Contracting with a Naïve Time-Inconsistent Agent: To exploit or not to exploit?

Murat Yılmaz* †

June 18, 2015

Abstract

In a repeated principal-agent model with moral hazard, in which the agent has $\beta\delta$-preferences, we analyze the case where the agent is *naïve* in the sense that he is not fully aware of his inconsistent discounting. We consider the possibility of principal manipulating the naïveté of the agent. Surprisingly, when the principal wants to implement the high effort, there are no gains to the principal from the naïveté of the agent and the principal does not choose to deceive the agent. The principal's maximum utility is the same from a sophisticated agent and from a naïve agent.

Keywords : repeated moral hazard; time-inconsistency; $\beta\delta$-preferences; sophisticated agent; naïve agent. *JEL* **Classification Numbers:** D03, D82, D86.

*Boğaziçi University, Department of Economics, Bebek, Istanbul, 34342. Phone: +90 212 3597656. Fax: +90 212 2872453. Email: *muraty@boun.edu.tr.* Web site: *http://www.econ.boun.edu.tr/yilmaz*

†I am particularly indebted to Bart Lipman and Andy Newman for their guidance and support. I would like to thank Sambuddha Ghosh, Muhamet Yıldız, Alp Atakan and the seminar participants at Boston University, ASSET'11, All-Istanbul Workshop'11 and ESEM'12 for their useful comments. I would also like to thank the two anonymous referees for their useful comments and suggestions. I also acknowledge the financial support from Boğaziçi University, Scientific Research Projects [Bilimsel Araştırma Projeleri (BAP)], Project no 6354 (11D01P1).

1 Introduction

Recent literature shows that there is evidence for possible time-inconsistencies in agents' behavior.[1] It is often observed that agents tend to postpone actions that have immediate costs, but they do not like to postpone benefits. This is due to agents not putting much weight on the future benefits when the costs are immediate.[2] Examples include filing taxes, going to the gym, finishing an ongoing project. Time-inconsistency has been widely studied in the literature. There is a large amount of work dealing with time-inconsistent policies within strategic environments in the macroeconomics literature.[3] However, within microeconomic environments with time-inconsistency, most of the literature has focused on the individual decision making and there is a limited amount of work focusing on the strategic interaction. One such environment with strategic interaction is the one with a principal contracting with an agent where the agent has time-inconsistent preferences.

We focus on the standard repeated principal-agent problem where there is moral hazard, with the standard trade-off between incentives and insurance, and allow the agent to be time-inconsistent. This is considered in Yılmaz (2013), where the agent's time-inconsistency is modeled through widely used $\beta\delta$-preferences, the agent is assumed to be fully aware of his inconsistency, that is, he is *sophisticated*, and the optimal contract is characterized.[4] We provide a direct follow-up of Yılmaz (2013) by assuming that the agent is **not** fully aware of his time-inconsistency, that is, he is *naïve*.

When the agent is **naïve**, he mispredicts his inconsistency and overestimates his true β. If the principal is aware of the inconsistency of the agent, then she can possibly deceive

[1]See Elster (1979) and Thaler (1981).

[2]See Frederick, Loewenstein, ODonoghue (2002) and Loewenstein and Prelec (1992).

[3]See Kydland and Prescott (1977) for instance.

[4]$\beta\delta$-preferences are first developed by Phelps and Pollak (1968) and later used by Laibson (1997), O'Donoghue and Rabin (1999a,1999b,2001) among others. See Frederick, Loewenstein, O'Donoghue (2002), for an extensive overview of the literature. For details on the discounting scheme of a sophisticated agent with $\beta\delta$-preferences, please see Yılmaz (2013).

the agent to get information rents and exploit his naïveté. However, in a model of repeated moral hazard with a contracting stage followed by two stages with effort choices, we show that whenever the principal wants to implement the high effort, such an opportunity to manipulate the naïve agent does not provide the principal with information rents. And, the principal is indifferent between facing a naïve agent and facing a sophisticated agent, which is particularly surprising. This is because when the principal wants to implement the high effort, she may choose to deceive the agent by offering a contract in the contracting stage which makes the agent believe that he will pick low effort in the following stage, but when the agent arrives at the following stage, he realizes that he must choose the high effort, after learning his true β. However, deceiving the agent in this way is costly because such a contract must provide the agent with relatively low incentives for high effort at the contracting stage, through relatively higher expected continuation payoff for low output. However, arriving at the period where the agent actually picks an effort level, he learns his true β and finds it optimal to exert high effort. And for the latter to happen, the contract must offer strong incentives for high effort in the first period. Then, the discounted cost under such a contract becomes higher than the one under a contract that does not deceive the agent.

This paper is related to a number of other papers.[5] In the literature that is focusing on repeated principal-agent relationships, two very related papers are Rogerson (1985) and Lambert (1983), both of whom consider a repeated moral hazard problem with a time-consistent agent and show that the optimal contract exhibits memory.[6] Another important paper is Fudenberg, Holmstorm and Milgrom (1990) where an infinitely repeated principal-agent relationship is considered and the efficient long-term contract is shown to

[5]Here, we provide a brief review of the related literature. For a more detailed discussion of related papers, please see Yılmaz (2013).

[6]Also see Spear and Srivastava (1987) and Malcomson and Spinnewyn (1988).

be implemented by a sequence of short-term contracts.

Within the time-inconsistency literature, O'Donoghue and Rabin (1999b) introduce unobservable task-cost realizations into the moral hazard problem and assume that the agent is risk-neutral. Gilpatric (2008) assumes, in a principal-agent relationship with time-inconsistent agents, that the profit is fully determined by the effort. The current paper and Yılmaz (2013) are both distinguished from these papers since both allow the trade-off between risk and incentives.[7] DellaVigna and Malmendier (2004) consider a monopolistic firm who is designing an optimal two-part tariff where the consumer is time-inconsistent, and the principal knows whether the agent is aware of his inconsistency or not. Eliaz and Spiegler (2006) consider a monopoly who is contracting with dynamically inconsistent agents, and they characterize the optimal menu of contracts and show that it includes exploitative contracts for naïve agents. The current paper distinguishes itself from this paper through its modeling aspects and the result regarding exploiting possibilities. In a series of papers, O'Donoghue and Rabin (1999a, 2000, 2001) consider $\beta\delta$-preferences and focus on individual decision making rather than contractual relations.

Section 2 gives the specifics of the model and the timing of events. Section 3 studies the optimal contract when the principal implements high effort in each period. Section 4 discusses some extensions and section 5 concludes.

2 The Model

In a finitely repeated principal-agent model, a principal (she) is contracting with a time-inconsistent agent (he) to work on a two period project preceded by a contracting period. After the contracting period, in each of the two following periods the agent can exert a

[7] Also, see Balletta and Immordino (2013)

costly effort. The principal cannot observe the effort choices of the agent. The project, in each period, has an output which is publicly observed. The output in each period is stochastic and affected by the effort level picked by the agent in that period. A contract is a wage scheme $\{w_i, w_{ij}\}$ where w_i is the wage paid in the first period if the output is q_i in the first period, and w_{ij} is the wage paid in the second period if the first period output is q_i and the second period output is q_j.

Timing of events:

At $t = 0$, a contract is offered to the agent by the principal. Then, the agent accepts or rejects. If he rejects, the game ends and both the principal and the agent get zero utility. If he accepts, they move on to the next period. At $t = 1$, the agent chooses an effort level, e_1, which is not observed by the principal. The output, q_1, is realized which is observable by both the agent and the principal. The wage payment, w_{q_1}, is made to the agent. At $t = 2$, the agent chooses an effort level, e_2, which is not observed by the principal. The output, q_2, is realized which is observable by both the agent and the principal. The wage payment, w_{q_1,q_2}, is made to the agent.

Note that, in the contracting stage, $t = 0$, there is no effort decision. The only decision made by the agent is to accept or reject the contract offered by the principal. This can be motivated by thinking of getting a job offer in March but starting in September, just as most economics Ph.D. candidates experience. Thus, there are three periods with two effort decisions. Once the contract is accepted at $t = 0$, both agent and principal are committed to the contract until the end of period 2, that is, focus is on long-term contracts and renegotiation is not allowed.

Agent:

There are two effort levels, 0 and 1, and two outcomes, q_h and q_l with $q_h > q_l$. The agent receives utility $u(w)$ from the wage w, with $u' > 0$, and disutility $\psi(e)$ from exerting effort e. He is risk-averse, that is, $u'' < 0$. The disutility from exerting effort is given by $\psi(1) = \psi$ and $\psi(0) = 0$, where $\psi > 0$. His net utility in period $t \in \{1, 2\}$ is given by $v_t = u(w_t) - \psi(e_t)$. The agent's present value of a flow of future utilities as of period t, $PV(t)$, is

$$PV(t) = \begin{cases} v_t + \beta \sum_{s=t+1}^{2} \delta^{s-t} v_s & \text{for} \quad t = 0, 1 \\ v_2 & \text{for} \quad t = 2 \end{cases}$$

The agent is *time-inconsistent* when $\beta < 1$. A time-inconsistent agent can be fully aware, partially (or fully) unaware of his time inconsistency. Denoting the agent's belief about his true β by $\widehat{\beta}$, a time-inconsistent agent is *sophisticated* when $\widehat{\beta} = \beta < 1$, and *naïve* when $\beta < \widehat{\beta} \leq 1$. In what follows, the agent is **naïve**.

Denote the probability of getting *high output* under *high effort,* $e = 1$ by $\Pr(q = q_h | e = 1) = p_1$ and the probability of getting *high output* under *low effort,* $e = 0$ by $\Pr(q = q_h | e = 0) = p_0$, where $p_1 > p_0$.

There is no lending or borrowing. So, the agent spends whatever wage he earns in a period within that period. Also, there is no limited liability.

Principal:

The principal is risk-neutral, time-consistent and has a discount factor, δ_P. She cannot observe the effort levels exerted by the agent but she knows that the agent is time-inconsistent and naïve.[8] The principal wants to implement high effort in each period.[9] The principal minimizes the expected cost of implementing high effort in both periods, subject to individual rationality (IR) and incentive compatibility (IC), over all possible contracts, $\{w_i, w_{ij}\}_{i,j \in \{h,l\}}$, where w_i is the wage paid in the first period if the output is q_i in the first

[8] This assumption is also present in, for instance, DellaVigna and Malmendier (2004) and O'Donoghue and Rabin (1999b). For a justification see Yılmaz (2013).

[9] See Lemma 2 in Appendix A of Yılmaz (2013) for a condition that guarantees this.

period, and w_{ij} is the wage paid in the second period if the first period output is q_i and the second period output is q_j. Her payoff from a contract, $\{(w_h, w_l), (w_{hh}, w_{hl}, w_{lh}, w_{ll})\}$, given that it is accepted and the agent exerts high effort in both periods is given by

$$p_1[q_h - w_h + \delta_P[p_1(q_h - w_{hh}) + (1 - p_1)(q_l - w_{hl})]]$$
$$+(1 - p_1)[q_l - w_l + \delta_P[p_1(q_h - w_{lh}) + (1 - p_1)(q_l - w_{ll})]]$$

3 Optimal Contract

When the agent is naïve, his anticipation of his future self's behavior is not correct.[10] At $t = 0$, he believes that he has a discounting given by $(1, \widehat{\beta}\delta, \widehat{\beta}\delta^2)$ where $\beta < \widehat{\beta} \leq 1$, so he underestimates his inconsistency (equivalently, overestimates his β). In the contracting stage, his decision to accept or reject the contract depends on the set of wages offered and on the actions he thinks his future selves will take. Hence, IR should be based on his (incorrect) anticipation of his future actions. However, IC should consider only the agent's actual behavior at the effort stage. This is because the principal knows agent's exact discounting scheme and, from $t = 1$ on, the agent behaves according to his true β, *after learning his true* β.[11] More precisely, the principal will pick a set of wages which makes the agent believe that his future selves will pick a certain effort scheme, the *artificial* one, the effort levels that the agent believes at $t = 0$, his future selves (the self at $t = 1$ and the self at $t = 2$) will pick at $t = 1$ and $t = 2$, denoted by $\{a_1, (a_2^i)_{i\in\{h,l\}}\}$. The same wage scheme will actually implement the *effective* effort scheme, denoted $\{e_1, (e_2^i)_{i\in\{h,l\}}\}$,

[10] We consider a naïve agent who, looking at the contract, does not infer anything about his future self's type. Instead he sticks to his prior belief about the future self. For a discussion on the possibility of the agent learning about his true β by looking at the contract at the contracting stage, please see the discussion section.

[11] The agent learns his true β at $t = 1$. If the agent never learns his true β, then he'd always act according to $\widehat{\beta}$, his overestimated β. Thus, he would be equivalent to a sophisticated agent with δ and $\widehat{\beta}$.

where $e_1 = 1$ and $e_2^i = 1$ for $i \in \{h, l\}$.

The principal's expected profit (expected cost) is based on the set of wages and on the effective effort schedule, not on the artificial one. The agent uses the discounting $(1, \widehat{\beta}\delta, \widehat{\beta}\delta^2)$ at the contracting stage expecting that his discounting from $t = 1$ on will be $(1, \widehat{\beta}\delta)$; however, he will end up behaving according to $(1, \beta\delta)$ from $t = 1$ on, upon learning his true β. Since at $t = 1$, the agent learns his true β, there is no other artificial effort level to consider. Since there is no further time-inconsistency at $t = 2$, the effective second period effort will be the same as the artificial second period effort, from $t = 0$ point of view. That is, $a_2^i = e_2^i = 1$ for $i \in \{h, l\}$. Therefore, the only artificial effort level that matters is $a_1 \in \{0, 1\}$.

Before starting to solve this problem, we solve for the second period payments first. Denote the utilities of the agent, in the second period, from wages with $u(w_{q_i}) = u_i$, and $u(w_{q_i q_j}) = u_{ij}$ where $i, j \in \{h, l\}$, and denote the inverse of the utility function by $h(u)$, assuming it's continuously differentiable. For a given first period output level $q_i \in \{q_h, q_l\}$, the agent's continuation payoff will be $p_1 u_{ih} + (1 - p_1) u_{il} - \psi$. Once the principal promises the agent a utility of Eu_i, the continuation of the optimal contract for the second period will be given by the solution to the following problem

$$\min_{u_{ih}, u_{il}} p_1 h(u_{ih}) + (1 - p_1) h(u_{il})$$

subject to

$$\begin{aligned} u_{ih} - u_{il} &\geq \frac{\psi}{p_1 - p_0} \\ p_1 u_{ih} + (1 - p_1) u_{il} - \psi &\geq Eu_i \end{aligned}$$

As in Yılmaz (2013), this is a static problem and both constraints bind. Thus, for a given first period output, q_i, the second period payoffs to the agent are

$$\begin{aligned} u_{ih} &= \psi + Eu_i + (1 - p_1)\frac{\psi}{p_1 - p_0} \\ u_{il} &= \psi + Eu_i - p_1\frac{\psi}{p_1 - p_0} \end{aligned}$$

If the agent has been promised Eu_i for the second period when the first period output realization is q_i, then u_{ih} and u_{il} are defined to be $p_1 u_{ih} + (1 - p_1)u_{il} - \psi = Eu_i$ for $i \in \{h, l\}$. Also, denote the cost of implementing the high effort level in the second period, given that the promised second period utility is Eu_i, by $C_2(Eu_i)$, that is,

$$C_2(Eu_i) = p_1 h(u_{ih}) + (1 - p_1)h(u_{il})$$

Now, we turn to the possibility of manipulating the agent's naïveté through the first period effort level. The principal has an opportunity to manipulate the agent's naïveté by making him believe that his future self at $t = 1$ exerts no effort, but when he actually arrives at $t = 1$ he exerts high effort as the principal desires. That is, the principal may set a contract such that $a_1 = 0$ and $e_1 = 1$. The principal's problem in the case is given by

$$\min_{\{u_i, Eu_i\}_{i \in \{h,l\}}} p_1[h(u_h) + \delta_P C_2(Eu_h)] + (1 - p_1)[h(u_l) + \delta_P C_2(Eu_l)]$$

subject to

$$\begin{aligned} &(IR^0) \quad p_0[u_h + \delta Eu_h] + (1 - p_0)[u_l + \delta Eu_l] \geq 0 \\ &(IC_a^0) \qquad u_h - u_l + \delta\widehat{\beta}(Eu_h - Eu_l) \leq \tfrac{\psi}{p_1 - p_0} \\ &(IC_e^0) \qquad u_h - u_l + \delta\beta(Eu_h - Eu_l) \geq \tfrac{\psi}{p_1 - p_0} \end{aligned}$$

Call this problem $\mathbf{P_0}$, where IR^0 ensures that the agent accepts the contract given that he believes that his self at $t = 1$ will pick low effort. IC_a^0 convinces the agent, from $t = 0$ perspective, that he will choose low effort at $t = 1$. Finally, IC_e^0 ensures that when the agent actually arrives at $t = 1$ and learns his true β, he changes his mind and picks high effort.

Alternatively, the principal can simply choose to implement high effort in both periods, without trying to make the agent believe that his future self will pick low effort at $t = 1$. That is, $a_1 = 1$. In this case, the principal's problem is

$$\min_{\{u_i, Eu_i\}_{i\in\{h,l\}}} p_1[h(u_h) + \delta_P C_2(Eu_h)] + (1 - p_1)[h(u_l) + \delta_P C_2(Eu_l)]$$

subject to

$$\begin{aligned} &(IR^1) \quad p_1[u_h + \delta Eu_h] + (1 - p_1)[u_l + \delta Eu_l] \geq \psi \\ &(IC_a^1) \qquad u_h - u_l + \delta\widehat{\beta}(Eu_h - Eu_l) \geq \tfrac{\psi}{p_1 - p_0} \\ &(IC_e^1) \qquad u_h - u_l + \delta\beta(Eu_h - Eu_l) \geq \tfrac{\psi}{p_1 - p_0} \end{aligned}$$

Call this problem $\mathbf{P_1}$. The two lemmas below show that the constraints with $\widehat{\beta}$, IC_a^1 and IC_a^0, do not bind.

Lemma 1 *In $\mathbf{P_1}$, IC_e^1 binds and IC_a^1 does not bind.*

Proof. Attaching multipliers λ to the IR^1, μ_a to IC_a^1 and μ_e to IC_e^1, we get $\frac{\delta_P}{\delta}C'(Eu_h) = \lambda + \frac{\widehat{\beta}\mu_a + \beta\mu_e}{p_1}$ and $\frac{\delta_P}{\delta}C'(Eu_l) = \lambda - \frac{\widehat{\beta}\mu_a + \beta\mu_e}{p_1}$. Note that $\mu_a \geq 0$ and $\mu_e \geq 0$ and C' is strictly increasing.[12] Thus, $Eu_h \geq Eu_l$. Suppose $Eu_h = Eu_l$. Then, it must be $\mu_a = \mu_e = 0$. But, we also have $h'(u_h) = \lambda + \frac{\mu_a + \mu_e}{p_1}$ and $h'(u_l) = \lambda - \frac{\mu_a + \mu_e}{1 - p_1}$. Thus, we get $u_h = u_l$. Then, the left hand side of IC_e^1 is zero, but the right hand side of it is strictly positive, which is a contradiction. Thus, $Eu_h > Eu_l$. Now, both IC_e^1 and IC_a^1 cannot be binding at the same time, since $\widehat{\beta} > \beta$. We also just showed that both cannot be non-binding. Thus,

[12] C' being strictly increasing is established in the proof of Proposition 3 in Yılmaz (2013).

exactly one of the two must bind. Since $Eu_h > Eu_l$, we conclude that IC_e^1 binds and IC_a^1 does not bind. ■

Lemma 2 *In* $\mathbf{P_0}$*, if* $Eu_h \neq Eu_l$*, then* IC_e^0 *binds and* IC_a^0 *does not bind.*

Proof. IC_a^0 and IC_e^0 together imply $\delta\widehat{\beta}(Eu_h - Eu_l) \leq \delta\beta(Eu_h - Eu_l)$. Hence $Eu_h \leq Eu_l$. By our assumption $Eu_h < Eu_l$. Then, both IC_e^0 and IC_a^0 cannot be binding since $\widehat{\beta} > \beta$. Both constraints cannot be non-binding as well. If they are both non-binding then $\mu_a = \mu_e = 0$. But then $h'(u_h) - h'(u_l) = \lambda(\frac{p_0}{p_1} - \frac{1-p_0}{1-p_1}) = \lambda(\frac{p_0-p_1}{p_1(1-p_1)}) < 0$ since $p_0 < p_1$. Hence $u_h - u_l < 0$. But then the left hand side of IC_e^0 is negative, which is a contradiction. Therefore, exactly one of the IC_a^0 and IC_e^0 should be binding. Suppose IC_a^0 binds and IC_e^0 does not. Then $\mu_a > 0$ and $\mu_e = 0$. But then we get $h'(u_h) - h'(u_l) = \lambda(\frac{p_0-p_1}{p_1(1-p_1)}) - \mu_a(\frac{1}{p_1(1-p_1)}) < 0$ since $p_0 < p_1$ and $\mu_a > 0$. Again $u_h - u_l < 0$, which again gives a contradiction. Thus, if $Eu_h \neq Eu_l$, then IC_e^0 binds and IC_a^0 does not bind. ■

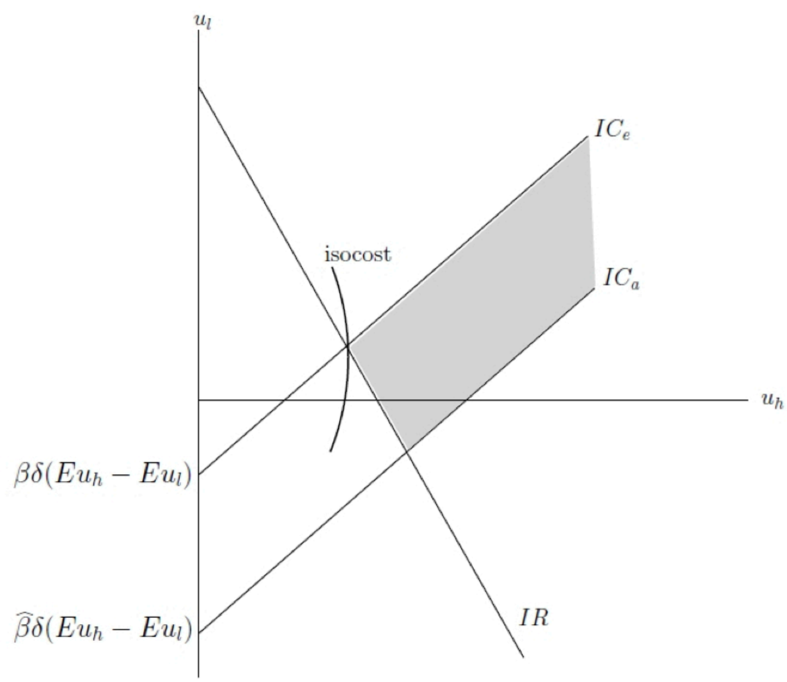

Figure 1

The intuition behind Lemma 2 can be seen in Figure 1 above. When $Eu_h < Eu_l$, the shaded region gives the set of contracts that are feasible. Among those the one at the intersection of IC_e^0 and IR^0 is the one that has the minimum cost, since the isocosts are steeper than the IR^0.

Proposition 1 *If the principal implements high effort in both periods,* $e_1 = 1$ *and* $e_2^h = e_2^l = 1$*, she is indifferent between facing a sophisticated time-inconsistent agent and facing a naïve time-inconsistent agent. There are no gains from exploiting the agent's naïveté.*

Proof. When principal wants to implement $e_1 = 1$ and $e_2^h = e_2^l = 1$, showing that $a_1 = 1$ is optimal will prove this result because in this case, the contract for sophisticated agent and the contract for naïve agent with $a_1 = 1$ are going to be exactly the same. This is due to the principal's problem when facing a sophisticated agent is same as $\mathbf{P_1}$ by Lemma 1. Now, to see $a_1 = 1$ is optimal, first assume that in $\mathbf{P_0}$, $Eu_h \neq Eu_l$. Then, by Lemma 1 and Lemma 2 above, IC_a does not bind for either problem $\mathbf{P_0}$ and $\mathbf{P_1}$. Also note that both IR and IC_e bind regardless of a_1, in each problem $\mathbf{P_0}$ and $\mathbf{P_1}$. Now look at the following problem

$$\min_{\{u_i, Eu_i\}_{i\in\{h,l\}}} p_1[h(u_h) + \delta_P C_2(Eu_h)] + (1-p_1)[h(u_l) + \delta_P C_2(Eu_l)]$$

subject to

$$(IR) \quad (p_1 - \eta)[u_h + \delta Eu_h] + (1 - p_1 + \eta)[u_l + \delta Eu_l] \geq \psi(1 - \tfrac{\eta}{p_1 - p_0})$$
$$(IC_e) \quad u_h - u_l + \delta\beta(Eu_h - Eu_l) - \tfrac{\psi}{p_1 - p_0} = 0$$

If $\eta = 0$, then the problem above is exactly the problem $\mathbf{P_1}$. If $\eta = p_1 - p_0$, then it is exactly the problem $\mathbf{P_0}$. Then we have, $\frac{dEC(\eta)}{d\eta} = \lambda(u_h - u_l + \delta(Eu_h - Eu_l) - \frac{\psi}{p_1 - p_0})$, where λ is the Lagrange multiplier attached to the IR, and $EC(\eta)$ is the expected cost in the relevant problem. By IC_e in $\mathbf{P_1}$, we have $u_h - u_l + \delta\beta(Eu_h - Eu_l) - \frac{\psi}{p_1 - p_0} = 0$. Then, using the fact that $Eu_h - Eu_l > 0$ in $\mathbf{P_1}$, the derivative $\frac{dEC(\eta)}{d\eta}$ evaluated at $\eta = 0$ gives us a positive value. Since $p_1 - p_0 > 0$, and $EC(\eta)$ is increasing at $\eta = 0$, the cost will be higher at $\eta = p_1 - p_0$ than the one at $\eta = 0$; that is, the minimized cost in $\mathbf{P_0}$ is higher than the one in $\mathbf{P_1}$.

Now, assume that $Eu_h = Eu_l = Eu$ in $\mathbf{P_0}$. Then, $\mathbf{P_0}$ boils down to the following problem

$$\min_{u_h, u_l, Eu} p_1 h(u_h) + (1 - p_1) h(u_l) + \delta_P C_2(Eu)$$

subject to

$$p_0 u_h + (1 - p_0) u_l + \delta Eu \geq 0$$
$$u_h - u_l = \tfrac{\psi}{p_1 - p_0}.$$

Now, in $\mathbf{P_1}$, if we restrict the feasible set to $Eu_h = Eu_l$, then the minimized cost in this modified problem, call it $\mathbf{P_1'}$, is at least as big as the minimized cost in the original $\mathbf{P_1}$. Then, the constraint set of the problem $\mathbf{P_1'}$ is given by the constraints $p_1 u_h + (1 - p_1) u_l + \delta Eu \geq \psi$ and $u_h - u_l = \frac{\psi}{p_1 - p_0}$. However, these two constraints are equivalent to the ones in the problem $\mathbf{P_0}$ with $Eu_h = Eu_l = Eu$. To see this, plug $u_h - u_l = \frac{\psi}{p_1 - p_0}$ into $p_1 u_h + (1 - p_1) u_l + \delta Eu \geq \psi$, that is, into $p_1(u_h - u_l) + u_l + \delta Eu \geq \psi$ and get $p_0 u_h + (1 - p_0) u_l + \delta Eu \geq 0$, which is exactly the same condition in the reduced $\mathbf{P_0}$.

Thus, we have shown that in either case the minimized cost in $\mathbf{P_0}$ is at least as big as the minimized cost in $\mathbf{P_1}$. Thus, $a_1 = 1$ is optimal. With $a_1 = 1$, the principal's problem is the same as the problem for the sophisticated agent. And, there are no gains from making the agent believe that he will pick low effort level. ■

The intuition for this result is that when the contract makes the agent believe that he will pick low effort in the first period, it provides relatively low incentives for high effort at the contracting stage, through relatively higher expected continuation payoff for low output, that is, $Eu_h \leq Eu_l$. But when the agent arrives at the first period he learns his true discounting and finds it optimal to exert high effort, for which the first period incentives must be much stronger through high u_h and low u_l. Thus, for the principal the discounted cost is higher than the one where she does not deceive the agent. Thus, she is worse off with the contract where she deceives the agent. This is particularly striking because with a naive time-inconsistent agent the principal has the power to manipulate the agent through his misperception. However, such an opportunity to manipulate does not provide the principal with higher profits. In fact, the principal chooses not to deceive the agent at all.

We have just seen, through the proof of Proposition 1, that the principal offers the same contract to both naïve and sophisticated agents, and both are accepted, implementing high effort in each period. Also, note that for both naïve and sophisticated agents, the individual rationality constraints bind, where β does not enter.[13] Thus, from the contracting stage point of view ($t = 0$), both naïve and sophisticated agents expect the same overall payoff from the contract offered, which is in fact zero. In the first period ($t = 0$), naïve agent learns his true beta, thus the expected payoff from the first period point of view are also the same for both type of agents, since the contracts are the same. From the second period point of view ($t = 2$), both agents again expect the same payoff, since there is no further inconsistency and the payments are the same. Thus, in terms of agent's payoff, naïveté does not make the agent worse off relative to the sophisticated agent.

[13]The outside option is assumed to be zero. Thus, β cancels out, since there is no payoff realized in the contracting stage. For a discussion on this assumption see Yılmaz (2013).

4 Discussion

Learning from the contract:

If the naive agent can learn from the contract about his true β, then the principal can either offer a contract that reveals no information or a contract that reveals all the information about the agent's true β.[14] However, we already showed that when the naive agent is not capable of learning from the contract, the principal is indifferent between a sophisticated agent (who knows his true β) and a naive agent (who mispredicts his true β) in terms of the cost of implementing high effort in each period. In the light of this result, the principal is not going to be better off with revealing no information rather than revealing all the information. To see this, suppose the optimal contract that the principal offers to a sophisticated agent is $\{w_i^{SO}, w_{ij}^{SO}\}_{i,j\in\{h,l\}}$. If the principal offers this contract to the naive agent, the naive agent will learn his true β and hence will behave as a sophisticated agent. Thus, this contract, when offered to a naive agent, will implement high effort in each period. If, however, the principal offers another contract to the naive agent, say $\{w_i^{NA}, w_{ij}^{NA}\}_{i,j\in\{h,l\}}$, that reveals no information then the agent will keep behaving according to his prior belief, $\widehat{\beta}$. But then, this contract will have a higher (or at least as high as) cost of implementing high effort in each period. This is because, the contract $\{w_i^{SO}, w_{ij}^{SO}\}_{i,j\in\{h,l\}}$ is the one with the lowest cost of implementing the high effort in each period when the agent does not learn, by Proposition 1. Thus, the contract $\{w_i^{NA}, w_{ij}^{NA}\}_{i,j\in\{h,l\}}$ will make the principal weakly worse off compared to the case where the principal reveals all the information by offering $\{w_i^{SO}, w_{ij}^{SO}\}_{i,j\in\{h,l\}}$. Thus, when the naive agent can learn from the contract, the optimal contract will be $\{w_i^{SO}, w_{ij}^{SO}\}_{i,j\in\{h,l\}}$, the agent will learn his true β and behave sophisticatedly, and high

[14] Here, we focus on revealing all or no information options, because revealing some (but not all) information just changes the level of misperception of the agent and does not make him sophisticated. And Lemma 1 and Lemma 2 tell us that the level of misperception is not important.

effort will be implemented in each period.

Implementing the low effort level:

I assumed that the principal wants to implement high effort level in each period. The possibility of exploiting the agent's naïveté when the principal wants to implement low effort in one or both of the periods is worth to discuss here. To keep it simple, we consider the case where the principal wants to implement low effort in the first period and high effort in the second period. Since, period $t = 2$ is the last period and there is no further discounting, hence no more time-inconsistency, it is plausible to keep high effort in the last period.

The principal can offer a contract that makes the agent believe that he will pick high effort in the first period, from the perspective of the contracting stage. But when the first period arrives the agent, after learning his true β, realizes that he must pick low effort. Thus, the artificial and the effective effort levels are $\{a_1 = 1, a_2^i = 1\}$ and $\{e_1 = 0, e_2^i = 1\}$ for each $i \in \{h, l\}$, respectively. The principal's optimization problem in this case is

$$\min_{\{u_i, Eu_i\}_{i\in\{h,l\}}} p_0[h(u_h) + \delta_P C_2(Eu_h)] + (1 - p_0)[h(u_l) + \delta_P C_2(Eu_l)]$$

subject to

$$(IR^0) \quad p_1[u_h + \delta Eu_h] + (1 - p_1)[u_l + \delta Eu_l] \geq \psi$$

$$(IC_a^0) \qquad u_h - u_l + \delta\widehat{\beta}(Eu_h - Eu_l) \geq \frac{\psi}{p_1 - p_0}$$

$$(IC_e^0) \qquad u_h - u_l + \delta\beta(Eu_h - Eu_l) \leq \frac{\psi}{p_1 - p_0}$$

On the other hand, if the principal tries not to manipulate the agent and just simply implements low effort in the first period without making him believe, from $t = 0$ perspective, that he will pick high effort in the first period, then the principal's problem is

$$\min_{\{u_i, Eu_i\}_{i\in\{h,l\}}} p_0[h(u_h) + \delta_P C_2(Eu_h)] + (1-p_0)[h(u_l) + \delta_P C_2(Eu_l)]$$

subject to

$$\begin{aligned}
(IR^0) &\quad p_0[u_h + \delta Eu_h] + (1-p_0)[u_l + \delta Eu_l] \geq 0 \\
(IC_a^0) &\qquad u_h - u_l + \delta\widehat{\beta}(Eu_h - Eu_l) \leq \tfrac{\psi}{p_1 - p_0} \\
(IC_e^0) &\qquad u_h - u_l + \delta\beta(Eu_h - Eu_l) \leq \tfrac{\psi}{p_1 - p_0}
\end{aligned}$$

Allowing only for non-negative wages, the cost minimizing contract for the latter problem is clearly $u_h = u_l = Eu_h = Eu_l = 0$. Thus, the minimized cost is 0. However, in the former problem the contract $u_h = u_l = Eu_h = Eu_l = 0$ is not feasible since IR^0 is not satisfied. Thus, the minimized cost in the former problem is strictly positive. Thus, the principal does not gain from manipulating the agent's naïveté, if the wages are non-negative. This is because, to make the agent believe that he will pick high effort in the first period, there must be some incentives provided for high effort through strictly positive wages for high output.

More than two effort levels:

If there are more than two effort levels and if the principal wants to implement an intermediate level of effort, then there are more possibilities in terms of deceiving the agent. The contract could make the agent believe that his future self will pick a higher effort level, but when time arrives, he actually chooses the one principal wants. This type of contract may make the principal better off than the contract which does not deceive the agent at all.

Endogenous time-inconsistency:

In our model, through given β and δ, time-inconsistency is assumed to be exogenous, that is, the agent cannot choose to be inconsistent or not. However, the agent may choose to be time-inconsistent or not. One can think of time-inconsistency as the change in the optimal plan over time, more precisely, a plan for the future may be optimal now, but when the future arrives, it may no longer be optimal, even though no new information is received. Within the current context, one plausible modelling choice is to have two stages, first a contracting stage (no effort choice) and then the effort stage (where effort is actually picked and exerted), and to make it possible for the agent to find (looking at the contract) high effort optimal from the contracting stage point of view, and then possibly find low effort optimal in the effort stage. Moreover, in order for the time-inconsistency to be *endogenous*, one may need to assume that there exists a commitment device, which the agent can use to make sure that he does not pick low effort in the second stage (when he faces a contract that tells him that high effort is optimal, from the contracting stage point of view). The agent can actually avoid time-inconsistency, if he likes, by commitment, and time-inconsistency would be endogenous. To incorporate the distinction between sophisticated agent and naïve agent, we need to assume that the naïve agent is not aware of the possible change in the optimal plan, when he is at the contracting stage. Thus, he will not feel any urge for commitment.

The main difficulty that can arise within this specification is the calculation of optimal contracts for both the sophisticated and naïve agents. When facing a sophisticated agent, the principal must take into account that there is a commitment opportunity for the agent, and whether to push the agent towards commitment or not, besides giving incentives for high effort. When facing a naïve agent, the principal must take potential time-inconsistency into account, that is, the possibility that the low effort may be optimal for the agent in the second stage, knowing that the agent has no commitment device.

Intuitively, when the principal faces a naïve agent, since there is time-inconsistency, the principal must provide relatively more incentives for the agent to pick high effort when the effort stage comes. The agent is likely to find low effort optimal in the effort stage, even though he may have found high effort optimal at the contracting stage. Thus, it will be relatively more costly to implement high effort. When the principal faces a sophisticated agent, however, there will potentially be a commitment by the agent and the principal may find it optimal to offer a contract which induces the agent to choose commitment. This may also be costly since this contract also needs to give incentives for high effort and there may be cost of commitment. However, the optimal contract for the naïve agent already implements high effort for the sophisticated agent in case there is no commitment. Thus, it must be weakly cheaper for the principal to implement high effort with the sophisticated agent than it is with the naïve agent, if the commitment is not too costly. Based on this intuition, my main result that says the principal is indifferent between facing a sophisticated agent and a naïve agent would be altered in a way that the principal would prefer to face a sophisticated agent rather than a naïve agent, potentially for low levels of cost of commitment, if any. This would not be surprising since time-inconsistency, which is bad for the principal, is endogenous, and sophisticated agent is capable of avoiding the time-inconsistency. My main result also points out that the principal is optimally *not* exploiting the naïve agent, and under the specification above this dimension of my result is still robust.

5 Conclusion

In a repeated moral hazard problem with a time-inconsistent agent, where the agent's time-inconsistency is captured by $\beta\delta$*-preferences*, the possibility to exploit the agent's misperception is considered. Unlike Yılmaz (2013), where the time-inconsistent agent is sophisticated, the time-inconsistent agent is naïve. With a naïve agent the principal's problem is more involved relative to the case where he is sophisticated, because she can deceive the agent and potentially get information rents. Principal may make the agent believe that he will pick the effort level other than the one she wants to implement, but when the time comes the agent actually picks the effort level the principal wants to implement. However, when the agent is naïve, the principal, implementing high effort in both periods, does not gain from exploiting the agent's naïveté. This follows from the result that the possibility of deceiving the agent through payments in the contract is costly. And, she is indifferent between facing a naïve agent and facing a sophisticated agent.

REFERENCES

1. Balletta, L. & Immordino, G. (2013). "On repeated moral hazard with a present biased agent", CSEF Working Paper n.341.

2. Dellavigna, S., & Malmendier, U. (2004). "Contract Design and Self-Control: Theory and Evidence", *Quarterly Journal of Economics*, Vol.119, pp. 353-402.

3. Elster, J. (1979). "Ulysses and the Sirens: Studies in Rationality and Irrationality", *Cambridge: Cambridge University Press*, Vol.1.

4. Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). "Time Discounting and Time Preference: A Critical View", *Journal of Economic Literature*, Vol. 40, pp. 351-401

5. Fudenberg, D., Holmstrom, B., & Milgrom, P. (1990). "Short-Term Contracts and Long-Term Agency Relationships", *Journal of Economic Theory*, Vol. 51, No.1.

6. Gilpatric, S. (2008). "Present-Biased Preferences, Self-Awareness and Shirking", *Journal of Economic Behavior and Organization*, Vol. 67, 3-4, pp. 735-754.

7. Kydland, F. E., & Prescott, E.C. (1977), "Rules Rather than Discretion: The Inconsistency of Optimal Plans", *Journal of Political Economy*, Vol.85, pp. 473-492.

8. Laibson, D. (1997). "Golden Eggs and Hyperbolic Discounting", *The Quarterly Journal of Economics*, Vol. 112, No.2., pp. 443-477.

9. Lambert, R. A. (1983). "Long-Term Contracts and Moral Hazard", *The Bell Journal of Economics*, Vol. 14, No.2., pp. 441-452.

10. Malcomson, J. M., & Spinnewyn, F. (1988). "The Multiperiod Principal-Agent Problem", *The Review of Economic Studies*, Vol.55, No.3., pp. 391-407.

11. O'Donoghue, T., & Rabin, M. (2001). "Choice and Procrastination", *The Quarterly Journal of Economics*, Vol. 116, No.1., pp. 121-160.

12. O'Donoghue, T., & Rabin, M. (1999a). "Doing It Now or Later", *American Economic Review*, Vol. 89, No.1., pp. 103-124.

13. O'Donoghue, T., & Rabin, M. (1999b). "Incentives for Procrastinators", *The Quarterly Journal of Economics*, Vol. 114, No.3., pp. 769-816.

14. Phelps, E. S., & Pollak, R. A. (1968). "On Second-Best National Saving and Game-Equilibrium Growth", *The Review of Economic Studies*, Vol.35, No.2., pp. 185-199.

15. Rogerson, W. P. (1985). "Repeated Moral Hazard", *Econometrica,* Vol.53,No.1., pp. 69-76.

16. Spear, S. E., & Srivastava, S. (1987). "On Repeated Moral Hazard with Discounting", *The Review of Economic Studies*, Vol.54, No.4., pp. 599-617.

17. Thaler, R. (1981). "Some Empirical Evidence on Dynamic Inconsistency", *Economics Letters*, Vol.8, No.3, pp. 201-207.

18. Yılmaz, M. (2013). "Repeated Moral Hazard with a Time-Inconsistent Agent", *Journal of Economic Behavior and Organization*, Vol.95, pp.70-89.