $w_t^\Delta < k_H + X_t^H(\hat{\gamma}^H)$. $c_H < (1+\beta)k_H/(1-\beta)$ implies that $k_H > \underline{w}_H > \overline{w}_L$, which along with $X_t^H(\hat{\gamma}^H) \geq X_t^L(\hat{\gamma}^L)$ establishes that $\gamma_t^L \neq d$ implies $\hat{\gamma}_t^H = s$.

Now suppose that $\gamma_t^L = s$ for all t, so $\hat{\gamma}_t^H = s$ for all t. Then for any t, $\gamma_t^H \neq w$ only if $c_H - k_H \leq \beta[\Sigma_{\tau=t}^{\infty}(1/2)^{\tau-t}\ w_\tau^\Delta + c_H - k_H]$ or $\Sigma_{\tau=t}^{\infty}(1/2)^{\tau-t+1}\ w_\tau^\Delta \geq (1-\beta)(c_H - k_H)/(2\beta) = \underline{w}_H$ (since $c_H < (1+\beta)k_H/(1-\beta)$. But the logic of Lemma 3 implies that $\gamma_t^L \neq d$ for all t only if $\Sigma_{\tau=t}^{\infty}(1/2)^{\tau-t+1}\, w_\tau^\Delta \leq \overline{w}_L$ for all t. Given $\underline{w}_H > \overline{w}_L$, the result follows.

Proof of Proposition 7. (i) Lemma 6 implies that there must be some period in which L's complete the task for sure because otherwise H's wait forever. Define $D \equiv \min\{t|\gamma_t^L = d\}$ as the deadline.

For $k_H \geq k_L$ and $c_H < (1+\beta)k_H/(1-\beta)$, it is possible to have $\gamma_t^H \neq w$ for some $t < D$. To maximize the likelihood of H's being selective *before* the deadline, it is optimal to make the incentive scheme as steep as possible. This means that we want $w_t^\Delta = \overline{w}_H$ for $t \geq D$ and $w_t^\Delta = \overline{w}_L$ for $t < D$; that is, we want a $(\overline{w}_L, \overline{w}_H)$-deadline scheme. (Note: this result relies on the fact that for $t < D$ we cannot have an incentive scheme that is steeper "on average" than $w_t^\Delta = \overline{w}_L$ for all $t < D$. This fact follows from the same logic as in the proof of Lemma 3 part (i)—it is straightforward to show that if $\gamma_D^L = d$ and $\gamma_t^L \neq d$ for all $t < D$, then for any $\tau < D$ we must have $\Sigma_{t=\tau}^{D-1}(1/2)^{t-\tau}\, w_t^\Delta \leq \Sigma_{t=\tau}^{D-1}(1/2)^{t-\tau}\, \overline{w}_L$.)

Now consider behavior under the $(\overline{w}_L, \overline{w}_H)$-deadline scheme with deadline D. It is obvious that $\gamma_t^L = s$ for $t < D$ and $\gamma_t^L = d$ for $t \geq D$. It is also obvious that $\gamma_t^H = s$ for all $t \geq D$. Consider γ_t^H for $t < D$. First, note that $\hat{\gamma}_t^H = d$ for all $t \geq D$ (since for all $t \geq D$, $w_t^\Delta = \overline{w}_H > k_H$), and that $\hat{\gamma}_t^H = s$ for all $t < D$ (which follows from the proof of Lemma 6). That is, for all $t < D$, H's perceive that they will be selective before the deadline and complete the task for sure at the deadline. Then in period $D - n'$, $n' \in \{1,2,\ldots\}$, we have $\gamma_t^H = s$ if and only if $c_H - k_H \leq \beta[\Sigma_{j=0}^{n'-1}(1/2)^j\, \overline{w}_L + c_H + k_H - \Sigma_{j=0}^{n'-1}(1/2)^j\, k_H]$ or $(1-\beta)c_H/\beta - k_H/\beta \leq k_H - \Sigma_{j=0}^{n'-1}(1/2)^j\,(k_H - \overline{w}_L)$. Hence, given the definition of $\hat{d}$, $\gamma_t^H = w$ for $t < D - \hat{d}$, and $\gamma_t^H = s$ for $t \geq D - \hat{d}$ (note that n in the proposition corresponds to $n' - 1$ here).

Finally, consider the optimal deadline. Using the same method as in the proof of Proposition 4, we get the following equation for expected total costs, which we denote by a function $\hat{Z}$:

$$\hat{Z}(D,\hat{d}) = \begin{cases} \pi[c_L - (1-(1/2)^{D-1})k_L \\ \quad + (1-(1/2)^{D-1})x^\Delta] & \text{for } D > \hat{d}+1 \\ \quad + (1-\pi)[c_H - k_H + (D-\hat{d})x^\Delta] & \\ \pi[c_L - (1-(1/2)^{D-1})k_L \\ \quad + (1-(1/2)^{D-1})x^\Delta] & \text{for } D \leq \hat{d}+1. \\ \quad + (1-\pi)[c_H - k_H + x^\Delta] & \end{cases}$$

For $D \leq \hat{d} + 1$, $\hat{Z}$ is decreasing in D, so we must have the optimal deadline $D^{**} \geq \hat{d} + 1$. For $D > \hat{d} + 1$, it is straightforward to show that the optimal deadline is D^* defined in Proposition 6 provided that $D^* > \hat{d} + 1$. Hence, the optimal deadline $D^{**} = \min\{D^*, \hat{d} + 1\}$.

(ii) Essentially identical to the proof of Proposition 4 part (ii), and so omitted. □

Proof of Proposition 8. We first prove the analog of Lemma 2 for sophisticates: a stationary incremental wage w^Δ can induce $\bar{\gamma}^i = (s,s, \ldots)$ for $i \in \{L, H\}$ if and only if $\underline{w}^i \leq w^\Delta \leq \overline{w}^i$, where $\underline{w}^i$ and $\overline{w}^i$ are given by $\underline{w}^i = (1 - \beta)c_i/(2\beta) - (1 - \beta)k/(2\beta)$ and $\overline{w}^i = (1 - \beta)c_i/(2\beta) + (1 + \beta)k/(2\beta)$. Since sophisticates correctly predict future behavior, any w^Δ that induces $\bar{\gamma}^i = (s,s, \ldots)$ implies $p^t(\bar{\gamma}^i) = 2w^\Delta$ and $\zeta_i^t(\bar{\gamma}^i) = c_i - k$ for all t. Hence, for stationary incremental wage w^Δ we have $\bar{\gamma}^i = (s,s, \ldots)$ if and only if $c_i - k \leq \beta[2w^\Delta + c_i - k] \leq c_i + k$, or $(1 - \beta)c_i/(2\beta) - (1 - \beta)k/(2\beta) \leq w^\Delta \leq (1 - \beta)c_i/(2\beta) + (1 + \beta)k/(2\beta)$.

Defining $\underline{w}^i$ and $\overline{w}^i$ as above, we next prove the analog of Lemma 3 for sophisticates: any incentive scheme W that induces $\bar{\gamma}^i = (s,s, \ldots)$ satisfies $\underline{w}^i \leq \Sigma_{t=1}^{\infty}(1/2)^t\, w_t^\Delta \leq \overline{w}^i$. Since sophisticates correctly predict future behavior, any incentive scheme W that induces $\bar{\gamma}^i = (s,s, \ldots)$ implies that $p^1(\bar{\gamma}^i) = 2\,\Sigma_{t=1}^{\infty}(1/2)^t\, w_t^\Delta$ and $\zeta_i^1(\bar{\gamma}^i) = c_i - k$. Then we have $\bar{\gamma}_1^i = s$ if and only if $c_i - k \leq \beta[2\,\Sigma_{t=1}^{\infty}(1/2)^t\, w_t^\Delta + c_i - k] \leq c_i + k$, which yields $\underline{w}^i \leq \Sigma_{t=1}^{\infty}(1/2)^t \leq \overline{w}^i$.

These two results imply that a stationary incentive scheme can induce efficiency for both types if $\underline{w}^H \leq \overline{w}^L$, and no incentive scheme can induce efficiency for both types if $\underline{w}^H > \overline{w}^L$. It is straightforward to show that $\underline{w}^H \leq \overline{w}^L$ if and only if $c_H - c_L \leq 2k/(1 - \beta)$, and the result follows. □

CORNELL UNIVERSITY
UNIVERSITY OF CALIFORNIA AT BERKELEY

REFERENCES

Ainslie, George, "Impulse Control in Pigeons," *Journal of the Experimental Analysis of Behavior,* XXI (1974), 485–489.

——, "Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control," *Psychological Bulletin,* LXXXII (1975), 463–496.

——, "Self-Reported Tactics of Impulse Control," *International Journal of the Addictions,* XXII (1987), 167–179.

——, "Derivation of 'Rational' Economic Behavior from Hyperbolic Discount Curves," *American Economic Review,* LXXXI (1991), 334–340.

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

Ainslie, George, and Nick Haslam, "Self-Control," in George Loewenstein and Jon Elster, eds., *Choice Over Time* (New York, NY: Russell Sage Foundation, 1992a).

Ainslie, George, and Nick Haslam, "Hyperbolic Discounting," in George Loewenstein and Jon Elster, eds., *Choice Over Time* (New York, NY: Russell Sage Foundation, 1992b).

Ainslie, George, and Richard Herrnstein, "Preference Reversal and Delayed Reinforcement," *Animal Learning and Behavior,* IX (1981), 476–482.

Akerlof, George, "Procrastination and Obedience," *American Economic Review,* LXXXI (1991), 1–19.

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

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

Funder, David, and Jack Block, "The Role of Ego-Control, Ego-Resiliency, and IQ in Delay of Gratification in Adolescence," *Journal of Personality and Social Psychology,* LVII (1989), 1041–1050.

Goldman, Steven, "Intertemporally Inconsistent Preferences and the Rate of Consumption," *Econometrica,* XLVII (1979), 621–626.

——, "Consistent Plans," *Review of Economic Studies,* XLVII (1980), 533–537.

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

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

Laibson, David, "Essays in Hyperbolic Discounting," Ph.D. Thesis, Massachusetts Institute of Technology, 1994.

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

——, "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics,* CXII (1997), 443–477.

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

O'Donoghue, Ted, and Matthew Rabin, "Procrastination in Preparing for Retirement," Cornell University and University of California at Berkeley, 1998a.

O'Donoghue, Ted, and Matthew Rabin, "Choice and Procrastination," Cornell University and University of California at Berkeley, 1998b.

O'Donoghue, Ted, and Matthew Rabin, "Doing It Now or Later," *American Economic Review,* LXXXIX (1999), 103–124.

Phelps, E. S., and Robert Pollak, "On Second-Best National Saving and Game-Equilibrium Growth," *Review of Economic Studies,* XXXV (1968), 185–199.

Pollak, Robert, "Consistent Planning," *Review of Economic Studies,* XXXV (1968), 201–208.

Prelec, Drazen, "Decreasing Impatience: Definition and Consequences," Harvard Business School, 1989.

Strotz, Robert, "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies,* XXIII (1956), 165–180.

Thaler, Richard, "Mental Accounting and Consumer Choice," *Marketing Science,* IV (1985), 199–214.

——, "Some Empirical Evidence on Dynamic Inconsistency," in Richard Thaler, ed., *Quasi Rational Economics* (New York, NY: Russell Sage Foundation, 1991).