We examine a mechanism design problem in which the principal owns a project which requires work effort by an agent, but agents may have time-inconsistent, present-biased preferences and imperfect foresight or “naïveté”. Present-biased agents value future effort and payoffs differently at the time of agreeing to a contract than when they must choose their effort, which leads to a second-best outcome for the principal. Limited self-awareness or “sophistication” among agents enables the principal to screen some present-biased agents from the contracting set, but at the cost of generating shirking behavior.
I Introduction

A recent body of work has begun to examine the behavior of individuals with a time-inconsistent preference for procrastination—that is, individuals who do not merely discount the future but who tend to pursue immediate gratification contrary to what their "long-run selves" would prefer. An important goal of this literature is to understand the implications not just of such present-biased preferences, but also of individuals’ self-awareness. That is, does it matter whether a person knows he has a tendency to procrastinate? One set of questions that arises in this context is, how is the mechanism design problem affected for a principal who wishes to induce optimal performance of some task by an agent or agents who might be present-biased?

O'Donoghue and Rabin [1999b] consider a setting in which the principal wishes to induce an agent to complete an unpleasant task at some time in the future. Delay is costly to the principal, but the agent faces stochastic costs of completing the task which imply that it is efficient for the agent to delay when costs are high. They show that if the agent has private information about the distribution of task-costs the principal can induce first-best efficient behavior by time-consistent agents, but may not be able to induce first-best behavior by time-inconsistent procrastinators.

Whereas O'Donoghue and Rabin consider a case in which a single agent, who may be present-biased, is matched to the principal and examine when the task will be completed, we focus on a setting where the principal hires an agent from a pool of potential agents, some of whom are present-biased, and address whether the agent who agrees to a contract will in fact choose to complete a task which must be undertaken at a particular time, and whether he will exert the efficient level of effort. The problem we study is thus really whether present-biased
preferences may lead to shirking, rather than procrastination. Central to this problem is whether the principal, who cannot observe agents’ types, can offer a contract that successfully screens out highly present-biased agents. If not, the contract design question for the principal is how to induce optimal effort by agents when some have time-consistent preferences while others do not.

The model presented here is quite similar to a classic hidden information principal-agent setting in which the principal wishes to hire an agent to perform a one-time task, the agent’s effort is fully observable, but after the contract is signed there is a random realization of the agent’s disutility from work effort which the principal does not observe. That is, the agent may find that he is well suited to the task and the cost of effort is low, or he may find that he is not and effort is more costly. The contract design problem for the principal is that, in order to generate an incentive for the agent to truthfully reveal the state of the world and exert higher effort if the cost of effort is low, the contract must compel the agent to bear some of the risk, which is inefficient because the agent is commonly assumed to be risk averse while the principal is risk neutral. One of the goals of this paper is to illustrate that the possibility an agent is present-biased has analogous implications for contracting when the task must be completed prior to the observation of the resulting profit. Because of the time lag separating the work effort and payment to the agent (which is made after profit is observed), present-bias increases the disutility of effort relative to the future payment. As with other hidden information settings, we find that optimal contract design induces first-best efficient effort by the "high" type agents (in this case those with time-consistent preferences), the common "no distortion at the top" result, while low-type (present-biased) agents will be induced to exert suboptimal effort. Furthermore, time-

---

1 See, for example, Shavell (1979).
2 Note that in this setting if agents are not risk averse the principal can achieve an efficient outcome by selling the project to the agent. A similar issue arises in our context: if agents are no more impatient than the principal then selling the project may be optimal. This issue is explore in Appendix A.
consistent agents earn rents which are necessary to make the decision to exert greater effort incentive compatible. Interestingly, we find that while present-biased agents are typically held to their second period present-bias participation constraint, they nevertheless earn positive utility when viewed from the perspective of their long run time-consistent preferences, and in fact do just as well as time-consistent agents.

Where the implications of present-bias differ markedly from other hidden information contracting problems is in the consideration of self-awareness among agents. If some agents are aware at the time of contracting that they will be present-biased when the time comes to complete task, the principal may do well to offer a contract that screens these self-aware (or sophisticated) present-biased agents from the contracting pool. However, the trade-off is that, by designing a contract that will be unattractive to self-aware present-biased agents, those that lack such self-awareness may agree to the contract but later shirk because they find the incentive payment insufficient to overcome their desire to put off costly effort. Nevertheless, we show that in some settings the principal may do best to accept some shirking in order to reduce rents to less present-biased agents.

The following section describes present-biased preferences and the role they will play in the contracting problem we study. Section 3 presents a fairly simple model in which agents are of only 2 types, present-biased or time-consistent (but not allowing for varying degrees of present-bias), and demonstrates how present-biased preferences lead to a second-best contracting outcome. We then develop a richer model in section 4 that allows for the degree of present-bias to vary continuously in the agent population and, importantly, permits us to explore the implications of partial self-awareness among agents of their time-inconsistent preferences. We
conclude in section 5 by considering the welfare implications of present-bias for agents in our model.

2 Present-Biased Preferences

O'Donoghue and Rabin [1999a] coined the term "present-biased" preferences to refer to a class of time-inconsistent preferences in which individuals, when considering trade-offs between two future moments, give greater weight to the earlier moment as it gets closer. Unlike an individual with time-consistent preferences for whom the marginal rate of substitution between any two periods $t+1$ and $t$ is a constant, for one with present-biased preferences the marginal rate of substitution between two periods may depend on the time at which a choice is evaluated.\(^3\) One method of modeling present-biased preferences is "hyperbolic discounting" in which an individual's discount function exhibits a relatively high discount rate over short time horizons and a relatively low discount rate over long time horizons.\(^4\) A simpler model for capturing the salience of the present over the future merely incorporates an additional parameter, $\beta$, into the standard time-consistent model of intertemporal preferences to capture an individual's bias in favor of the present over all future periods. This framework, which has been referred to as "quasi-hyperbolic discounting" or "$(\beta, \delta)$-preferences", was first suggested by Phelps and Pollack [1968], has been broadly utilized in recent literature and we employ it throughout our analysis.\(^5\) Let $u_t$ be the instantaneous utility an individual receives at period $t$. Then the following utility function, $U_t$, represents his intertemporal preferences at time $t$:

\(^3\) Caillaud & Jullien [2000].

\(^4\) For a discussion of the hyperbolic discount function and evidence supporting its application see Loewenstein and Prelec [1992].

\(^5\) Examples of papers utilizing this preference structure are O'Donoghue and Rabin [1999a, 1999b], Laibson [1997] and Caillaud and Jullien [2000].
The parameter $\delta$ is the usual time-consistent discounting term while $\beta \in (0,1]$ captures the possible time-inconsistent preference for present-period utility compared to all future periods.

The standard approach to modeling individuals with time-inconsistent preferences is to consider a person at each time period a separate agent who maximizes utility with regard to his current preferences while his "future selves" will determine future behavior according to the preferences that then prevail [O'Donoghue and Rabin, 1999a]. An important question that follows from this assumption is what does a person believe about his future selves' preferences? At one extreme individuals may be fully self-aware or "sophisticated", meaning they have perfect foresight regarding their future preferences and know exactly what their future selves will prefer even though these preferences differ from those of the current self. At the other extreme individuals may be described as naïve because they have imperfect foresight and they believe their future preferences will be identical to their current preferences, systematically failing to realize that as an event becomes closer in time their preferences will change. Economists have debated whether to model individuals as naïve or sophisticated. The assumption of sophistication entails rational expectations of future behavior, but naiveté captures the often observed human tendency to put off unpleasant tasks today in the belief that tomorrow we will have the willpower or self-discipline to do them.6 Present-biased preferences reflect a self-control problem: viewed from a prior (or "long-run") perspective, a present-biased person wants to behave relatively patiently, but at the moment a choice must be made he wants to behave impatiently
The issue of naïveté versus sophistication is thus a question of awareness of one’s self-control problem. O’Donoghue and Rabin (2001b) review evidence of peoples’ awareness of self-control problems such as procrastination and find evidence of both naïve and sophisticated behavior, although evidence regarding self awareness is far more limited than that indicating the existence of the self-control problem itself. As they suggest, it is likely that people are somewhat sophisticated and partially aware of their tendency to procrastinate, but that they fail to foresee the full extent of future self-control problems.

We will initially consider the two extreme cases: agents who are either completely sophisticated or completely naïve, labeled "sophs" and "naifs" by O'Donoghue and Rabin [1999a]. In our setting a soph is an agent whose type is private information hidden from the principal but known to the agent. A naif is one whose type is a complete unknown—the principal has no information about a naif’s type while the agent himself is mistaken in believing all his future selves will be time-consistent although his current self is not. The contract design problem we study arises because the imperfect foresight of naïve present-biased agents prevents any contract from screening them out of the set of agents in the contracting pool while remaining acceptable to time-consistent agents. It is important to note from the outset that an individual with present-biased preferences, and in particular a naif who fails to foresee his future self-control problems, is qualitatively different from one who is simply less patient due to greater time-consistent discounting. The critical difference from the standpoint of the principal’s mechanism design problem is that a naif may accept a contract offer in the belief that he will complete the assigned task, but later fail to do so as a result of his self-control problem. This problem does not arise with time-consistent individuals, regardless of their impatience or

---

6 For much more discussion of the implications of modeling individuals as naïve or sophisticated and the arguments for each see O'Donoghue & Rabin [1999a].
discount rate. We show in section 4 that present-biased preferences have important implications for contracting behavior even when the population exhibits only a small degree of naiveté.\footnote{This result, that even a little naïveté results in very different consequences than full sophistication, has also been observed in other contexts by O’Donoghue and Rabin (1999a, 2001a, 2001b) and Della Vigna (2003).}

We assume that agents have $(\beta, \delta)$-preferences and that all agents have a common time-consistent discounting parameter $\delta$ which is known to the principal. This assumption is not critical to our model but simplifies the presentation. The critical assumption is that total per-period discounting by naïve agents exceeds that of time-consistent agents, i.e. $\beta \delta_{\text{naif}} < \delta_{\text{TC}}$. This is simply the condition for the existence of the problem we study, which is that the principal would do better if she could observe agents’ types and hire only time-consistent agents, but she cannot and naïve present-biased agents will accept any offer that is acceptable to a time-consistent agent. We assume the principal does not have present-biased preferences and we represent her time-consistent discount parameter by $\rho$.

In what follows we consider a one-time interaction between a principal and agent. The principal owns a project which must be done at a particular time and requires the labor of an agent. The interaction between the principal and agent takes place over three periods. In period 1 the principal is randomly matched with a single agent from a population of potential agents who find the contract acceptable but differ in type (that is, some are TCs, some are naifs) which the principal cannot observe. The task to be performed must be accomplished in period 2, but the owner of the project will not observe the profit from the work done until period 3. The contract specifies how the agent will be compensated in period 3 and the wage can depend only on information available to both parties.\footnote{We assume up-front payments cannot be recaptured if effort is unsatisfactory and therefore all incentive pay must take place in period 3 when profit is revealed.} This sequence captures the fact that the payoff from work
is often not realized until some time after the effort is undertaken and represents a circumstance for the agent of immediate costs with delayed rewards when evaluated at the time that the choice of effort level is made in period 2.

The principal's problem in this environment is that an agent with present-biased preferences will have a different marginal rate of substitution between her payoff in period 3 and in period 2 when evaluated at period 2 when the choice of effort is made, than at period 1, when the contract is signed. Sophs would anticipate this and will not sign a contract if completing the task in period 2 will not be desirable for their period 2 "self". Naifs, however, will fail to foresee that at period 2 they will prefer to avoid the unpleasant task despite the loss of payment in period 3. The contract therefore must account for the possibility that an agent is a naif and create incentives which maximize the principal’s expected payoff given this possibility.

3 The Contracting Model with Two Discrete Types of Agents

The principal and the agent contract in period 1 over a one-dimensional signal, the profit of the firm, $\pi$, which is observable in period 3. The firm's profit is a deterministic function of the effort exerted by the agent, $e$. Agents have $(\beta, \delta)$-preferences as described earlier. For the moment we will assume that there are two types of agents: TCs, for whom $\beta=1$, and naifs, for whom $\beta<1$ but who, being naïve, believe they will have time-consistent preferences in the future. The share of naifs in the population is represented by $\alpha$. Later we discuss how the possibility of some agents being sophs affects our results. Initially we assume all naifs have identical preferences (equivalent $\beta$ values) in order to examine a simple two-type model. In the next section we consider the case in which there is a continuous distribution of $\beta$ in the population. Agents are identical in all respects other than time-preference (e.g. reservation utilities are equal) and their
types are not observable to the principal. Agents are assumed to be risk-neutral. We denote the payment the agent receives in period three from the principal by $w \geq 0$. The cost to the agent of exerting effort $e$ in period 2 is represented by $c(e)$. We normalize the agent's reservation utility to be zero in each period.

An agent's utility function evaluated at period 1, when he chooses whether to contract with the principal, is

$$U = -\beta c(e) + \beta \delta^2 w.$$ 

In period 2 when the agent must choose his effort level his utility function is

$$U = -c(e) + \beta \delta w.$$ 

For time-inconsistent agents for whom $\beta < 1$ the marginal rate of substitution between payments in period 3 and effort in period 2 differs between the time the contract is signed and the time effort is chosen due to the immediacy of unpleasant effort in period 2.

The principal, whom we assume to be risk-neutral as well, simply maximizes the expected difference between her profit, $\pi$, and payment to the agent, $w$, in period 3. Since we have assumed profit is fully determined by effort, uncertainty arises only from the fact that the agent may choose different effort levels depending on his type. We assume that although the principal cannot observe the agents’ types she knows the distribution of types in the population of agents.

Effort and profit can take on any positive real number value with $c'(e) > 0, c''(e) > 0, \pi'(e) > 0, \pi''(e) < 0$. In order to assure an interior solution we assume

---

9 Note that our assumption that the payment from the principal to the agent in period 3 is non-negative is not trivial. If the contract could specify that the agent pay the principal a penalty fee for low effort then the principal could ensure that any agent who signed a contract in period 1 would exert high effort in period 2 even if doing so provided
\[ \pi'(e) \to \infty \text{ as } e \to 0 \]

\[ \pi'(e) \to 0 \text{ as } e \to \infty . \]

The principal's mechanism design problem is now to design a contract that induces the effort from each type of agent that maximizes her payoff. If agents' types were observable, the principal would simply offer a contract which maximized

\[
\max_e \pi(e) - w \text{ subject to } \beta \delta \geq \frac{c(e)}{\beta \delta},
\]

the solution to which is defined by the first-order condition

\[
\pi'(e) - \frac{c'(e)}{\beta \delta} = 0 .
\]

Note that this condition implies that the optimal effort level decreases as \( \beta \) decreases since the more present-biased an agent's preferences, the more costly it is to induce effort in period 2 because payment will not be made until period 3. So the principal will wish to offer a contract that induces lower effort in a naïf agent than in a TC.

We turn now to the hidden information case. Note first that in some circumstances the principal may do better, and a more efficient outcome may be obtainable, if the project is sold to an agent. In a classic hidden information setting an efficient outcome can be obtained if agents are risk-neutral by the project being sold to an agent, and contracting entails a trade-off between optimal risk sharing and generating incentives for the agent to exert high effort.\(^{10}\) In Appendix A we consider the possibility of selling the project and the tradeoff that exists in this context

\(^{10}\) See Shavell (1979).

---

negative utility to the period 2 present-biased naïf. In such a case naïfs' preferences would result in their signing contracts that they would later regret, but would not harm the principal.
between agent incentives and optimal ownership as a function of investor patience or
discounting.

We will abuse notation slightly by using $\beta$ to refer to the specific present-bias preference
parameter of a naif, and we will not include the $\beta$ term when considering TCs for who it has a
value of 1. Any contract that satisfies a naif's participation constraint in period 2 will provide a
TC with positive utility (since TCs will value the future payment more highly than will naifs).
Therefore, if it is to be incentive compatible for TCs to put forth greater effort than naifs, the
contract must enable them to earn rents. Define $e_{TC}$ to represent the effort level the principal
wishes to induce in TCs and $e_N$ to be the effort level she wishes to induce in naifs. In order to
achieve participation by both types and ensure TCs don't pretend to be naifs, the contract must
satisfy the naifs' individual rationality constraint for period 2 and the TCs incentive compatibility
constraint as follows

\[ \text{IRC}_{\text{naif}}: \quad -c(e_N) + \beta \delta w(e_N) \geq 0 \quad \text{or} \quad (4) \]

\[ w(e_N) \geq \frac{c(e_N)}{\beta \delta} \quad \text{and} \quad (5) \]

\[ \text{ICC}_{\text{TC}}: \quad -c(e_{TC}) + \delta w(e_{TC}) \geq -c(e_N) + \delta w(e_N) \quad \text{or} \quad (6) \]

\[ w(e_{TC}) - w(e_N) \geq \frac{c(e_{TC}) - c(e_N)}{\delta}. \quad (7) \]

If the incentive compatibility constraint for TCs is satisfied, the individual rationality constraint
for TCs does not bind. Furthermore, because the incentive compatibility constraint for TCs
binds, the incentive compatibility constraint for naifs must be slack since the cost of additional
effort to both types is equal but naifs value the additional future payment less than TCs.
The principal will maximize her payoff when the two constraints hold with equality.

Substituting the expression for $w(e_N)$ into the incentive compatibility constraint for TCs gives the following expression

$$w(e_{TC}) = \frac{c(e_{TC})}{\delta} + \frac{(1 - \beta)c(e_N)}{\beta\delta}.$$  \hfill (8)

The second part of the right-hand-side sum represents information rents to TC agents.

Using the results above for $w(e_N)$ and $w(e_{TC})$ we can write the principal's maximization problem as

$$\max E[\pi - w] = (1 - \alpha) \left[ \pi(e_{TC}) - \left( \frac{c(e_{TC})}{\delta} + \frac{(1 - \beta)c(e_N)}{\beta\delta} \right) \right] + \alpha \left[ \pi(e_N) - \frac{c(e_N)}{\beta\delta} \right].$$  \hfill (9)

The first-order conditions are

$$e_{TC}: \pi'(e_{TC}) - \frac{c'(e_{TC})}{\delta} = 0$$  \hfill (10)

$$e_N: -(1 - \alpha)\frac{(1 - \beta)c'(e_N)}{\beta\delta} - \frac{\alpha c'(e_N)}{\beta\delta} + \alpha\pi'(e_N) = 0.$$  \hfill (11)

These conditions define the solution to the principal's mechanism design problem. Optimal effort levels are described by

$$\pi'(e_{TC}) = \frac{c'(e_{TC})}{\delta}$$  \hfill (12)

$$\pi'(e_N) = \frac{c'(e_N)}{\beta\delta} \left[ 1 + \frac{(1 - \alpha)(1 - \beta)}{\alpha} \right].$$  \hfill (13)

We can now state three propositions that characterize the optimal contract in this environment.
PROPOSITION 1: TCs earn rents that increase with the share of naifs in the population, but the effect of the degree of present-bias in naifs is ambiguous. (All proofs in Appendix B.)

The first part of this result comes from the fact that as $\alpha$ increases $e_N$ increases, which increases rents. However, the effect of $\beta$ is ambiguous because although $e_N$ increases with $\beta$ the direct effect of $\beta$ on rents is negative.

PROPOSITION 2: The optimal contract induces efficient effort by a TC agent but too little effort by a naïf.

PROPOSITION 3: The magnitude of the distortion from the naïf's first-best effort increases with the degree of naïf's present-bias (i.e. decreases with $\beta$), and decreases with the share of naifs in the population.

We now turn to examining how the presence of sophs affects the contracting problem. To understand this case, we start by recognizing that if a soph agrees to a contract in period 1 he will behave in period 2 in exactly the same manner as a period 2 naïf. This is because we assume both have the same degree of present-bias (or value of $\beta$) and because in period 2 no belief about a future choice impacts the individual's expected payoff, therefore it does not matter what the individual's beliefs are about his future preferences. For this reason, if the contract does not screen sophs from the contracting set, the principal will contract in period 1 with each type of potential agent with probability equal to that type's share in the population and it is straightforward to describe the principal's optimal contract design. Because sophs and naifs behave identically in period 2 the principal is concerned only with the joint share of these types
in the population, rather than the share of naifs alone as discussed in section 2. Let $\gamma$ represent the share of sophs in the population with $\alpha$ indicating the share of naifs as before. Then the principal's expected payoff is just as shown earlier but with $(\alpha + \gamma)$ replacing $\alpha$ and with $e_N$ now representing the effort induced from either or naif or soph:

$$E[\pi - w] = (1 - \alpha - \gamma) \left[ \pi(e_{TC}) - \left( \frac{c(e_{TC})}{\delta} + \frac{(1 - \beta)c(e_N)}{\beta \delta} \right) \right] + (\alpha + \gamma) \left[ \pi(e_N) - \frac{c(e_N)}{\beta \delta} \right], \quad (14)$$

where

$$\pi'(e_{TC}) = \frac{c'(e_{TC})}{\delta} \quad \text{and} \quad (15)$$

$$\pi'(e_N) = \frac{c'(e_N)}{\beta \delta} \left[ 1 + \frac{(1 - \alpha - \gamma)(1 - \beta)}{\alpha + \gamma} \right] \quad (16)$$

Note that the assumptions we made in section 2 that $\pi'(e) \to \infty$ as $e \to 0$ and $\pi'(e) \to 0$ as $e \to \infty$ ensured an interior solution in which the principal always did best to offer a contract that induced some positive effort by naifs as well as TCs. However, these conditions no longer ensure such a solution because with sophs in the population the principal can increase the share of TCs with whom she contracts in period 1 by offering a contract a soph will not accept. In the case discussed in section 2 with only TCs and naifs in the population the principal could not construct a contract to screen out naifs, and thereby affect the relative share of each type with which she contracted, because any contract acceptable to a TC is also acceptable to a period 1 naif. Now the principal has a choice: she can offer a contract which will be accepted by all types and induce positive effort from all types thereby generating a payoff to the principal given by (14) above, or she can screen out sophs by offering a contract which provides negative utility to a period 2 present-biased agent if he chooses any positive effort level. It is precisely a soph's
awareness of his future present-biased preferences that makes it possible to screen them out. Of course such a contract that screens out sophs does not screen out naifs but results in their choosing $e=0$ in period 2 which is costly to the principal. This may nevertheless be optimal. The advantage to the principal is that she can now induce optimal effort by TCs while holding them to their binding individual rationality constraint (thus eliminating information rents) and reduce the total share of present-biased agents with whom she contracts by offering $w = \frac{c(e_{TC})}{\delta}$ if $\pi(e_{TC})$ is observed and $w = 0$ otherwise. The expected payoff to the principal from offering such a contract which results in hiring a TC with probability $\frac{(1-\alpha - \gamma)}{(1-\gamma)}$ and a naif with probability $\frac{\alpha}{(1-\gamma)}$ will be

$$E[\pi - w] = \frac{(1-\alpha - \gamma)}{(1-\gamma)}\left[\pi(e_{TC}) - \frac{c(e_{TC})}{\delta}\right] + \frac{\alpha}{(1-\gamma)}(0). \tag{17}$$

**PROPOSITION 4:** When sophs are present in the population of agents, for a given $\beta$, $\delta$ and $\gamma > 0$, for a sufficiently small $\alpha$ the principal prefers to offer a contract that will screen out sophisticated present-biased agents (who will choose not to accept the contract). In this case a TC will earn his reservation utility and a naif will agree to a contract but later choose not to perform the task.

To see that the payoff from a contract that screens out sophs given by expression (17) may be greater than that from a contract that is accepted by all types given by expression (14), let $\alpha$ go to zero while $\gamma$ remains positive. In this case in which sophs become the dominant type of
present-biased individuals, the principal does better to offer a contract which they will reject. The principal then receives a higher payoff if she contracts with a TC. Against this she balances the cost of receiving a zero payoff if the agent she contracts with is a naif, but if naifs are relatively rare that trade-off may be worth it. The greater the share of sophs relative to naifs in the population and the more present-biased these time-inconsistent agents are, the more likely the principal will optimize by designing a contract that will induce effort only from time-consistent agents.

4 Contracting with a Continuum of Types and Partial Self-Awareness

Here we consider the model when, rather than two discrete $\beta$-types of agents, the distribution of agent types in the population is continuous. This richer model will enable us to examine how partial self-awareness affects the contracting outcome. Let $F(\beta)$ represent the distribution of agent types on $[0, \beta]$ with $f(\beta)$ representing the density function on $[\beta, 1]$. However, because some agents may be less than fully aware of their present-bias, let the distribution of agents’ beliefs concerning their types be represented by $G(\beta)$. We will assume that the distribution $F(\beta)$ is smooth and continuous with no mass points (no mass in the distribution at $\beta=1$ implies that all agents are at least somewhat present-biased). Note that if all agents were fully sophisticated the distribution of beliefs, $G(\beta)$, would be identical to the actual distribution $F(\beta)$, whereas if all agents were completely naïve the $G(\beta)$ distribution would simply be a mass point at $\beta=1$. In fact, as in the previous section, it will be useful to first examine this case of full naïveté among agents.

As we’ve seen, when all agents are fully naïve it is not possible for any contract to screen out more present-biased agents. Therefore the optimal contract will support positive effort by
agents of all types.\footnote{The assumption made earlier to ensure an interior solution also imply that that the optimal contract in this case supports positive effort by all agents.} As before the agent's utility function at the time of his choice of effort in period 2 is

\[ u = -c(e) + \beta \delta w. \tag{18} \]

In order to fit our model into the standard framework in which an agent's type interacts with his choice of signal rather than the payoff, we re-normalize the expression by dividing through by \( \beta \delta \) and denote this function by \( V \)

\[ V = -\frac{c(e)}{\beta \delta} + w. \tag{19} \]

It will convenient to denote \( v(e) = -\frac{c(e)}{\beta \delta} \).

Note that the agent's utility function satisfies the Spence-Mirrlees single crossing condition

\[ \frac{\partial}{\partial \beta} \left( \frac{\partial V}{\partial e} \right) = \frac{c'(e)}{\delta \beta^2} > 0. \tag{20} \]

This condition together with the fact that we have quasi-linear preferences ensures that any piecewise continuous decision function \( e(\cdot) \) satisfying \( de/d\beta \geq 0 \) is implementable, i.e. there exists a wage \( w(e) \) such that \((c(e), w(e))\) is incentive compatible.

Following Mirrlees [1971] the principal's optimization program can be written as

\[
\max_{e(\cdot)} \int_{E} \left[ \pi(e) + v(e) - \frac{1-F(\beta)}{f(\beta)} \frac{\partial v}{\partial \beta} \right] f(\beta) d\beta
\]

s.t. \( e(\cdot) \) is non-decreasing: \( \beta_2 > \beta_1 \Rightarrow e(\beta_2) \geq e(\beta_1) \).
The monotonicity constraint can be ignored so long as the distribution of types satisfies the monotone hazard rate condition

\[
\frac{d}{d\beta} \left( \frac{f(\beta)}{1 - F(\beta)} \right) \geq 0,
\]  

which holds for all unimodal distributions and which we will assume holds. The solution to the relaxed optimization program is then given by

\[
\frac{\partial \pi(e)}{\partial e} + \frac{\partial v(e)}{\partial e} = \frac{1 - F(\beta)}{f(\beta)} \frac{\partial^2 v(e)}{\partial e \partial \beta}, \quad \text{or}
\]

\[
\pi'(e) - \frac{c'(e)}{\beta \delta} = \frac{1 - F(\beta)}{f(\beta)} \frac{c'(e)}{\beta^2 \delta}
\]

which can be re-written as

\[
\pi'(e) = \frac{c'(e)}{\beta \delta} \left[ 1 + \frac{1 - F(\beta)}{f(\beta) \beta} \right].
\]

This expression is quite similar to that derived for naifs in the case of two discrete types in equation 13. Again we see no distortion at the top (as \(\beta\) approaches 1) with progressively greater distortion from first-best for lower values of \(\beta\).

With this result, we can now characterize the trade-offs involved in contract design when agents are neither fully naïve nor sophisticated but exhibit limited self awareness. An individual is partially self-aware (or partially naïve) if he recognizes he has a degree of present-bias but underestimates its magnitude: i.e. if \(\tilde{\beta}\) represents his belief then \(\beta < \tilde{\beta} < 1\). A natural way to model partial self-awareness in the population as a whole is to assume that the distribution of agents’ beliefs \(G(\beta)\) first-order stochastically dominates the actual distribution of types \(F(\beta)\). That is, for any \(\beta\) on the support of the distribution \(F(\beta)\) there is a larger share of the population

\[\text{This formulation follows O'Donoghue and Rabin (2001a).}\]
who believe they are less present-biased than this than the share in the population who actually are. This is a very weak assumption of naiveté that would be satisfied by a small share of the population simply underestimating the magnitude of their present-bias. In other words, this condition requires neither that all present-biased agents are even partially naïve, nor that any are fully naïve.

The implications of this for optimal contracting can be described quite intuitively. Just as in the case examined earlier with only two \( \beta \)-types when some of the present-biased agents are sophisticated, the presence of some self-awareness in the population gives the principal some ability to screen the agent population. Indeed, if the entire population were fully sophisticated then the principal would do best to offer a contract supporting effort only by an agent with \( \beta \) arbitrarily close to 1 which paid negligible rents to the agent. The sophistication of the agent population would ensure that all agents with significant present-bias would be screened from the contracting set. However, the presence of only limited self awareness in the population implies that such an attempt to screen all but the least present-biased agents would lead to a high probability of hiring an agent who shirks. Unlike the 2-type case, with a continuum of \( \beta \)-types in the population the principal does not face a binary choice whether to screen or not. Rather, she can offer a contract that supports effort by all agents above a critical point we will denote \( \hat{\beta} \). By doing so, she will screen out all agents who recognize that they are sufficiently present-biased that they would shirk when the time came to perform the contractual task, i.e. a share of the population represented by \( G(\hat{\beta}) \) (but note that this is not the same as the share of the population actually below \( \hat{\beta} \), which is \( F(\hat{\beta}) \)). By raising the lower bound of the contractual support the principal raises the effort she can expect from a contracting agent and reduces the rents that must be paid to induce higher effort levels from agents with little present-bias. Against these benefits
the principal of course trades-off the fact that naïveté in the population leads to shirking.

Specifically, those agents who believe they are less present-biased than \( \hat{\beta} \), and therefore accept the contract, but are incorrect, will shirk. The share of agents who will shirk is \( F(\hat{\beta}) - G(\hat{\beta}) \).

In this setting the principal chooses both how much of the agent population to screen by choosing \( \hat{\beta} \), and the optimal contract to maximize her payoff within the contracting set subject to the constraint that the contract cannot provide positive utility to any agent with \( \beta < \hat{\beta} \).\(^{13}\) Define \( H(\beta) \) on \( \beta \in [\hat{\beta},1] \) as the distribution of agents who will accept the contract and not shirk; i.e. it is the distribution of actual types \( F(\beta) \) conditional on \( \beta > \hat{\beta} \). Then we know from the solution to the contract design problem above (25) that the optimal contract for a given \( \hat{\beta} \) will be implicitly defined by

\[
\pi'(e) = \frac{c'(e)}{\beta \delta} \left[ 1 + \frac{1 - H(\beta)}{h(\beta) \beta} \right].
\]

The expected payoff to the principal if the agent has \( \beta > \hat{\beta} \) and therefore does not shirk is then

\[
\int_{\beta} \left[ \pi(e) - \frac{c(e)}{\beta \delta} - \frac{1 - H(\beta) c(e)}{h(\beta) \delta \beta^2} \right] h(\beta) d\beta.
\]

Note that this expression increases with \( \hat{\beta} \) because raising the lower bound on the degree of present-bias within the set of agents whose work effort is supported by the contract increase the expected effort from the agent and reduces the rents paid to an agent with little present-bias. If the agent has \( \beta < \hat{\beta} \) and agrees to the contract but then exerts no effort the principal receives a payoff of zero. Note that the size of the contracting set after eliminating those who are screened

\(^{13}\) Note that we assume here that the principal is able to commit not to renegotiate the contract at period 2. Without this commitment she would have an incentive to rewrite the contract in period 2 to support effort from all agents within the contracting set and the principal would fail to screen highly present-biased agents.
is $[1 - G(\hat{\beta})]$, so the share of the contracting set who shirk is $[F(\hat{\beta}) - G(\hat{\beta})]/[1 - G(\hat{\beta})]$ and the share of the contracting set who do not shirk is $[1 - F(\hat{\beta})]/[1 - G(\hat{\beta})]$. With this we can state the principal’s payoff as a function of her choice of how much to screen the agent population and the optimal contract given that choice:

$$\frac{1 - F(\hat{\beta})}{1 - G(\hat{\beta})} \int_{\hat{\beta}} \left[ \pi(e) - \frac{c(e)}{\beta \delta} - \frac{1 - H(\beta)}{h(\beta)} \frac{\delta \beta^2}{\beta^2} \right] h(\beta) d\beta.$$  \hspace{1cm} (28)

It is difficult to generalize about the optimal choice of $\hat{\beta}$ because without more restrictive assumptions on the distributions of agents’ types and beliefs the effect of increasing $\hat{\beta}$ on the share of the contracting pool who do not shirk may be very complex. However, we know that $[1 - F(\hat{\beta})]/[1 - G(\hat{\beta})]$ = 1 when $\hat{\beta} = \beta$ (i.e. no agents shirk when the contract does seek to screen anyone) and the fact that $G$ first-order stochastically dominates $F$ implies that $[1 - F(\hat{\beta})]/[1 - G(\hat{\beta})] < 1$ for at least some range of $\beta$. In order to ensure an interior solution we need a stronger but fairly intuitive assumption that $[1 - F(\hat{\beta})]/[1 - G(\hat{\beta})]$ is decreasing in $\hat{\beta}$, i.e. that the share of agents that shirk rises as the principal seeks to screen more of the population.

**PROPOSITION 5:** Assume $[1 - F(\hat{\beta})]/[1 - G(\hat{\beta})]$ is continuous, differentiable, decreasing in $\hat{\beta}$, and approaches zero as $\hat{\beta} \to 1$ and 1 as $\hat{\beta} \to \beta$. Then the principal’s optimal contract offer has an interior solution on $\hat{\beta}$. The principal will do best to offer a contract that screens some agents from the contracting set, but results in a positive probability of shirking by the agent who is hired.
This result indicates that the presence of present-biased agents who are not fully sophisticated regarding their future preferences leads to two sources of inefficiency in the second-best contract. Within the set of agents whose effort is supported by the contract, the principal trades-off suboptimal effort among more present-biased agents against rents to less present-biased agents, as discussed earlier. But here we also see that, because of the benefit to the principal of screening more present-biased agents from the contracting set, a positive probability exists that an agent will accept the contract but shirk when the time comes to perform the task.

5 Conclusion

When agents have present-biased preferences their effort is more costly than time-consistent agents if payment must be made after the effort takes place. Consequently, the first-best effort level for present-biased agents in such a circumstance is less than for otherwise identical agents with time-consistent preferences. However, the problem for the principal contracting with present-biased agents is not merely that inducing effort is more costly. When the principal cannot identify whether an agent's preferences are biased in favor of the present, she cannot offer a contract which will implement the first-best effort levels for all agents. Instead she must trade-off rents to agents with little or no present bias against inefficiently low effort by more present-biased agents. When agents are partially sophisticated the principal may do better to design a contract that screens out the most present-biased agents and mitigates the suboptimal effort and rents within the set of agents whose effort is supported by the contract, but at the cost of resulting in a positive probability that the agents will shirk.
Is it costly to an agent to be present-biased? Welfare comparisons for agents with present-biased preferences are inherently fraught with problems since the individual's preferences differ at different points in time, raising the question of how to weight the utility of the different period selves relative to each other. We have noted that in our simple 2-type model, when the principal chooses to contract with an agent rather than sell the project, if a TC agent is hired he typically earns rents while a present-biased agent will receive his reservation utility (the exception to this occurs if sophs are present in the population and the principal offers a screening contract which eliminates TC rents). However, this statement refers only to the welfare of the period 2 self. It is not clear that when analyzing the welfare effect on an agent of having present-biased preferences we should adopt the period 2 self's preferences, which give greater weight to the cost of effort relative to the later reward, than the preferences of any of the agent's other selves.

O'Donoghue and Rabin [1999a and 2001b] have argued that when making welfare comparisons one should do so from the perspective of the agent's prior or “long-run” perspective, reflecting the fact that present-biased preferences are seen as capturing a self-control problem which does not represent the individual's "true" preferences (this perspective is also taken less explicitly by Akerlof [1991]). The individual’s long-run utility is thus formally defined as

\[ U^0(u_1, u_2, \ldots, u_T) = \sum_{\tau=1}^{T} \delta^{T-\tau} u_{\tau} \]

We take this approach here so that the welfare of a present-biased individual is measured by his long-run utility.

Perhaps surprisingly, we find that when viewed from this perspective, agents with present-biased preferences receive positive utility in the contracting model (again ignoring the screening case) and in fact receive precisely the same utility as a TC agent. Evaluated at period 1, the long-run utility of a naïf from the contracting equilibrium is:
The utility of a TC agent is

\[ U_{TC} = -\delta c(e_N) + \frac{\delta^2 c(e_N)}{\beta \delta}, \]  

or

\[ U_{TC} = \delta c(e_N) \left( \frac{1}{\beta} - 1 \right). \]

This result comes from the fact that the wage/effort pairs for a naïf and a TC must lie on the same indifference curve for a time-consistent agent because the incentive compatibility constraint for a TC binds. In effect the presence of present-biased agents in the population gives both the present-biased and the time consistent agent bargaining power. By rendering it a credible threat that the agent will agree to a contract which just satisfies the time-consistent participation constraint but later choose not to complete the task, present-biased preferences prevent the principal from writing a contract that provides the present-biased agent with zero long-run utility. The fact that types are unobservable then requires that the contract pay rents to a time-consistent agent in order to induce him to truthfully reveal his type (i.e. exert effort \( e_{TC} \)).

The principal, when trading off the extra cost of inducing effort from a present-biased agent (which will be positive utility to the agent's long-run self) against rents to a time consistent agent, optimizes by choosing effort levels which equate the total long-run utility to each type.

---

14 Also note that this result holds for the case of a continuum of types as well. All \( \beta \)-types whose effort is supported by the contract will receive equal long-run time consistent utility because all wage/effort pairs lie on the same indifference curve.
Appendix A: The Principal’s Decision to Contract or Sell

Having solved the principal’s contract design problem, we can examine the question of whether the payoff from contracting with an agent is greater than what the principal could receive from simply selling the project. We first show that efficient ownership under full information requires that, since ownership must be determined in period 1 and the cost of purchasing the project is borne at that time whereas the profit is not realized until period 3, the project should be owned by the most patient party. Thus, if the principal’s discount parameter $\rho$ is greater than the time-consistent agents’ discount parameter $\delta$, the principal should retain ownership of the firm.

Consider first the principal’s payoff if she contracts with an agent under full information. In this case she will hire a TC agent and induce effort $e_{TC}$ by offering to pay $w = c(e_{TC})/\delta$ in period 3 if $\pi(e_{TC})$ is observed. Thus the principal’s payoff (evaluated at period 1) from contracting under full information is

$$V_{p}^{FI} = \rho \left( \pi(e_{TC}) - \frac{c(e_{TC})}{\delta} \right).$$  \hspace{1cm} (25)

If instead the principal sells the firm she will make a take-it-or-leave-it offer to sell at a price which just induces a TC agent (who will be willing to pay the most) to buy. A TC agent will be willing to pay up to the value to him of owning the project, which of course he will maximize by choosing effort $e_{TC}$. Therefore the value to a TC at period 1 of buying the project, and thus the price at which the principal can sell it, is

$$V_{TC} = \delta \pi(e_{TC}) - \delta e(e_{TC}).$$  \hspace{1cm} (26)

The difference between the payoff to the principal from contracting less the price at which she could sell the project is

$$V_{p}^{FI} - V_{TC} = \left( \rho \pi(e_{TC}) - \delta^2 \pi(e_{TC}) - \frac{c(e_{TC})}{\delta} \right).$$  \hspace{1cm} (27)

It follows directly from equation (27) that, with full information, the project will be owned by the more patient party. Efficiency requires that the principal retain ownership if $\rho > \delta$. 

26
As we have shown, when information about agents’ types is unobservable to the principal her optimal contract will provide an expected payoff lower than under full information because she may hire a present-biased agent for whom inducing effort is more costly and because the presence of such agents leads to a second-best contract design in which a TC agent will earn rents while a present biased agent will be induced to exert suboptimal effort. Selling the project avoids the inefficiency due to contracting under imperfect information; the project will always be sold to a TC agent who will maximize his payoff after purchasing the project by choosing the optimal effort level. But, if agents have a lower time-consistent discounting parameter than the principal ($\delta < \rho$), the transfer of ownership is itself inefficient. The efficiency cost of the transfer will be borne by the principal in the sense that she will receive a lower sale price than the payoff she could have received by retaining ownership if she could observe types and hire a TC agent. Consequently, when all agents’ time-consistent discounting is greater than the principal she must trade off the efficiency cost of contracting under imperfect information against the efficiency cost of selling the firm to a less-patient owner. In this circumstance, either the firm will be sold although the sale leads to inefficient ownership because doing so avoids the contracting problem, or an agent will be hired to work on the project even though hidden information about agents’ preferences for procrastination leads to second-best contracting.\textsuperscript{15} Whether principal does better to sell or contract depends on which inefficiency is greater: the cost of suboptimal contracting or non-optimal ownership.

\textsuperscript{15} The analysis in this section has characterized the trade-off between contracting and selling the project as one in which imperfect contracting is weighed against inefficient transfer of ownership to a less-patient agent. In a more general sense there may be many costly obstacles to the sale of a project from a principal to an agent. Ownership of a project may involve ownership of various pieces of capital which may be better utilized by being shared among many individuals undertaking different tasks, each of whom cannot individually own the same capital. The principal may have a lower cost of capital than the pool of agents. Or there may simply be substantial transactions costs associated with transferring ownership of the project.
Appendix B: Proofs

Proof of Proposition 1: Equation (8) shows that a TC agent will earn rents equal to
\[ \frac{(1 - \beta)c(e_N)}{\beta \delta} \]. We first must show that \( e_N \) (and thus \( c(e_N) \)) increases with \( \alpha \) and \( \beta \). Equation (13) which defines \( e_N \) can be re-written as
\[ \frac{\pi'(e_N)}{c'(e_N)} = \frac{1}{\beta \delta} \left[ 1 + \frac{(1 - \alpha)(1 - \beta)}{\alpha} \right], \]
which makes clear that \( e_N \) decreases with the right-hand-side of this expression. The partial derivative of the right-hand-side with respect to \( \alpha \) is
\[ \frac{\partial}{\partial \alpha} \left( \frac{(1 - \alpha)(1 - \beta)}{\alpha} \right) > 0, \]
so \( e_N \) increases with \( \alpha \).

The partial derivative of the right-hand-side with respect to \( \beta \) is:
\[ \frac{(\alpha - 1)}{\beta \delta \alpha} \left[ 1 + \frac{(1 - \alpha)(1 - \beta)}{\alpha} \right] \left( -\frac{1}{\beta^2 \delta} \right) < 0, \]
so \( e_N \) increases with \( \beta \).

Now we can sign the partial derivatives of the TCs rents.
\[ \frac{\partial}{\partial \alpha} \left[ \frac{(1 - \beta)c(e_N)}{\beta \delta} \right] = \frac{(1 - \beta)}{\beta \delta} c'(e_N) \frac{\partial e_N}{\partial \alpha} > 0, \]
which indicates that rents are increasing in \( \alpha \) since \( c(e_N) \) increases with \( \alpha \).
\[ \frac{\partial}{\partial \beta} \left[ \frac{(1 - \beta)c(e_N)}{\beta \delta} \right] = \frac{(1 - \beta)}{\beta \delta} c'(e_N) \frac{\partial e_N}{\partial \beta} - \frac{c(e_N)}{\beta^2 \delta}, \]
since both terms on the right-hand side are positive we cannot sign this expression for generic cost and profit functions. Because a higher \( \beta \) has the direct effect of decreasing rents but acts indirectly to increase rents by increasing \( e_N \), the overall effect of \( \beta \) on TCs rents is ambiguous.

Proof of Proposition 2: This follows directly from equations (12) and (13). Note that the second term in the summation in brackets in equation (13) is strictly positive.
Proof of Proposition 3: From equation (13) we have that the distortion from the first-best is indicated by the term \(\frac{1-\alpha}{\alpha}(1-\beta)\alpha > 0\). To prove the proposition we show that this term is decreasing in \(\alpha\) and \(\beta\).

\[
\frac{\partial}{\partial \alpha} \left[ (1-\alpha)(1-\beta) \right] = \frac{(\beta-1)}{\alpha^2} < 0.
\]

\[
\frac{\partial}{\partial \beta} \left[ (1-\alpha)(1-\beta) \right] = \frac{(\alpha-1)}{\alpha} \leq 0.
\]

Proof of Proposition 4: From equations (17) and (20) we can see that the principal's expected payoff from offering a contract that screens out sophs, \(\Pi_S\), minus her expected payoff from a non-screening contract, \(\Pi_N\), is given by

\[
\Pi_S - \Pi_N = \frac{(1-\alpha-\gamma)}{(1-\gamma)} \left[ \pi(e_{TC}) - \frac{c(e_{TC})}{\delta} \right] - (1-\alpha-\gamma) \left[ \pi(e_{TC}) - \frac{c(e_{TC})}{\delta} + \frac{(1-\beta)c(e_N)}{\beta\delta} \right] - (\alpha+\gamma) \left[ \pi(e_N) - \frac{c(e_N)}{\beta\delta} \right]
\]

which can be reduced to

\[
\frac{\gamma(1-\alpha-\gamma)}{(1-\gamma)} \left[ \pi(e_{TC}) - \frac{c(e_{TC})}{\delta} \right] - (\alpha+\gamma) \left[ \pi(e_N) - \frac{c(e_N)}{\delta} \right] + \frac{(1-\beta)c(e_N)}{\beta\delta}.
\]

When the expression is positive the principal prefers to screen out sophs. To prove the proposition we show that the limit of this expression as \(\alpha\) approaches 0 is strictly positive:

\[
\lim_{\alpha \to 0} (\Pi_S - \Pi_N) = \gamma \left[ \pi(e_{TC}) - \frac{c(e_{TC})}{\delta} \right] - \lim_{\alpha \to 0} \left( \pi(e_N) - \frac{c(e_N)}{\delta} \right) + \lim_{\alpha \to 0} \frac{(1-\beta)c(e_N)}{\beta\delta},
\]

which is the sum of two positive numbers (the bracketed term must be positive since, as noted before, \(e_{TC}\) maximizes the difference between profit and effort cost), so as the share of naifs in the population becomes very small the principal does best to offer a screening contract.
Proof of Proposition 5: The principal seeks to maximize her expected payoff represented by (28). As noted in the text, the integral in this expression is positive and increasing in $\hat{\beta}$. This must be true because the $H$ distribution by construction is supported on $\beta \in [\hat{\beta}, 1]$, so as the principal increases her choice of $\hat{\beta}$ she compresses the distribution toward higher $\beta$ values which increases the expected payoff represented by the integral. The assumptions on $[1 - F(\hat{\beta})] / [1 - G(\hat{\beta})]$ then ensure that the value of $\hat{\beta}$ that maximizes the principal’s payoff lies on $(\underline{\beta}, 1)$. 
References


