The Ratchet Effect Re-examined: A Learning Perspective

V Bhaskar[*]

University College London

April 2014

Abstract

We study dynamic moral hazard where principal and agent are symmetrically uncertain about job difficulty. Since effort is unobserved, shirking leads the principal to believe that the job is hard, increasing the agent's continuation value. So deterring shirking requires steeper incentives, which induce the agent to over-work today, since he can quit if the principal believes that the job is easy. With continuous effort choices, *no interior effort is implementable* in the first period. The agent's continuation value function is non-differentiable and convex, since the principal makes the agent indifferent between his discrete (participation) choices in the second period. The problem can be solved if the agent's participation decision is made continuous, or if there are long-term commitments, and we provide conditions for the first order approach to work. However, the impossibility result recurs in other agency models that combine discrete and continuous choices.

Keywords: ratchet effect, moral hazard, learning, envelope theorem, first-order approach. JEL codes: D83, D86.

[*] Thanks to Sunjoo Hwang, George Mailath, Meg Meyer, Alessandro Pavan, Caroline Thomas and seminar audiences at Arizona State, Austin, Bonn, Chicago, Duke-UNC, Mannheim, Michigan, Northwestern and UCL for helpful comments.

1 Introduction

The ratchet effect is one of the earliest problems noted by modern incentive theory, and was prominent in discussions of Soviet planning (Berliner, 1957). If the factory met or exceeded its plan target, the target for subsequent years was increased, reducing current effort incentives for the manager (Weitzman, 1980). The problem also arises in capitalist firms, as Milgrom and Roberts (1990) note. When a firm installs new equipment, firm and workers have to learn what is the appropriate work standard. It is efficient to use future information to adjust the standard. But this reduces work incentives today. Time and motion studies may reduce the degree of uncertainty regarding the technology, and ameliorate the effect, but their role is limited in contexts where a worker's performance improves with experience. Mathewson (1931), Roy (1952) and Edwards (1979) are workplace stud ies that document output restriction by workers.[1] The ratchet effect also arises in marketing, since salesmen are often paid bonuses that depend on exceeding targets, that are adjusted based on past performance. It also arises in a regulatory context, where both the regulator are uncertain about the effects of new technology (see Meyer and Vickers, 1997).

Theoretical work on optimal contracts in the presence of the ratchet effect usually assumes that agent already has private information. It studies dynamic mechanism design without commitment, and examines how the principal induces the agent to reveal his private information. This work includes Lazear (1986), Gibbons (1986), Freixas, Rochet & Tirole (1985), Laffont & Tirole (1986) and Carmichael and Macleod (2000). Lazear (1986) argues that high powered incentives are able to overcome the ratchet effect, and without any efficiency loss, assuming that the worker is risk neutral. Gibbons (1986) shows that Lazear's result depends upon an implicit assumption of long term commitment; in its absence, one cannot induce efficient effort provision by the more productive type. Laffont and Tirole (1988) prove a general result, that one cannot induce full separation given a continuum of types. Malcomson (2013) shows that the no full-separation result also obtains in a relational contracting setting, where the principal need not have all the bargaining power, as long as continuation play following full separation is efficient. A comprehensive discussion of the problem is found in Laffont and Tirole (1993), who consider both the case of binary types

[1] Interestingly, they find that workers collectively enforce norms of lower output, and sanction individuals who break the norm, highlighting the limitations of yardstick competition in overcoming the ratchet effect.

and of a continuum of types. We will relate our findings to their discussion after presenting our main result, theorem 7.

This paper studies the ratchet effect in a situation where both employer and the worker are learning about the technology, with ex ante symmetric information. When uncertainty pertains not to the worker's innate characteristic, but rather to the nature of the job or match specific productivity — as, for example, when new machinery is introduced – learning becomes important. If the worker does acquire private information, this takes time, and one must allow for contractual remedies that could address this at the outset. In hidden information environments, it is well known that contracting at the ex ante stage, before the agent has private information, is more efficient than ex post contracting. Indeed, if the agent is risk neutral, ex ante contracting enables full efficiency. One might expect similar results in our setting.

Milgrom and Roberts (1990) present an illuminating (albeit somewhat informal) discussion of the ratchet effect in a learning context. They assume a linear technology and normally distributed shocks, and argue that the ratchet effect implies that incentives need to be more high powered at the beginning of the relationship. Long term commitments alleviate the problem, but it may be hard to stick to these commitments since they are likely to be inefficient ex post. While their discussion is extremely insightful, they do not make explicit their assumptions.

This paper also relates to the literature on dynamic moral hazard with learning/experimentation. Holmstrom's (1982,1999) career concerns model is pioneering in this regard.[2] A crucial difference is that in the present paper, learning relates to the nature of the job rather than the agent's talent, and does not affect the outside option of the agent. More recently, there has been increased interest in agency models with learning, where the uncertainty also pertains to the nature of the project. Bergemann and Hege (1998, 2005), Manso (2011), Horner and Samuelson (2009) and Kwon (2011) analyze agency models with binary effort and binary signals. One key difference is that these papers usually assume limited liability, so that the agent's participation constraint does not bind, as it does for the main part of our analysis. Bhaskar and Mailath (2014) analyze a model of the ratchet effect with binary effort, without limited liability. The signal structure is similar to that assumed here, and the dynamic incentive problem arises since the agent can increase his continuation

[2] Extensions of the career concerns model include Gibbons and Murphy (1992) and Dewatripont, Jewitt and Tirole (1999).

value by shirking. They examine how the costs of incentivizing high effort vary with the length of the interaction, and show that the difficulty of the incentive problem increases at least linearly with time horizon, so that inducing high effort consistently becomes unprofitable.

There is also recent work that examines learning in agency models with private actions in continuous time and continuum action spaces – see De Marzo and Sannikov (2011), Cisternas (2012) and Prat and Jovanovic (2013). We will see that a crucial difference is that our paper allows both continuous choices (in the effort dimension) and discrete choices (regarding participation), whereas other papers allow either only discrete or only continuous choices.

We study optimal contracts where the principal and the agent are ex ante symmetrically uncertain about the difficulty of the job, and learn this over time. To focus on the ratchet effect, we assume that the principal cannot commit to long term contracts, but chooses short-term contracts optimally. In our baseline model, there is no limited liability and the principal has all the bargaining power, and thus the agent need not be paid any more than his outside option. Furthermore, since uncertainty pertains to the nature of the job, the outside option does not depend upon what is learned regarding the job. The ratchet effect arises from the possibility that the agent can manipulate the beliefs of the principal, by shirking. Consider a contract where the principal seeks to induce an interior effort level e^*. If the agent shirks and chooses $e < e^*$, then his beliefs will differ from that of the principal, since the principal updates her beliefs assuming that e^* has been chosen. If the agent becomes more pessimistic than the principal, he incurs no loss – the job only pays him his reservation utility in equilibrium, and he can quit if he earns less. If he becomes more optimistic than the principal, he earns a rent. Under fairly general conditions, we show that there must be some signal such that the agent is more optimistic than the principal when he shirks. Thus the agent raises his continuation value by shirking a little, and incentives have to be high powered in order to deter shirking. However, if the principal provides high powered incentives, this makes it profitable for the agent to over-work, i.e. to choose $e > e^*$, since by doing so, he increases his current payoff. This may cause the job to be more unattractive tomorrow, but since the agent is always free to quit, he incurs no loss in consequence. Thus, at any interior effort level e^*, the left hand derivative of the agent's continuation value function is negative, while the right hand derivative is non-negative. Thus the first order conditions for implementing e^* can never be satisfied, as long as e^* is in the interior of the agent's choice set.

The underlying reason for the failure of implementability is that we have an agency model where the agent makes choices from a continuum set (effort) as well as a discrete choice set (participation), so that the standard envelope theorem does not hold. If the agent's participation decision is continuous – e.g. because his outside option is uncertain and private information – then the continuation value is smooth at e^*, and the first order conditions for implementing interior effort are no longer inconsistent. Similarly, the problem does not arise in limited liability models with continuous effort choices, where the agent always has positive rents. We analyze such a limited liability model, and provide conditions under which the first-order approach (to the principal's optimization problem) can be applied to our dynamic context. Substantively, we find that under limited liability, the ratchet effect takes a different form as compared to a model where participation constraints bind. Indeed, for some parameter values, it is possible that the agent has increased incentives to work, so that dynamic considerations reduce the incentive problem. Similarly, if both principal and agent can commit to a binding long-term contract, implementing interior effort is possible.

Notwithstanding these positive result, we show that the impossibility result also arises in other agency settings that combine discrete and continuous choices, and where the principal cannot make long-term commitments. We analyze a limited liability model where effort is chosen a continuum in the initial period, and from a discrete choice set in the final period. If the agent faces discrete choices in the final period, then in equilibrium, the principal makes him indifferent between these choices, i.e. the incentive constraint binds. Any variation in the agent's beliefs causes this indifference to be broken, and in different ways, depending upon whether beliefs are more optimistic or pessimistic. This gives rise to a perverse kink in the agent's continuation value function, and ensures that the first order conditions for implementing interior effort in the first period cannot be satisfied. Thus the problem identified in this paper is more general than the specific contexts that we examine. Here again, the problem can be resolved if the principal can commit to the second period contract, and can also commit not renegotiate it.

The rest of this paper is as follows. Section 2 sets out our basic model of the ratchet effect. Section 3 shows that if we make the participation decision continuous, this eliminates discrete choice, and helps overcome the problem of implementability. This is possible if the agent's future reservation utility is random and private information, if the agent has limited liability, or if there is full bilateral commitment. Section 4 modifies the limited liability model by allowing for discrete choices in the final period, and show that it gives rise to non-implementability, just as in the baseline model with discrete participation.

2 The Model

Our model combines moral hazard with uncertainty regarding job difficulty. There are two states of the world $\omega \in \{B, G\}$, with the job being good (easy) in G, and bad (hard) in B, with $\lambda \in (0,1)$ denoting the common prior that $\omega = G$. The uncertainty concerns how difficult it is to succeed on *this* job. Importantly, learning does *not* affect outside option of the agent, which is fixed, and normalized to 0. The agent chooses effort $e \in [0,1]$, at cost $c(e)$, which is increasing, strictly convex and differentiable. The agent learns knowing his own effort choice and a realized public signal, $y \in Y$, where $Y := \{y^1, y^2, \ldots, y^K\}$ is a finite set of signals. The principal learns knowing only the signal, since the agent's effort is not public (i.e., it is not observed by the principal). The agent's flow utility from a wage payment $w \in \mathbb{R}$ and choosing $e \in [0,1]$ is $u(w) - c(e)$, where u is strictly increasing and strictly concave (so the agent is risk averse). We assume that wage payments are unrestricted, so that there is unlimited liability (section 4) examines limited liability).

A spot contract specifies the wage payment as a function of the realized signal. It is more convenient to work with utility schedules, so we write a spot contract as $u := (u^1, \ldots, u^K)$, where u^k is the utility the agent will receive after signal y^k. Let $w(.) = u^{-1}(.)$ denote the inverse function corresponding to $u(.)$.

The principal is risk neutral and her flow utility is $y - w(u)$. In each period, the principal makes a take-it-or-leave-it offer of a spot contract to the agent. If the agent refuses, the relationship is dissolved and the game ends.

The probability of signal y^k at action e and state $\omega \in \{B, G\}$ is $p^k_{e\omega}$. Consider first the extremal efforts, $e \in \{0, 1\}$. We will assume that a "high" signal is both a signal of the good state and of high effort. We capture this by the following assumption.

Assumption 1

1. There exists an *informative* signal, i.e., $\exists y^k \in Y$ such that $p^k_{0B} \neq p^k_{1G}$. For any

informative signal $y^k \in Y$,

$$\min\left\{p^k_{0B}, p^k_{1G}\right\} < p^k_{0G}, p^k_{1B} < \max\left\{p^k_{0B}, p^k_{1G}\right\}.$$

2. Signals have full support: $p^k_{e\omega} > 0$ for all k, e, ω.

The probability of signal y^k at a belief μ on G is $p^k_{e\mu} = \mu p^k_{eG} + (1-\mu) p^k_{eB}$. Order the signals so that $p^k_{0\mu}/p^k_{1\mu}$ is decreasing in k. Partition the set of signals into a set of "high" signals Y^H, "low" signals Y^L, and neutral $Y \setminus (Y^H \cup Y^L)$ by defining

$$y^k \in Y^H \iff p^k_{1G} > p^k_{0B},$$

$$y^k \in Y^L \iff p^k_{1G} < p^k_{0B}.$$

Assumption 1 implies

$$y^k \in Y^H \iff p^k_{1G} > p^k_{1B}, y^k_{0G} > p^k_{0B} \iff p^k_{0\mu} > p^k_{0\mu},$$

$$y^k \in Y^L \iff p^k_{1G} < p^k_{1B}, y^k_{0G} < p^k_{0B} \iff p^k_{1\mu} < p^k_{0\mu}.$$

In other words, high signals arise with higher probability when either the agent exerts effort *or* the state is good.

Our second assumption extends the information structure to all effort levels in $[0,1]$.With a continuum of effort levels, we need to employ the first-order approach to solve for the optimal contract, even in the static case. We therefore assume the Hart-Holmstrom (1987) sufficient conditions for the validity of this approach, and our assumption is an adaptation of their conditions.

Assumption 2: For any $y^k \in Y$, and any $\omega \in \{G, B\}, p^k_{e\omega} = e p^k_{1\omega} + (1-e) p^k_{0\omega}$.[3]

We assume that the principal and agent interact for two periods – two periods suffice to make the main points of our paper (since our main result is a negative one, it will be immediate that it also extends to any finite length interaction). The agent discounts future payoffs at rate $\delta \in (0,1]$. The principal's discount factor is possibly

[3] Hart and Holmstrom (1987) assume a linear cost of effort and that the probability of y^k is a convex combination of two distributions, a "good" one and a "bad". They assume that the weight on the good distribution is an increasing and concave function of effort. To see that our parametrization is equivalent to theirs, define a new effort variable, $c(e)$. This gives linear costs and a concave weighting function.

different, but will play little role in the analysis. Since neither the principal nor the agent can commit in period one regarding the contract in period two, payments must satisfy incentive compatibility and individual rationality period by period.

We study the dynamic game induced by this contracting problem, and solve for perfect Bayesian equilibria that satisfy sequential rationality, with beliefs given by Bayes rule. Sequential rationality ensures that the contract offered by the principal at $t = 2$ is optimal, and so the agent's participation constraint binds. We do not have to deal with out of equilibrium beliefs, since we have a game with no observable deviations by the informed party. Since effort choice by the agent is private and public signals have full support, the principal does not see an out of equilibrium action, except when the game ends by the agent refusing the contract (at which point, beliefs are moot). Deviations by the uninformed party (the principal) have no implications for beliefs.

2.1 Deterministic effort

We begin by considering pure strategy equilibria, where the effort choice by the agent in period one is deterministic. If the agent chooses effort e^* at $t = 1$, and output y^k is realized, then the common belief of the principal and agent at $t = 2$ is denoted by $\mu^k_{e^*}$. Sequential rationality implies that the principal offers a profit maximizing contract at $t = 2$. We assume that the project is profitable at all beliefs at $t = 2$, so that the principal always induces the agent to participate.

2.1.1 The Final Period

Let μ denote the belief of the principal, given the observed output y^k and the induced first period effort, e^*. Since we are focusing on a pure strategy equilibrium, the principal's second order beliefs are degenerate – she believes that the agent also has beliefs μ.

Definition 1 *Effort $\hat{e}$ is* implementable *at $t = 2$ if there exists a spot contract u such that $\hat{e}$ is optimal for the agent under belief μ. A period 2 contract $(\hat{e}, \hat{u})$ is optimal given belief μ if it maximizes the principal's profits $\mathbf{E}_{e,\mu}(y - w(u))$ over all (e, u) such that e is implementable.*

Claim 2 *In the final period, for any public belief μ, every effort $e \in [0, 1]$ is implementable. The profit maximizing contract , $(\hat{e}(\mu), u(\mu))$, satisfies the first order* *conditions for the principal's maximization problem, and the agent's individual rationality constraint binds. The agent's incentive constraint binds if $e > 0$. For any contract u, there is a unique effort level that maximizes the agent's utility.*

We omit the proof this claim, since it is straightforward, and almost identical to the argument in Hart and Holmstrom (1987). Sequential rationality implies that the principal always chooses the optimal contract at $t = 2$, given any belief μ. Assume that for any μ, the effort induced by the principal, $\hat{e}(\mu)$, is non-zero – this assumption is mild if $c_e(0) = 0$.

We now analyze the agent's payoff in the final period when his belief is π and differs from the principal's belief μ. Let $\hat{V}(\pi, \mu) = \max_e (p_{e\pi}.u_\mu - c(e))$ denote the payoff to the agent, conditional on accepting the job and choosing effort optimally. Since the agent will quit when he gets less than his outside option, let $V(\pi, \mu) = \max\{\hat{V}(\pi, \mu), 0\}$ denote his payoff given optimal participation. $V(\mu, \mu) = 0$ when $\pi = \mu$ since the agent's participation constraint binds under the optimal contract if he has the same beliefs. Finally, when the principal and agent's beliefs differ, V is computed under the distribution $p_{e\pi}$, i.e. this reflects the fact that the agent has the correct beliefs, since the difference in beliefs arises due to the fact that the agent knows his actual effort choice at $t = 1$.

Lemma 3 *$V(\pi, \mu) > 0$ if $\pi > \mu$, $V(\pi, \mu) = 0$ if $\pi \leq \mu$. $\hat{V}(\pi, \mu)$ is a differentiable function of π. $V(.)$ is convex in π.*

Proof. From claim 2, the optimal contract in the final period, u, must satisfy the first order conditions for $\hat{e}$ to be optimal at μ, i.e. [4]

$$u.(p_{1\mu} - p_{0\mu}) = c_e(\hat{e}) > 0. \tag{1}$$

In an optimal static contract, utility payments u^k must be increasing, since they are ordered in terms of the likelihood ratio. Thus u can be written as $u = z.\mathbf{1} + \tilde{u}$, where $z.\mathbf{1}$ is a vector where each component equals z, and $\tilde{u}^k > 0$ if $y^k \in Y^H$, and if $\tilde{u}^k < 0$ if $y^k \in Y^L$.The agent's payoff from his optimal effort choice at π is no less than his payoff from choosing $\hat{e}$ at π, which equals

$$[u.p_{\hat{e}\pi} - c(\hat{e})] - [u.p_{\hat{e}\mu} - c(\hat{e})] = (\pi - \mu)u.(p_{\hat{e}G} - p_{\hat{e}B}) = (\pi - \mu)\tilde{u}.(p_{\hat{e}G} - p_{\hat{e}B}),$$

[4] If $\hat{e}(\mu) = 1$, then equation (1) applies to the left hand derivative of $c(e)$ at 1.

since $\mathbf{1}.(p_{\hat{e}G} - p_{\hat{e}B}) = 0$. Assumption 1 implies that $p^k_{\hat{e}G} - p^k_{\hat{e}B} > 0$ if $y^k \in Y^H$ and $p^k_{\hat{e}G} - p^k_{\hat{e}B} < 0$ if $y^k \in Y^L$, and thus $\tilde{u}.(p_{\hat{e}G} - p_{\hat{e}B}) > 0$.

Letting $\tilde{e}(\pi)$ denotes the optimal effort choice at belief $\pi, \hat{V}(\pi, \mu) = p_{\tilde{e}(\pi)\pi}.u - c(\tilde{e}(\pi)$.The derivative with respect to π equals

$$\begin{aligned} \frac{d\hat{V}(\pi,\mu)}{d\pi} &= \left(p_{\tilde{e}(\pi)G} - p_{\tilde{e}(\pi)B}\right).u + \frac{d\tilde{e}}{d\pi}\left[\left(p_{1\pi} - p_{0\pi}\right).u - c_e(\tilde{e}(\pi))\right] \\ &= \left(p_{\tilde{e}(\pi)G} - p_{\tilde{e}(\pi)B}\right).u, \end{aligned} \tag{2}$$

since the second term is zero by the envelope theorem.

Given any $\pi > \mu$ and any $\pi', \hat{V}(\pi', \mu)$ is bounded below by the linear function $(\pi - \mu)\tilde{u}.(p_{\tilde{e}(\pi)G} - p_{\tilde{e}(\pi)B})$, and is thus convex in π. Since V equals the maximum of $\hat{V}$ and 0, it is also convex in π. ■

2.1.2 The First Period

Suppose that the principal seeks to induce effort level e^* at $t = 1$. If the agent deviates and chooses e different from e^*, then the principal and agent will have different second period beliefs after output y^k. The principal will have belief $\mu^k_{e^*}$, while the agent will have belief π^k_e. Thus the expected second period continuation value of the agent from choosing e when the principal induces e^* equals

$$W(e, e^*) = \sum_{y^k \in Y} p^k_{e\lambda} V(\pi^k_e, \mu^k_{e^*}).$$

Each term under the summation sign is non-negative, since $V(\pi^k_e, \mu^k_{e^*}) \geq 0$, given that the agent can always quit. Thus $W(e, e^*)$ is strictly positive as long as there is some y^k such that $\pi^k_e > \mu^k_{e^*}$.The following two lemmata show that downward deviations are strictly profitable, in terms of increasing agent's continuation value.

Lemma 4 *There exist a partition of Y into Y^D, Y^U and $Y \setminus (Y^U \cup Y^D)$ such that for any $e, e^* \in [0, 1]$ with $e < e'$: $\pi^k_e > \mu^k_{e^*}$ if $y^k \in Y^D, \pi^k_e < \mu^k_{e^*}$ if $y^k \in Y^U$ and $\pi^k_e = \mu^k_{e^*}$ if $y^k \in Y \setminus (Y^U \cup Y^D)$.*

Proof.

$$\begin{aligned}\pi_e^k - \mu_{e^*}^k &= \frac{\lambda p_{eG}^k}{p_{e\lambda}^k} - \frac{\lambda p_{e^*G}^k}{p_{e^*\lambda}^k} \\ &= \frac{\lambda}{p_{e\lambda}^k p_{e^*\lambda}^k}\left(p_{eG}^k p_{e^*\lambda}^k - p_{e^*G}^k p_{e\lambda}^k\right).\end{aligned}$$

Using the fact that p_{eG}^k is a convex combination of p_{1G}^k and p_{0G}^k (and similarly $p_{e^*G}^k, p_{e^*\lambda}^k$ and $p_{e\lambda}^k$), this can be re-written as

$$\pi_e^k - \mu_{e^*}^k = \frac{\lambda(1-\lambda)}{p_{e\lambda}^k p_{e^*\lambda}^k}(e^* - e)\left[p_{0G}^k p_{1B}^k - p_{0B}^k p_{1G}^k\right]. \tag{3}$$

For any $e < e^*$,the sign of $\pi_e^k - \mu_{e^*}^k$ only depends on the sign of $p_{0G}^k p_{1B}^k - p_{0B}^k p_{1G}^k$, proving the lemma. ■

The following lemma is key for our results, since it shows that Y^D is non-empty – there is at least one signal such that the agent is more optimistic than the principal when he shirks. We prove a more general result, that the agent is on average more optimistic than the principal, since it is of independent interest, especially under limited liability, where the agent's participation constraint need not bind.

Lemma 5 $\mathbf{E}_{0,\lambda}(\pi_0^k) > \mathbf{E}_{0,\lambda}(\mu_1^k)$, *so that* Y^D *is non-empty.*

Proof. From the martingale property of beliefs, $\mathbf{E}_{0,\lambda}(\pi_0^k) = \mathbf{E}_{1,\lambda}(\mu_1^k) = \lambda$, i.e.

$$\sum_Y p_{0\lambda}^k \pi_0^k = \sum_Y p_{1\lambda}^k \mu_1^k.$$

Subtract $\sum_Y p_{0\lambda}^k \mu_1^k$ from both sides to get

$$\sum_Y p_{0\lambda}^k(\pi_0^k - \mu_1^k) = \sum_Y (p_{1\lambda}^k - p_{0\lambda}^k)\mu_1^k.$$

Since $\lambda \sum_Y (p_{1\mu}^k - p_{0\mu}^k) = 0$,

$$\sum_Y p_{0\lambda}^k(\pi_0^k - \mu_1^k) = \sum_Y (p_{1\lambda}^k - p_{0\lambda}^k)(\mu_1^k - \lambda).$$

Under assumption A1, for any k, $(p_{1\lambda}^k - p_{0\lambda}^k)$ has the same sign as $(\mu_1^k - \lambda)$ – i.e. a signal that has higher probability under high effort is also informative of the job being easier. Since there is some informative signal, we conclude that $\sum_Y p_{0\lambda}^k(\pi_0^k - \mu_1^k) > 0$, i.e. the expectation of the difference in beliefs under the experiment $e = 0$ is strictly positive. Thus there must be some signal y^k such that $\pi_0^k > \mu_1^k$. ■

Lemma 5 follows from assumption 1 and implies that the agent can always increase his continuation value by shirking, since there is at least one signal where he is more optimistic than the principal. Thus the ratchet effect obtains under a fairly general information structure – most existing work assumes either binary or normal signals. Assumption 1 plays a similar role in Bhaskar and Mailath (2014), which examines in the long run consequences of belief manipulation. Lemmata 4 and 5 are robust – one can have some signals that violate assumption 1, as long as the probability of these signals is small. Since the inequalities in the proof of the lemmata are strict, the result will continue to apply if we have a small perturbation of an information structure that satisfies assumption 1.

If Y^U is empty, so that downward effort deviations cause the agent to become weakly more optimistic after every signal realization, the signal structure satisfies *uniform optimism.*. The following binary example, with $Y = \{H, L\}$, and the probability of H given by the table below, is illustrative:

	$e = 1$	$e = 0$
G	p	$q + \theta$
B	$p - \gamma$	q

Table 1: Binary signals: $\Pr(H|e, \omega)$

Let $p > q$, so that H is in fact a high signal. This signal structure satisfies Assumption 1 if $\theta, \gamma \in (0, p - q)$, so that $e = 1$ makes H more likely independent of the state, and also $\omega = G$ makes H more likely independent of the effort level.

For our first parametrization, let $\gamma \simeq 0$ and $\theta \simeq p - q$, i.e. effort is more effective in state B. In this case, lower effort makes for a more informative experiment. If the agent shirks, he is more optimistic than the principal when he succeeds, and less optimistic when he fails, i.e. $Y^D = \{H\}$ and $Y^U = \{L\}$.

For a second parametrization, suppose that $\gamma \simeq p - q$ and $\theta \simeq 0$, so that effort has a larger effect on the probability of high output in state G. Higher effort makes for a more informative experiment. If the agent shirks, he is more optimistic than the principal after L and less optimistic after , i.e. $Y^D = \{L\}$ and $Y^U = \{H\}$.

Finally, consider the case where $\gamma \simeq \theta$, so that effort has similar effects in both states. If the agent shirks, he is more optimistic than the principal after both signals, so that $Y^D = \{H, L\}$. This signal structure satisfies uniform optimism.

In the light of lemma 4 we may re-write $W(e, e^*)$ as

$$W(e,e^*) = \left\{ \begin{array}{l} \sum_{y^k \in Y^D} p^k_{e\lambda} V(\pi^k_e, \mu^k_{e^*}) \text{ if } e < e^* \\ \sum_{y^k \in Y^U} p^k_{e\lambda} V(\pi^k_e, \mu^k_{e^*}) \text{ if } e > e^*. \end{array} \right.$$

The overall payoff of the agent from effort choice e in the first period, given a contract u that seeks to induce effort level e^*, equals $v(e, e^*; u) := u.p_{e\lambda} - c(e) + \delta W(e, e^*)$.

Definition 6 *Effort e^* is* implementable *in period* 1 *if there exists a spot contract u such that e^* maximizes $v(e, e^*; u)$.*

Theorem 7 *Assume that optimal effort at $t = 2$ is not zero, for all relevant beliefs. If $e^* \in (0, 1)$, then e^* is **not** implementable at $t = 1$.The extremal efforts 0 and 1 are implementable.*

The proof is as follows. We evaluate the left-hand and right-hand derivatives of $W(e, e^*)$ at $e = e^* \in (0, 1)$, and show that these are inconsistent with the first order conditions for implementing e^*.The left hand derivative is given by

$$\left.\frac{\partial W^-(e,e^*)}{\partial e}\right|_{e=e^*} = \sum_{y^k \in Y} \frac{\partial p^k_{e\lambda}}{\partial e} V(\pi^k_{e^*}, \mu^k_{e^*}) + \sum_{y^k \in Y^D} p^k_{e^*\lambda} \left.\frac{\partial V^+(\pi^k_e, \mu^k_{e^*})}{\partial \pi^k_e}\right|_{\pi^k_e = \mu^k_{e^*}} \left.\frac{\partial \pi^k_e}{\partial e}\right|_{e=e^*}.$$

Since $V(\pi^k_{e^*}, \mu^k_{e^*}) = 0$, and since (by lemma 3) the right-hand derivative of V is the derivative of $\hat{V}$, this equals

$$\left.\frac{\partial W^-(e,e^*)}{\partial e}\right|_{e=e^*} = \sum_{y^k \in Y^D} p^k_{e^*\lambda} \left.\frac{\partial \hat{V}(\pi^k_e, \mu^k_{e^*})}{\partial \pi^k_e}\right|_{\pi^k_e = \mu^k_{e^*}} \left.\frac{\partial \pi^k_e}{\partial e}\right|_{e=e^*}. \qquad (4)$$

The derivative of $\hat{V}$ equals $\left(p_{\tilde{e}(\mu)G} - p_{\tilde{e}(\mu)B}\right).u > 0$. From equation 3, $\left.\frac{\partial \pi^k_e}{\partial e}\right|_{e=e^*}$ has the same sign as $\left[p^k_{1G}p^k_{0B} - p^k_{1B}p^k_{0G}\right]$, which is strictly negative if $y^k \in Y^D$. Since each term in the summation is strictly negative, $\left.\frac{\partial W^-(e,e^*)}{\partial e}\right|_{e=e^*} < 0$.

The right-hand derivative, $\frac{\partial W^{+}(e,e^{*})}{\partial e}\Big|_{e=e^{*}}$, is bounded below by zero, since $W(e,e^{*}) \geq 0$ and $W(e^{*},e^{*}) = 0$. If there is uniform optimism, $\frac{\partial W^{-}(e,e^{*})}{\partial e}\Big|_{e=e^{*}} = 0$, since Y^{U} is empty. If Y^{U} is non-empty, then similar arguments as for the left-hand derivative show that $\frac{\partial W^{+}(e,e^{*})}{\partial e}\Big|_{e=e^{*}} > 0$. Figures 1a. and 1b. graph the two possible shapes for the continuation value function $W(e,e^{*})$, as a function of e. In either case, the function is kinked at $e = e^{*}$, with the right-hand derivative being strictly larger than the left-hand derivative.

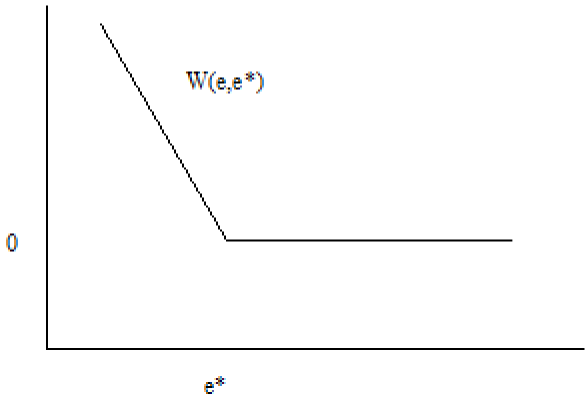


Fig. 1a.
W(e,e*) under uniform optimism

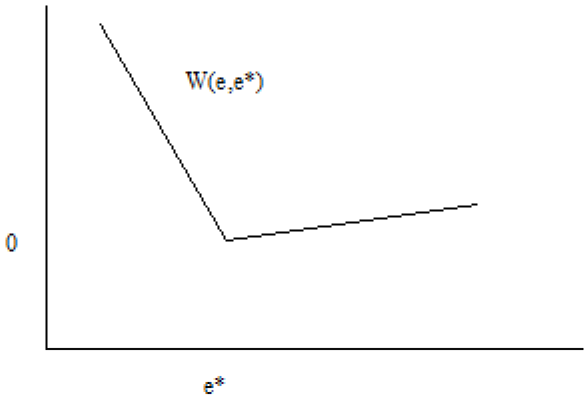


Fig. 1b.
W(e,e*) without uniform optimism

The agent's current payoff at $t = 1$ is a smooth function of effort, given the differentiability of the cost of effort and expected utility. Thus the first order conditions for e^{*} to be optimal for the agent at $t = 1$ are:

$$(p_{1\lambda} - p_{0\lambda})\,.u - c_{e}(e^{*}) + \delta\,\frac{\partial W^{-}(e,e^{*})}{\partial e}\Bigg|_{e=e^{*}} \geq 0.$$

$$(p_{1\lambda} - p_{0\lambda})\,.u - c_{e}(e^{*}) + \delta\,\frac{\partial W^{+}(e,e^{*})}{\partial e}\Bigg|_{e=e^{*}} \leq 0.$$

Since $\frac{\partial W^{-}}{\partial e}\Big|_{e=e^{*}} < \frac{\partial W^{+}}{\partial e}\Big|_{e=e^{*}}$, the two conditions cannot be simultaneously satisfied, thereby proving the main part of theorem.

The extremal efforts, 0 and 1, can be implemented, since one has to only deter deviations in one direction. For implementing $e = 0$, if the signal structure satisfies uniform optimism, then a constant utility schedule, with $u^k = c(0) \forall k$, is the optimal contract. However, if the signal structure does not satisfy uniform optimism, then the agent may have to be punished for higher output levels – he may have an incentive to deviate upwards, since he will be more optimistic than the principal after some output realizations. Let $W(e, 0)$ denote the agent's expected continuation value from effort level e given that the principal induces 0. It suffices to choose utility payments u such that $p_{e\lambda}.u - c(e) + W(e, 0)$ is maximized at $e = 0$. If $c(.)$ is sufficiently convex (to offset the convexity of the W function), then the first order condition suffices:

$$\left(p_{1\lambda} - p_{0\lambda}\right).u - c_e(0) + W_e^+(e, 0)\big|_{e=0} \leq 0.$$

Similarly, since $c_e(e)$ is continuous on a compact set, $c_e(1)$ exists, and so $e = 1$ can be implemented, by choosing u so that $(p_{1\lambda} - p_{0\lambda}.u - c_e(1) + W_e(1, 1) = 0$. This completes the proof of theorem 7.

The negative result in theorem 7 is striking: no interior effort level can be implemented in the first period. The ratchet effect is totally destructive of incentives. The ratchet effect implies that the agent can raise his continuation value by shirking a little relative to e^*. To overcome this, incentives today must be high powered, so that shirking reduces the agent's current payoff. However, this implies that the agent can increase his current payoff by over-working relative to e^* – this follows from the fact that current costs and benefits are smooth functions of effort. But over-working cannot *reduce* the agent's continuation value relative to e^*, since the agent can always quit. In other words, the principal can deter downward deviations, but this makes upward deviations profitable. Thus high powered incentives cannot overcome the ratchet effect, contrary to the suggestion in some of the literature.

The negative result obtains even though one has contracting before the agent has any private information. Thus ex ante contracting does not help, in contrast with models of hidden information, where it improves efficiency. The key reason appears to be the lack of inter-temporal commitment.

We now compare our result with those in models of the ratchet effect arising from ex ante private information – see Laffont and Tirole (1993, chapter 10) for a comprehensive discussion. With a continuum of types, the main result is that one cannot have full separation of types, so that there must be some pooling. With binary types, full separation may be possible, but may also be vulnerable to the "take the money and run" strategy, whereby the low type mimics the high type in the first period, and quits in the second period (this is somewhat similar to our finding in theorem 7). In this case, the equilibrium involves partial separation, and mixed strategies. As we will see, mixing may not resolve the problem in our model.

Our results do not depend upon the agent being risk averse. Suppose that the agent is risk neutral. As before, contracts are only for one period. In the final period, suppose that the belief is μ. Then the principal can make the agent the residual claimant of the project, by charging a fixed rental, $R(\mu)$. This must satisfy the individual rationality constraint, $\max_e [\mathbf{E}_{e,\mu}(y) - c(e)] - R(\mu) \geq 0$. Since the optimal contract maximizes $R(\mu)$ subject to this constraint, $R(\mu) = \max_e [\mathbf{E}_{e,\mu}(y) - c(e)]$. Thus the principal charges the agent a fee $R(\mu)$, that is increasing in μ under our assumptions. If the agent is offered $R(\mu)$, but has belief $\pi > \mu$, his payoff will be

$$V(\pi, \mu) = \max_e [\mathbf{E}_{e,\pi}(y) - c(e)] - R(\mu) = R(\pi) - R(\mu).$$

In particular, the derivative is given by

$$\left.\frac{dV^+(\pi, \mu)}{d\pi}\right|_{\pi=\mu} = \left(p_{\hat{e}(\mu)G} - p_{\hat{e}(\mu)B}\right).y > 0,$$

where $\hat{e}(\mu)$ is optimal effort given μ. Now consider the first period problem. Suppose that the principal wants to implement effort level e^*. The second period continuation value of the agent when he deviates to $e < e^*$ is given by $W(e, e^*) > 0$. The left hand derivative evaluated at $e = e^*$ is strictly negative, since $\left.\frac{dV^+(\pi,\mu)}{d\pi}\right|_{\pi=\mu} > 0$. Thus, in order to prevent downward deviations, the agent must be offered more high powered incentives than residual claimancy – his wage payments have to *more variable* than y. However, this implies that the agent has earn more than his outside option today by increasing his effort level beyond e^*, and quitting the job tomorrow, when signals in Y^D are realized. Thus no interior effort level is implementable even when the agent is risk neutral.

This problem can be solved if the agent can sign a long term contract, whereby he commits to buying the project for both periods. In this case, the agent will be willing to buy the project for

$$\max \left\{ \sum_k p^k_{e\lambda} \left\{ y^k + \delta \left[\mathbf{E}_{\hat{e}(\mu^k_e),\mu^k_e}(y) - c(\hat{e}(\mu^k_e)) \right] \right\} \right\}.$$

Thus long-term commitment can solve the ratchet effect, as we shall see further in sub-section 3.3.

2.2 Random effort

A natural question is whether interior effort levels are implementable with positive probability. That is, can the principal design a contract where the agent randomizes over effort levels in period 1, using a mixed strategy σ. In the static context (e.g. in period 2), random efforts can never be implemented, since the optimization problem induced by a contract u is strictly concave.

Randomization by the agent at $t = 1$ induces asymmetric information at $t = 2$, and may be a way for the principal to commit to pay some rents to the agent in the second period. Let $S(\sigma)$ denote the support of σ. On observing output y^k, the principal now believes that the agent has beliefs in the set $\{\mu_e^k\}_{e \in S(\sigma)}$, where the probability he assigns to μ_e^k is

$$\theta(\mu_e^k) = \frac{\sigma(e) p_{\lambda e}^k}{\sum_{e' \in S(\sigma)} \sigma(e') p_{\lambda e'}^k}.$$

The principal's problem at $t = 2$ is a screening/mechanism design problem with moral hazard, and without quasi-linear utilities (since the agent is risk-averse). The agent has beliefs in the set $\{\mu_e^k\}_{e \in S(\sigma)}$, and the principal assigns probabilities $\theta(\mu_e^k)$ to each of these beliefs. A general analysis of this problem is interesting but involved, and would take us far from the focus of this paper. We therefore restrict ourselves to the question of whether randomization resolves the implementability problem.

Let us now consider the screening problem more generally. Let M be an arbitrary set of types, where $\mu \in M$ is a belief regarding ω, and belongs to $[0, 1]$. Let θ denote a probability distribution on M. Let $\mathbb{R}^K \cup \emptyset$ denote the set of feasible contracts that can be offered by the principal. That is the principal may either offer a utility vector u specifying signal contingent payments or she may offer the null contract $\emptyset$, and not employ the agent. A direct mechanism ζ specifies an element of $\mathbb{R}^K \cup \emptyset$ for each member in the set M. In addition, we assume that ζ always contains $\emptyset$, regardless of what the principal offers, so that the agent always has the option of not taking the job. Fix a direct mechanism, and let $u_\mu \neq \emptyset$ denote the contract for type μ, who the principal would like to participate. Let $\hat{e}(\mu)$ denote the payoff maximizing effort choice for the agent given belief μ and contract u_μ. We shall call $\hat{e}(\mu)$ the effort induced by the contract u_μ.

The participation constraint for the agent implies $p_{\hat{e}(\mu)\mu}.u_\mu - c(\hat{e}(\mu)) \geq 0$. The truth-telling constraint, that type μ does not prefer a contract u_π is

$$p_{\hat{e}(\mu)\mu}.u_\mu - c(\hat{e}(\mu)) \geq \max_e \left[p_{e\mu}.u_\pi - c(e)\right].$$

A direct mechanism ζ is incentive-compatible if for any μ, the participation constraint is satisfied, and the truth-telling constraint is satisfied relative to any $\pi \in M$.

Consider a contract u_μ such that μ finds it optimal to participate. If $\pi > \mu$, then the payoff of type π from accepting contract μ is

$$\begin{aligned} \tilde{V}(\pi, \mu) &= \max_e \left[p_{e\mu}.u_\pi - c(e)\right] \\ &\geq p_{\hat{e}(\mu)\pi}.u_\mu - c(\hat{e}(\mu)) = (\pi - \mu)(p_{\hat{e}(\mu)G} - p_{\hat{e}(\mu)B}).u_\mu + \tilde{V}(\mu, \mu). \end{aligned}$$

If $\hat{e}(\mu) > 0$, $\tilde{V}(\pi, \mu) > \tilde{V}(\mu, \mu)$, since the inner product $(p_{\hat{e}(\mu)G} - p_{\hat{e}(\mu)B}).u_\mu$ has the same sign as $(p_{1\mu} - p_{0\mu}).u_\mu$ under assumption 1. Thus if type μ is induced to participate and exert positive effort, then incentive-compatibility implies that type $\pi > \mu$ must be given a strictly positive rent. Conversely, if either type μ does not participate or if $\hat{e}(\mu) = 0$, then the contract u_μ does not impose any cost in terms of additional rent for type $\pi > \mu$. (If $\hat{e}(\mu) = 0$, the agent's risk-aversion implies that the cost-minimizing contract u_μ is a constant vector, and thus $\tilde{V}(\pi, \mu) = \tilde{V}(\mu, \mu)$).

With asymmetric information, it may be optimal to exclude low types in the set M, as a way of reducing the rents paid to types with higher beliefs. Alternatively, the principal can allow their participation, but induce zero effort, since this has identical effects in terms of the rents that must be paid to higher belief types. We shall now assume that for any μ, the principal gets strictly higher profits from exclusion than from inducing zero effort – this could be due to fixed costs, that are not incurred if the agent is not employed. Thus, $\hat{e}(\mu) > 0$ for any u_μ that induces participation. We shall also assume that for any set M and distribution θ on M, the probability that the agent participates under the optimal mechanism ζ is non-zero. [5]

Suppose that the principal induces a random effort level σ at $t = 1$. Let (M^k, θ^k) denote the type/distribution pair following signal realization y^k. Sequential rational-

[5] Recall that when the principal was certain regarding μ, we assumed that the principal induced participation and non-zero effort at any μ.

ity implies that the principal offers an optimal mechanism ζ^k at (M^k, θ^k). Let $\bar{V}^k(\mu)$ denote the payoff of type μ under ζ^k. Let $\bar{\mu}^k$ denote the lowest type that is induced to participate by ζ^k. Let $\pi > \bar{\mu}^k$ be an arbitrary belief, i.e. π need not be an element of M^k, and let $\bar{V}^k(\pi)$ denote its optimal payoff under ζ^k. This is at least at least $(\pi - \mu)(p_{\hat{e}(\mu)G} - p_{\hat{e}(\mu)B}).u_{\bar{\mu}} > 0$. Thus the right-hand derivative of $V^k(\pi)$ with respect to π, evaluated at $\pi = \bar{\mu}^k$ is $(p_{\hat{e}(\bar{\mu}^k)G} - p_{\hat{e}(\bar{\mu}^k)B}).u_{\bar{\mu}^k} > 0$. (The left-hand derivative is zero, since types below $\bar{\mu}^k$ choose not to participate).

Define the ex-ante expected continuation value from choosing e, given that the principal induces σ by $\bar{W}(e, \sigma) = \sum_Y p^k_{e\lambda} \bar{V}^k(\mu^k_e)$. The overall payoff from e given a first period contract u that induces σ is $v(e, \sigma, u) = p_{e\lambda}.u - c(e) + \delta \bar{W}(e, \sigma)$.

Theorem 8 *Let σ be a probability distribution over effort levels where* $\max S(\sigma) \in (0, 1)$. *If the signal structure satisfies uniform optimism, σ is not implementable.*

Proof. Let v^* denote the overall payoff to the worker from any strategy in the support of σ. Let $< e_n >, e_n \in S(\sigma) \forall n$, converge to $\bar{e} = \max S(\sigma)$. Since $v(e_n, \sigma, u) = v^* \forall n$, and since both current payoffs and continuation values are continuous in $e, v(\bar{e}, \sigma; u) = \lim_{n \to \infty} v(e_n, \sigma; u) = v^*$. Thus $\bar{e}$ must be optimal for the agent.

Under uniform optimism, if $e > e', \mu^k_e < \mu^k_{e'} \forall k$. Let $\tilde{e}^k$ denote the supremum of the effort levels in set of included types after signal y^k. Let $\tilde{e} = \max\{\tilde{e}^k\}_{k=1}^K$. We show that $\tilde{e} = \bar{e}$. Clearly, $\bar{W}(e, \sigma) = 0$ if $e \geq \tilde{e}$ since the principal never induces the participation of any type $e > \tilde{e}$ after any signal. Thus the right hand derivative of $v(e, \sigma)$ at $\tilde{e}$ equals the derivative of $p_{e\lambda}.u - c(e)$, and this must be less than or equal to zero for $\tilde{e}$ to be optimal. If it is negative, then $\tilde{e}$ cannot be optimal since it is profitable to reduce effort, and thus $\tilde{e}$ maximizes $p_{e\lambda}.u - c(e)$. Since $p_{e\lambda}.u - c(e)$ is strictly concave, $\tilde{e}$ must be the unique maximizer, and so we deduce that $\tilde{e} = \bar{e}$, and thus $\tilde{e} \in (0, 1)$.

If $e < \tilde{e}, \mu^k_e > \mu^k_{\bar{e}}$ for any k, and $\tilde{V}(\mu^k_e, \mu^k_{\bar{e}}) > 0$ since $\hat{e}(\mu^k_{\bar{e}}) > 0$. Hence $\bar{W}(e, \sigma) = > 0$. Since the right-hand derivative of $\bar{V}(\mu^k_e)$ at $\bar{\mu}^k_e$ is strictly positive, the left-hand derivative of $\bar{W}(e, \sigma)$ at $\bar{e}$ is negative. The right-hand derivative $\bar{W}(e, \sigma)$ at $\bar{e}$ is zero, and so $\bar{e}$ cannot satisfy the first order conditions for maximizing $v(.)$. ■

This theorem has the following implication. Assume that there is uniform optimism, and suppose that $c_e(0) = 0$. Suppose that the principal offers a contract u at $t = 1$ with positive incentives, so that $e = 0$ does not maximize the agent's current payoff. However, $c(1)$ is sufficiently large so that it is a dominated choice given u.

Theorem 8 implies that the continuation game induced by this contract does not have a perfect Bayesian equilibrium.

The theorem assumes that at $t = 2$, inducing zero effort is less profitable for the principal than exclusion, so that any participating type exerts positive effort. If we did not make this assumption, we can still show that the conclusions of the theorem hold for any random effort σ with finite support. The theorem also assumes uniform optimism, and it remains an open question as to whether a random effort level with upper bound $\bar{e} < 1$ can be sustained when Y^U is non-empty, i.e. there are signals such that the agent is more pessimistic after a downward deviation. The non-participation of some effort types is necessary for a possibility result – if the principal ensures the participation of all types, $\bar{W}$ has a kink at $\bar{e}$, and the impossibility result applies. The critical factor that removes such a kink is if the type that chooses $\bar{e}$ is induced to participate after signals in Y^U, but not after signals in Y^D. This ensures that downward deviations from $\bar{e}$ do not increase the continuation value $\bar{W}$, thereby overcoming the problem of the perverse kink in the value function at $\bar{e}$. However, such an equilibrium must also satisfy additional constraints, implied by sequential rationality, and such a construction may not be feasible.

3 Solutions: Rents or Commitment

The impossibility result in theorem 7 is drastic – the first order conditions for implementing any interior effort cannot be satisfied. We now examine the underlying reason for this result, and also how it may be overcome. The underlying reason is that the agent has discrete choices – stay on the job or quit – due to which there is a failure of the standard envelope theorem. In consequence, the maximum value function of the agent fails to be differentiable, as a function of his beliefs. Envelope theorems exist for problems involving discrete choices (see Milgrom and Segal, 2002), but they do not deliver differentiability of the maximum value function. Now, kinks the maximum value function arise in other contexts, e.g. when the consumer has discrete choices, but they occur only at an isolated set of prices, and are therefore rare – any convex function is differentiable almost everywhere. However, in our agency context, the principal *designs* the contract so as to make the consumer indifferent between his discrete choices. Thus non-differentiability of the maximum value function is inevitable, at precisely the point that is relevant.

We now show that the impossibility result can be overcome – in the sense that the first order conditions for implementing interior effort can be satisfied – if make the agent's participation decision a continuous one. This is the case if the agent's reservation value at $t = 2$ is randomly drawn at the beginning of that period, and is private information or if the agent has limited liability, where the agent always gets rents. Finally, full commitment can also solve the problem.

3.1 Rents via Limited Liability

Consider first the situation where utility payments are bounded below, due to the limited liability of the agent (see e.g. Innes, 1990). In this case, the participation constraint may not bind in the final period, and so the agent gets rents. Let us suppose that this is always the case, i.e. for all public beliefs $\mu^k_{e^*}$ that arise in the second period, the optimal contract u offered by the principal is such that $V(\mu, \mu) > 0$. Now if the agent makes a small deviation e different from e^*, and has belief π^k_e, $\hat{V}(\pi^k_e, \mu^k_{e^*})$ is still positive, regardless of whether the deviation is upwards or downwards. The agent will no longer find it optimal to choose $\hat{e}$ (the effort the principal seeks to induce); however, since the agent's effort choice is from a continuum, his continuation value function $V(\pi, \mu)$ will be smooth as a function of π. In consequence, $W(e, e^*)$ will be differentiable at $e = e^*$, and the first order conditions for implementing e^* can be satisfied by the appropriate choice of first period contract. Some complications remain, since $V(\pi, \mu)$ is convex in π, implying that $W(e, e^*)$ is convex in e. Nevertheless, if the cost of effort, $c\,(e)$, is sufficiently convex, the overall first period payoff function of the agent, $v(e, e^*, u)$, is concave, so that e^* is implementable.

Let $\underline{u}$ denote the lower bound of utilities given by limited liability. Given belief μ, the principal chooses $\hat{e}$ and $u = \left(u^k\right)_{k=1}^K, u^k \geq \underline{u}$, to maximize $p_{\hat{e}\mu}((y - w(u)))$ subject to the IR constraint $p_{\hat{e}\mu}.u - c(\hat{e}) \geq 0$, and incentive constraint $\left(p_{1\mu} - p_{0\mu}\right).u = c_e(\hat{e})$.

We shall assume that at any μ, at the optimal contract u, limited liability constraints bind, so that the participation constraint does not. Thus the agent gets a payoff $\hat{V}(\mu, \mu) = p_{\hat{e}\mu}.u - c(\hat{e}) > 0$. The expected continuation value function of the agent, $W(e, e^*)$:, is given by

$$W(e, e^*) = \sum p^k_{e\lambda} \hat{V}(\pi^k_e, \mu^k_{e^*}).$$

Since $V(\pi^k_e, \mu^k_{e^*}) = \hat{V}(\pi^k_e, \mu^k_{e^*})$ if the agent always gets rents and since the latter is differentiable, $W(e, e^*)$ is also differentiable at $e = e^*$:

$$\left.\frac{\partial W(e,e^*)}{\partial e}\right|_{e=e^*} = \sum_{y^k \in Y} \frac{\partial p^k_{e\lambda}}{\partial e} \hat{V}(\pi^k_{e^*}, \mu^k_{e^*}) + \sum_{y^k \in Y} p^k_{e^*\lambda} \left.\frac{\partial \hat{V}(\pi^k_e, \mu^k_{e^*})}{\partial \pi^k_e}\right|_{\pi^k_e = \mu^k_{e^*}} \left.\frac{\partial \pi^k_e}{\partial e}\right|_{e=e^*}. \quad (5)$$

Thus the first order condition for implementing e^* can be satisfied by a suitable choice of u:

$$(p_{1\mu} - p_{0\mu})\,.u + \left.\frac{\partial W(e,e^*)}{\partial e}\right|_{e=e^*} = c_e(e^*).$$

We now turn to second order conditions. Let us assume that the $\hat{V}(\pi, \mu)$ is non-negative at every π that arises, so that participation constraints are satisfied. So, the second derivative of W is given by

$$\begin{aligned}\frac{\partial^2 W(e,e^*)}{\partial e^2} &= \sum_{y^k \in Y} \frac{\partial^2 p^k_{e\lambda}}{\partial e^2} \hat{V}(\pi^k_{e^*}, \mu^k_{e^*}) + 2 \sum_{y^k \in Y} \frac{\partial p^k_{e\lambda}}{\partial e} \frac{\partial \hat{V}(\pi^k_e, \mu^k_{e^*})}{\partial \pi} \frac{\partial \pi^k_e}{\partial e} \\ &\quad + \sum_{y^k \in Y} p^k_{e\lambda} \frac{\partial^2 \hat{V}(\pi^k_e, \mu^k_{e^*})}{\partial \pi^2} \left(\frac{\partial \pi^k_e}{\partial e}\right)^2 + \sum_{y^k \in Y} p^k_{e\lambda} \frac{\partial \hat{V}(\pi^k_e, \mu^k_{e^*})}{\partial \pi^k_e} \frac{\partial^2 \pi^k_e}{\partial e^2}.\end{aligned}$$

Note that $\frac{\partial^2 p^k_{e\lambda}}{\partial e^2} = 0$ and

$$\frac{\partial \pi^k_e}{\partial e} = \frac{\lambda(1-\lambda)\left[p^k_{0G} p^k_{1B} - p^k_{0B} p^k_{1G}\right]}{\left(p^k_{e\lambda}\right)^2},$$

$$2 \frac{\partial p^k_{e\lambda}}{\partial e} \frac{\partial \pi^k_e}{\partial e} + p^k_{e\lambda} \frac{\partial^2 \pi^k_e}{\partial e^2} = 0.$$

Thus the second derivative simplifies to

$$\frac{\partial^2 W(e,e^*)}{\partial e^2} = [\lambda(1-\lambda)]^2 \sum_{y^k \in Y} \frac{\partial^2 \hat{V}(\pi^k_e, \mu^k_{e^*})}{\partial \pi^2} \frac{\left[p^k_{0G} p^k_{1B} - p^k_{0B} p^k_{1G}\right]^2}{\left(p^k_{e\lambda}\right)^3}. \quad (6)$$

This is positive (and generally strictly positive) since $\hat{V}$ is convex in π. We now examine the second derivative of $\hat{V}$ with respect to π. Re-write $\hat{V}$ as:

$$\hat{V}(\pi,\mu) = \hat{V}(\mu,\mu) + (\pi-\mu)\left(p_{\hat{e}G} - p_{\hat{e}B}\right).u + (\tilde{e}-\hat{e})\left(p_{1\pi} - p_{0\pi}\right).u - c(\tilde{e}(\pi)) + c(\hat{e}).$$

Using the envelope theorem, the derivative with respect to π equals

$$\frac{\partial \hat{V}(\pi,\mu)}{\partial \pi} = \left(p_{\hat{e}G} - p_{\hat{e}B}\right).u + (\tilde{e}-\hat{e})\,\rho.u(\mu), \tag{7}$$

where the vector ρ is defined as

$$\rho := p_{1G} + p_{0B} - p_{0G} - p_{1B}, \tag{8}$$

and $u(\mu)$ denotes the optimal contract in the second period given public belief μ. The second derivative of $\hat{V}$ equals

$$\frac{\partial^2 \hat{V}(\pi,\mu)}{\partial \pi^2} = \frac{d\tilde{e}}{d\pi}\rho.u(\mu) = \frac{\left[\rho.u(\mu)\right]^2}{c_{ee}(\tilde{e}(\pi))}, \tag{9}$$

since $\frac{d\tilde{e}}{d\pi} = \frac{\rho.u}{c_{ee}(\tilde{e})}$. Now generically, ρ will not equal the null vector. Furthermore, $u(\mu)$ is not a constant vector since we have assumed that the profit maximizing effort level $\hat{e}(\mu) > 0$. Since generically, $\rho.u \neq 0$, $\frac{\partial^2 \hat{V}(\pi,\mu)}{\partial \pi^2} > 0$, and so $\frac{\partial^2 W(e,e^*)}{\partial e^2} > 0$.

If $\rho = 0$, then W is linear and $v(e, e^*, u)$ is strictly concave, as long as $c(.)$ is strictly convex. So for a given convex c, then if ρ is sufficiently close to zero, then v is strictly concave. Conversely, if c is close to linear, then we can find ρ so that the second order conditions are not satisfied.

In order to examine more closely the degree of convexity of $\hat{V}$, we examine $u(\mu)$, i.e. the optimal contract in the second period. This will depend upon $\hat{e}(\mu)$, the profit maximizing effort choice at μ. Lemma 12 in the appendix shows that when the agent is risk-neutral, then for any value of $\hat{e}$, the optimal cost minimizing contract requires the principal to make payments depend only on a binary partition of the signal space Y, paying $u^K > \underline{u}$ after the signal (or signals) with the highest likelihood ratio, and $\underline{u}$ after all other signals. In the light of this lemma, we simplify the exposition by assuming there is a single signal with the highest likelihood ratio for all values of μ, and label this y^H.[6] This reduces to the case of binary signals as in table 1, and from

[6] Without this assumption, the partition of the signal space may depend upon μ, but this does not cause any essential difficulties, since there can be only finitely many such partitions.

the table, $\rho = (\gamma - \theta, \theta - \gamma)$.The second period incentive constraint is

$$(p^H_{1\mu} - p^H_{0\mu})(u^H - u^L) = c_e(\hat{e}(\mu)),$$

$$|\rho.u| = |(\theta - \gamma)|\,(u^H - u^L) = \frac{|(\theta - \gamma)|\,c_e(\hat{e}(\mu))}{(p^H_{1\mu} - p^H_{0\mu})}. \tag{10}$$

The second derivative of the overall payoff in the first period at e, when the principal seeks to implement e^* equals $\frac{\partial^2 W(e,e^*)}{\partial e^2} - c_{ee}(e)$. The first term depends upon $\frac{\partial^2 \hat{V}(\pi,\mu)}{\partial \pi^2}$, and this can be made as small as desired by making the cost function $c(.)$ sufficiently convex. Take the case of quadratic costs, $c(e, \phi) = \frac{\phi}{2}e^2$. In the appendix, we show that $c_e(\hat{e}(\mu))$ is decreasing in ϕ, so the numerator of 10 is decreasing in ϕ. Since the denominator equals ϕ, $\frac{\partial^2 \hat{V}(\pi,\mu)}{\partial \pi^2}$ can be made as small as required by a suitably large choice of ϕ. Thus for ϕ large enough, the first period payoff function $v(e, e^*)$ is concave in e, and every $e^* \in [0, 1]$ is implementable. We summarize our results in the following theorem.

Theorem 9 *Suppose that the agent's limited liability constraint binds in the second period, so that the agent always participates at all second period beliefs that arise. Assume that either signals are binary or the agent is risk neutral, and that effort costs $c(e)$ are quadratic. If the second derivative of $c(.)$ is sufficiently large, every effort level in $[0, 1]$ is implementable at $t = 1$, and the principal's optimal contract at $t = 1$ solves the first order conditions for maximizing her payoff.*

Proof. See appendix. ■

The above theorem provides a sufficient condition for the first-order approach to work in a dynamic context. To our knowledge, this is the first time that the first-order approach has been extended to a dynamic setting. The sufficient condition is not necessary – for example, take a quadratic cost function $c(e)$ where the second derivative is sufficiently large such that the second order condition holds as a strict inequality for every value of e. One can perturb this cost function slightly, so that the third-derivative is no longer exactly zero, and if the perturbation is small enough, the second order conditions will continue to be satisfied. In our numerical examples, the degree of convexity of W is very small, so that concavity of the agent's first period payoff appears to be a non-issue.

Having established the validity of the first order approach, we now turn to the economic implications of the ratchet effect under limited liability. Re-write the derivative of the agent's expected continuation value as

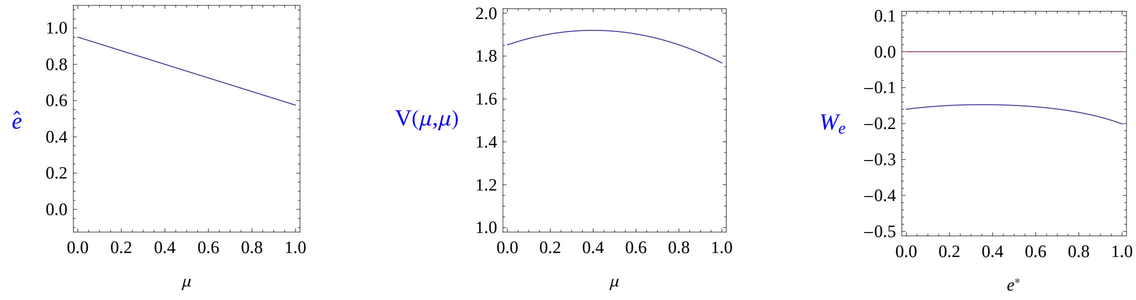
$$\left.\frac{\partial W(e,e^*)}{\partial e}\right|_{e=e^*} = \sum_{y^k \in Y} \frac{\partial p^k_{e\lambda}}{\partial e} \hat{V}(\pi^k_{e^*}, \mu^k_{e^*}) + \sum_{y^k \in Y^D} p^k_{e^*\lambda} \left.\frac{\partial \hat{V}(\pi^k_e, \mu^k_{e^*})}{\partial \pi^k_e}\right|_{\pi^k_e = \mu^k_{e^*}} \left.\frac{\partial \pi^k_e}{\partial e}\right|_{e=e^*}$$
$$+ \sum_{y^k \in Y^U} p^k_{e^*\lambda} \left.\frac{\partial \hat{V}(\pi^k_e, \mu^k_{e^*})}{\partial \pi^k_e}\right|_{\pi^k_e = \mu^k_{e^*}} \left.\frac{\partial \pi^k_e}{\partial e}\right|_{e=e^*}. \tag{11}$$

The importance of the ratchet effect depends on the magnitude (and sign) of the above derivative. The most important difference, as compared to the case where the agent's reservation utility constraint binds, arises due to the fact that now the agent earns some rents, and these rents vary with μ, the public belief. This is seen by comparing the above expression 11 with the expression for the left-hand derivative, 4 that we had earlier. While the second term is common in both and is negative – higher effort reduces V after signals in Y^D, there are two additional terms in the present case, that are zero in 4. The third term in 11 is positive, if Y^U is non-empty – shirking makes the agent more pessimistic after signals in Y^D, and he stays on the job with positive probability, and this reduces the benefit of shirking. Most importantly, the first term is not zero since $V(\mu,\mu)$ is not constant. In general, it may be increasing or decreasing in μ, due to the interaction of two factors. First, a higher value of μ makes it more likely that the agent will succeed, and thus incentive pay is more costly for the principal. This reduces $\hat{e}$ as a function of μ, and therefore reduces incentive pay and rents (even though the direct effect of an increase in μ is to increase the probability of success and rents). Second, from a revenue standpoint, the principal may have differential incentives to induce effort in different states of the world, and this also affects rents, since a higher value of $\hat{e}$ increases rents.

These considerations may be easier to comprehend in the context of some numerical examples. Take the binary signal setting of table 1, and consider three examples, showing qualitatively different parameter configurations.[7] In the figure labeled Ex-

[7] These parametrizations are meant to illustrate qualitative features, but do depend upon the specific parameter values. One important consideration in choosing these values is to ensure that the optimal second period effort is interior. We also assume that the agent is risk neutral and that the lower bound on wages is zero.

ample 1, $\gamma = \theta$, so that effort has the same effect on output in both states.[8] In this case, $\hat{e}(\mu)$ is decreasing in μ – the output benefit from incentivizing higher effort is invariant, but the principal pays a higher rent for the same incentive, as μ increases. This is depicted in the first panel of the figure. The agent's equilibrium (on-path) continuation value, $V(\mu, \mu)$, is declining in μ, since the effect on $\hat{e}$ outweighs the direct effect of μ – this is shown in the second panel. Furthermore, in this case, the agent if more optimistic than the principal when he shirks after either signal realization, and thus the third term in equation 11 is absent (Y^U is empty). Thus the ratchet effect here is stronger than in 4. Accordingly, we find that $\left.\frac{\partial W(e,e^*)}{\partial e}\right|_{e=e^*}$ is negative for all values of e^*, as is shown in the third panel. Thus the ratchet effect works to make it costlier to incentivize effort in the first period, much as in the traditional analysis.

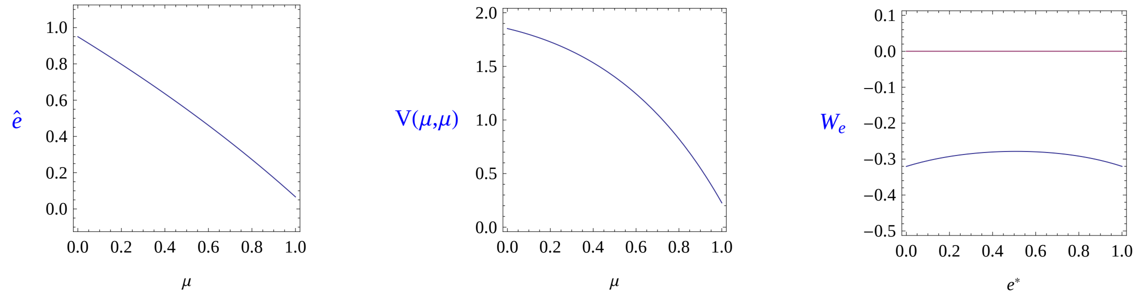


Example 1: $\gamma = \theta$.

In Example 2, $\theta > \gamma$.[9] This has the property that effort has a higher effect on output in the bad state. This reinforces the principal's incentive to depress effort as μ increases, so that $\hat{e}(\mu)$ falls more rapidly than in the first case, as shown in panel 1 of the figure. The agent's continuation value is decreasing in μ, as shown in the second panel. The ratchet effect is accentuated, i.e. $\left.\frac{\partial W(e,e^*)}{\partial e}\right|_{e=e^*}$ is negative, and larger in absolute value as compared to Example 1.

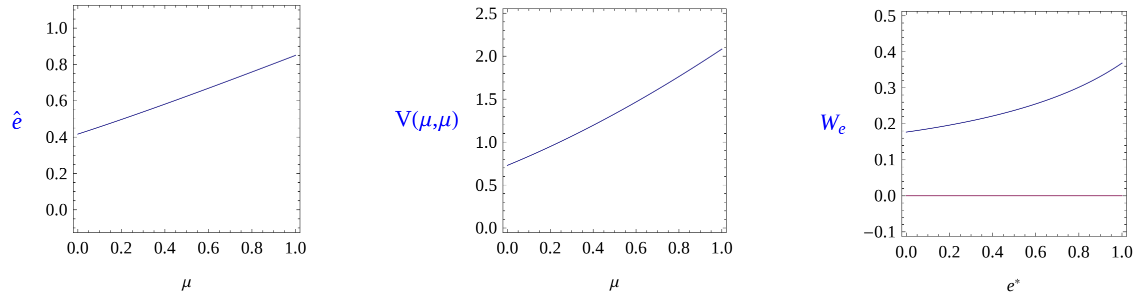
[8] The parameters for example 1 are: $p = 0.9, q = 0.2, \gamma = \theta = 0.3, y^H = 12, y^L = 0, \phi = 2$.

[9] Parameters for example 2 are: $p = 0.8, q = 0.2, \gamma = 0.2, \theta = 0.3, y^H = 12, y^L = 0, \phi = 2$.



Example 2: $\gamma < \theta$.

In Example 3, $\gamma > \theta$, so that effort has a larger effect on output in the state G rather than state B.[10] Now $\hat{e}(\mu)$ is increasing (even though a higher μ means that he pays more rents, and this depresses u^H). This is depicted in the first panel. 3. Consequently, $V(\mu, \mu)$ increases quite rapidly with μ, as shown in panel 2. In consequence, $\left.\frac{\partial W(e,e^*)}{\partial e}\right|_{e=e^*}$ is positive for all values of e^*, and increases with e^*,as is shown in the third panel. This contradicts the traditional analysis of the ratchet effect, since dynamic considerations make it cheaper to incentivize effort in the first period, as compared to the static model. Qualitatively, the results here are closer to those in a career concerns model, where the agent has an incentive to work rather than shirk. This example is particularly interesting, since a qualitatively similar parametrization has been popular in the literature (e.g. Bergemann and Hege (1998, 2005), Horner and Samuelson (2009) and Kwon (2012)). These papers find that dynamic considerations aggravate the incentive problem, whereas we find the opposite. One key difference is that in Bergemann-Hege and Horner-Samuelson, the project ends with the first success, which may be a reason why the incentive to over-work does not arise.



Example 3: $\gamma > \theta$.

[10] Parameters for example 3 are: $p = 0.9, q = 0.2, \gamma = 0.4, \theta = 0.2, y^H = 10, y^L = 0, \phi = 2$.

Table 2 summarizes our numerical results. The first row shows the sign of the ratchet effect, which takes its usual form in the first two examples, with $W_e < 0$. However in the third example, $W_e > 0$, so that the agent increases his continuation value by over-working. Turning to first period effort choice by the principal, we assume that she has a zero discount factor, in order to abstract from her experimentation motive. The second row shows the optimal first period effort induced by the principal when the agent has discount factor $\delta = 1$. The third row is a benchmark, and shows first period effort induced by the principal in the absence of the ratchet effect – this corresponds to the case where the agent is myopic and has $\delta = 0$. The ratchet effect reduces first period effort in the first two parametrizations, but raises it in the third. The last row shows the maximum value of the second derivative of the agent's continuation value $W(e, e^*)$, over all (e, e^*). This is very small relative to c_{ee}, which equals 2, and so the agent's first period payoffs are always globally concave in effort.

Table 2: First Period Effort Choice by Principal

	$\gamma = \theta$	$\gamma < \theta$	$\gamma > \theta$
ratchet effect, W_e	$-$ve	$-$ve	+ve
e^*, patient agent	0.69	0.48	0.81
e^*, myopic agent	0.76	0.55	0.63
$\max W_{ee}(.)$	0	0.001	0.02

3.2 Private Information

Suppose that the agent has unlimited liability, but assume that the agent's reservation value in the second period is random, and is revealed to the agent privately at the beginning of period 2. At $t = 1$, the agent views his future participation as random, and therefore continuous, as a function of his private belief π. Thus $V(\pi, \mu)$ is differentiable in π, ensuring differentiability of $W(e, e^*)$. We now set out the details of this approach.

Assume that the agent's reservation utility in the final period, v, is a random variable that has distribution function F and density f. Given that the principal has belief μ that the project is good, he chooses a utility level for the agent from the contract, $\bar{v}$, and $\hat{e}$ and $u = \left(u^k\right)_{k=1}^{K}$ to maximize $F(\bar{v})\mathbf{E}_{\hat{e},\mu}((y - w(u)))$, subject to the constraints $p_{\hat{e}\mu}.u - c(\hat{e}) \geq \bar{v}$, $\left(p_{1\mu} - p_{0\mu}\right).u = c_e(\hat{e})$.

We solve this problem in two steps. First, for a utility, $\tilde{v}$, that the principal provides, he can compute the optimal contract – this is the standard solution to moral hazard agency problem. Let $\Pi(\tilde{v}, \mu)$ denote the principal's profit when he provides $\tilde{v}$, conditional on the agent accepting the contract (i.e. conditional on $v \leq \tilde{v}$). $\Pi(\tilde{v})$ is decreasing, since the Lagrange multiplier on the IR constraint is strictly positive in the standard problem. Now the principal can choose $\tilde{v}$ to maximize $\Pi(\tilde{v}, \mu)F(\tilde{v})$.

Let $\bar{v}(\mu)$ denote the solution to this problem. We assume that F is sufficiently dispersed so that $\bar{v}(\mu)$ lies in the interior of the support of F. For future reference, note that if $\Pi(\tilde{v}, \mu)$ is increasing in μ, then the principal will induce greater participation, so that $\bar{v}(\mu)$ will be increasing in μ.

Suppose now that the agent has belief π different from μ. His payoff from accepting the contract equals $\hat{V}(\pi, \mu)$, (recall that $\hat{V}$ is the payoff conditional on accepting the job). He will accept the contract if $\hat{V}(\pi, \mu) \geq v$. His ex ante expected payoff, given optimal acceptance, is given by

$$V(\pi, \mu) = F\left(\hat{V}(\pi, \mu)\right)\hat{V}(\pi, \mu) + \int_{\hat{V}(\pi,\mu)} v f(v) dv. \tag{12}$$

Lemma 3 established that $\hat{V}(\pi, \mu)$ is differentiable. If F is continuous at $\hat{V}(\pi, \mu)$, then the left and right hand derivatives of $V(\pi, \mu)$ are equal and so

$$\frac{\partial V(\pi, \mu)}{\partial \pi} = F\left(\hat{V}(\pi, \mu)\right)\frac{\partial \hat{V}(\pi, \mu)}{\partial \pi}.$$

We conclude that V is differentiable at π as long as $F(.)$ is continuous at $\hat{V}(\pi, \mu)$, and so $W(e, e^*)$ is differentiable at $e = e^*$, with derivative

$$\left.\frac{\partial W(e, e^*)}{\partial e}\right|_{e=e^*} = \sum_{y^k \in Y} \frac{\partial p^k_{e\lambda}}{\partial e} V(\pi^k_{e^*}, \mu^k_{e^*}) + \sum_{y^k \in Y} p^k_{e^*\lambda} F\left(\bar{v}(\mu^k_{e^*})\right) \left.\frac{\partial \hat{V}(\pi^k_e, \mu^k_{e^*})}{\partial \pi^k_e}\right|_{\pi^k_e = \mu^k_{e^*}} \left.\frac{\partial \pi^k_e}{\partial e}\right|_{e=e^*}. \tag{13}$$

Thus there exists u such that the first order condition for implementing e^* are satisfied. The second derivative is

$$\frac{\partial^2 W(e, e^*)}{\partial e^2} = [\lambda(1-\lambda)]^2 \sum_{y^k \in Y} \frac{\partial^2 V(\pi^k_e, \mu^k_{e^*})}{\partial \pi^2} \frac{\left[p^k_{0G} p^k_{1B} - p^k_{0B} p^k_{1G}\right]^2}{\left(p^k_{e\lambda}\right)^3},$$

$$\frac{\partial^2 V(\pi^k_e, \mu^k_{e^*})}{\partial \pi^2} = F(\hat{V}(\pi, \mu))\frac{\partial^2 \hat{V}(\pi, \mu)}{\partial \pi^2} + f\left(\hat{V}(\pi, \mu)\right)\frac{\partial \hat{V}(\pi, \mu)}{\partial \pi}.$$

Both the terms in the above expression are positive. As compared to limited liability, now one has the additional second term, and to ensure that this is not too large, the distribution of reservation values must be sufficiently dispersed – otherwise, if $f(.)$ is large, W will be too convex. Conversely, if $f(.)$ is small, and if the cost function $c(e)$ is sufficiently convex, then the overall first period payoff function will be concave in effort, and one can employ the first order approach. We do not examine this problem in detail, noting only that the uncertainty regarding future reservation values must be large enough, i.e. a small perturbation of the underlying model will not suffice to ensure implementability.

How does the ratchet effect operate in this setting, as compared to the limited liability model? To answer this, compare the derivative of W under private information in 13 with that in 11, the expression under limited liability. The expressions are similar, but the agent's rents vary systematically with μ, in a way that is very different in the two cases. In particular, $V(\mu,\mu)$ is increasing in μ, since the principal optimally induces a higher degree of participation when the project is more profitable, so that $\bar{v}(\mu)$ will be increasing. (In contrast, under limited liability, $V(\mu,\mu)$ is often decreasing in μ). Since μ_e^k is larger for higher signals (those with a smaller likelihood ratio $\frac{p_{0\lambda}^k}{p_{1\lambda}^k}$), and since shirking reduces the probability of these signals, this term is also positive. In consequence, the incentives to shirk are also muted, due to the two countervailing effects that are not present when the agent's reservation utility is fixed.

3.3 Commitment

Our analysis highlights the importance of long term commitment – in its absence, the ratchet effect has very serious consequences. Milgrom and Roberts (1990) emphasize the role of commitment and discuss the Harvard Business School case study of Lincoln Electric. Lincoln Electric used piece rates, and had a policy of not revising the rate in the light of worker performance. This would provide the worker with rents when the job turned out to be a good one. However, if the job turns out to be a bad one, then the piece rate would not meet the worker's reservation utility, unless it was set very generously in the first instance. Thus, a policy of not revising piece rates requires either long term commitments on the worker's part as well, or else the firm to pay the workers rents ex ante, by meeting his IR constraint even in the bad state.

We now examine more formally the possibility of long-term commitment on both sides, and show that the stark impossibility result does not arise. Suppose that principal and agent commit to a two-period contract at the beginning of period one. The principal offers a contract $u := (u_1, u_2)$ where $u_1 \in \mathbb{R}^K$ and $u_2 : Y \to \mathbb{R}^K$. u_1 specifies the first period utility payments as a function of the first period signal realization y_1, and u_2 specifies second period payments as a function of (y_1, y_2), the pair of signal realizations across both periods. Let u_{2j} denote the j-th element of u_2, i.e. the vector of utilities that are promised after y^j is realized at $t = 1$.

Let $\mathbf{e} \in [0,1]^{K+1}$ denote a profile of efforts. The first component, e^*, denotes effort in period one, and the remaining K components, with generic element e^j, denote the effort in period 2 after signal y^j is realized at $t = 1$. For the contract u to induce $\mathbf{e}$, the agent's incentive constraints must be satisfied. In addition, an overall individual rationality constraint must be satisfied, so that his expected discounted payoff under the contract exceeds the discounted value of his outside option over the two periods. Fix e^*, and a first period signal realization y^j. e^j is implementable if and only if

$$\left(p_{1\mu^j_{e^*}} - p_{0\mu^j_{e^*}}\right).u_{2j} = c_e(e^j).$$

As before, for any $\mu^j_{e^*}$ and any e^j, we can find a u_{2j} that solves this equation. Let $V(\pi, \mu)$ denote the agent's continuation value when he has private belief π and the principal has public belief μ. Now at $\mu = \mu^j_{e^*}$ and given u_{2j},

$$V(\mu^j_{e^*}, \mu^j_{e^*}) = p_{e^j\mu^j_{e^*}}.u_{2j} - c(e^j).$$

Suppose that the agent has different beliefs π^j_e, since his first period effort choice was $e \neq e^*$. His continuation payoff is

$$V(\pi^j_e, \mu^j_{e^*}) = \max_{\tilde{e}} \left[p_{\tilde{e}\pi^j_e}.u_{2j} - c(\tilde{e})\right].$$

Since the agent is committed to the contract, there is no non-negativity constraint on his value function, and he gets $V(\pi^j_e, \mu^j_{e^*})$, for any π^j_e that he may have. His expected continuation value, as a function of his first period effort is $W(e, e^*) = \sum_Y p^k_{e\lambda} V(\pi^k_e, \mu^k_{e^*})$.

As we have already established, this is differentiable in e, and the derivative, evaluated at $e = e^*$, takes the same form as under private information, as set out in equation (13). The first order condition for implementing e^*, given any profile of second period efforts, $(e^j)_{j=1}^K$ and second period utilities, u_2, is satisfied if

$$(p_{1\lambda} - p_{0\lambda}) .u_1 + \left.\frac{\partial W(e, e^*)}{\partial e}\right|_{e=e^*} = c_e(e^*).$$

Again, there exists u_1 such that the first order conditions are satisfied.

The principal's overall problem is to choose $\mathbf{e}$ and u to maximize his profits, subject to the $K + 1$ first order conditions for the agent's incentive constraints, and the agent's overall individual rationality constraint. That is, one can apply the first-order approach to solve the principal's maximization problem. At the solution, we need to verify that the agent's second order condition is satisfied, i.e. he cannot benefit by making a large deviation from his e^*, his first period effort. As we have already seen, a sufficient condition is that $W(e, e^*)$ is not too convex relative to the convexity of the $c(e)$. We conjecture that if the agent's cost function is sufficiently convex, the agent's payoff will be globally concave in his first period effort, and thus the first order approach will be valid.

If both principal and agent can commit to a long term contract, the implementability problem does not arise. The properties of the optimal long term contract are of interest, but would take us away from the focus on the present paper. In general, the optimal contract will not be repetition of the static contract, but will condition non-trivially upon the first period signal, for two reasons. First, as public beliefs $\mu_{e^*}^j$ vary with the signal, the agent's incentive constraint will vary, and the contract must necessarily adjust. Second, optimal consumption smoothing between principal and agent implies the Lambert (1983) and Rogerson (1985) conditions on the inverses of the marginal utilities, and so continuation utilities must be history dependent.

More limited forms of commitment may also play a role, if full commitment is impossible. Suppose that the firm adopts a policy of paying Δ more than the worker's outside option – Bewley (1999) presents survey evidence showing that firms often adopt such policies. Now suppose that the worker deviates in the first period to $e < e^*$. If he becomes more pessimistic after signal y^k, and his payoff is between 0 and Δ, he will stay on the job. Since the agent will no longer quit if his private belief π is just a little below the public belief μ, $W(e, e^*)$ decreases as e is increased beyond e^*, and is smooth at e^*. Since $W(e, e^*)$ is decreasing and differentiable at $e = e^*$, the first order conditions for implementing e^* can be satisfied. Now the non differentiability is now at some $\tilde{e} >> e^*$. [11]

An alternative way to mitigate the ratchet effect is by replacing or rotating the worker. This may be costly if there is learning on the job, and so a degree of commitment is required to implement such a policy.[12]

To summarize, one interpretation of our impossibility result in theorem 7 is that it illustrates the importance of commitment. Full commitment contracts may be unrealistic, but even limited forms of commitment, such as rent-sharing or job-rotation may alleviate the problem.

4 Combining Discrete & Continuous Choices

We now modify the limited liability model, where participation is a non-issue, so that the agent has discrete choices in the second period, and continuous choices in the first period. We find that the non-implementability problem recurs. We believe that this is general feature of models that combine continuous and discrete choices that are subject to incentive constraints, but in the interests of economy, we consider the minimal model in this class. At $t = 1$, the agent chooses effort from $[0, 1]$, incurring a cost $c(e)$ that is strictly convex. At $t = 2$, the agent chooses from $\{0, 1\}$, with costs $c_1 > c_0$.[13] Other than the restriction to binary choices at $t = 2$, we assume that the information structure is as in the rest of the paper (i.e. the distribution $(p_{e\omega}), e\omega \in \{0, 1\} \times \{G, B\}$ is fixed across the two periods, and that the chosen state ω is also fixed across periods).

Our focus is on the conditions under which an interior effort level is implementable at $t = 1$. We begin the analysis in period 2, where there is some common belief μ. Assume that principal optimally induces $e = 1$ at $t = 2$, at this belief, and offers

[11] One has to therefore verify that choosing e^* is globally optimal, and this will only be true if Δ is large enough. Large upward deviations may be unprofitable if the cost of effort function $c(e)$ is sufficiently convex.

[12] There are many real-life examples of multi-plant firms that implement job rotation policies, with managers being given a maximum tenure at any single plant – for example, the Tata group of companies has a policy of rotating managers of its tea-plantations every three years. Ickes and Samuelson (1987) examine the role of job transfers in mitigating the ratchet effect arising from ex ante private information, assuming that the firm can commit to a transfer policy.

[13] We conjecture that similar results would hold in a multi-tasking model where the agent chooses from $[0, 1] \times \{0, 1\}$ in each period. However, a pre-requisite for analyzing such a model is the analysis of the static multi-tasking model that combines discrete and continuous choices, and that would take us far from the focus of the present paper.

a utility vector u. Since the incentive constraint at $t = 2$ must bind,[14] we have $(p_{1\mu} - p_{0\mu}).u = c_1 - c_0$, or

$$[\mu(p_{1G} - p_{0G}) + (1-\mu)(p_{1B} - p_{0B})]\,.u = c_1 - c_0.$$

If the agent has private belief π, he will find it optimal to choose $e = 1$ if

$$[\pi(p_{1G} - p_{0G}) + (1-\pi)(p_{1B} - p_{0B})]\,.u \geq c_1 - c_0.$$

The difference between the left-hand sides of the above two expressions equals

$$(\pi - \mu)(p_{1G} + p_{0B} - p_{1B} - p_{0G}).u = (\pi - \mu)\,(\rho.u)\,,$$

where $\rho := (p_{1G} + p_{0B} - p_{1B} - p_{0G})$, as before.

Our analysis depends upon the sign of $\rho.u$. Suppose $\rho.u > 0$. Then if $\pi > \mu$, it is optimal to choose $e = 1$, and if $\pi < \mu$, it is optimal to pick $e = 0$. The agent's expected utility may be written as a function of beliefs (π, μ) and can be written as

$$V(\pi,\mu) = V(\mu,\mu) + \begin{cases} (\pi - \mu)\,[p_{1G} - p_{1B}]\,.u \text{ if } \pi \geq \mu \\ (\pi - \mu)\,[p_{0G} - p_{0B}]\,.u \text{ if } \pi < \mu. \end{cases}$$

V is therefore a piecewise linear function of π, with slope V_π^- at $\pi < \mu$ and $V_\pi^+(u)$ at $\pi > \mu$. The difference between these slopes is $V_\pi^+ - V_\pi^- = \rho.u > 0$, since we assumed $\rho.u > 0$.On the other hand, if $\rho.u < 0$, the agent's optimal action is to choose $e = 0$ if $\pi > \mu$, and $e = 1$ if $\pi < \mu$. In this case it can be verified that $V_\pi^+ - V_\pi^- = -\rho.u > 0$. Finally, if $\Delta(u) = 0$, both actions are optimal at every belief and $V(\pi,\mu)$ is linear in π. We conclude therefore that, in general,

$$V_\pi^+ - V_\pi^- = |\rho.u|\,.$$

We now turn to the derivatives of the expected continuation value function, $W(e, e^*)$ at $e = e^*$. The left hand derivative is

[14] The incentive constraint must bind since otherwise the principal can reduce his costs by reducing payments when they exceed the limited liability constraint.

$$\begin{aligned}\left.\frac{\partial W^{-}(e,e^*)}{\partial e}\right|_{e=e^*} &= \sum_{y^k\in Y} \left.\frac{\partial p^k_{e\lambda}}{\partial e}\right|_{e=e^*} V(\pi^k_{e^*},\mu^k_{e^*}) + \sum_{y^k\in Y^D} p^k_{e^*\lambda} V^{+}_{\pi}(u(\mu^k_{e^*}))\left.\frac{\partial \pi^k_e}{\partial e}\right|_{e=e^*} \\ &\quad + \sum_{y^k\in Y^U} p^k_{e^*\lambda} V^{-}_{\pi}(u(\mu^k_{e^*}))\left.\frac{\partial \pi^k_e}{\partial e}\right|_{e=e^*}.\end{aligned}$$

The right hand derivative is

$$\begin{aligned}\left.\frac{\partial W^{+}(e,e^*)}{\partial e}\right|_{e=e^*} &= \sum_{y^k\in Y} \left.\frac{\partial p^k_{e\lambda}}{\partial e}\right|_{e=e^*} V(\pi^k_{e^*},\mu^k_{e^*}) + \sum_{y^k\in Y^D} V^{-}_{\pi}(u(\mu^k_{e^*}))p^k_{e^*\lambda}\left.\frac{\partial \pi^k_e}{\partial e}\right|_{e=e^*} \\ &\quad + \sum_{y^k\in Y^U} p^k_{e^*\lambda} V^{+}_{\pi}(u(\mu^k_{e^*}))\left.\frac{\partial \pi^k_e}{\partial e}\right|_{e=e^*}.\end{aligned}$$

The difference between the right-hand and left-hand derivatives equals

$$\sum_{y^k\in Y^D} -\left|\rho.u(\mu^k_{e^*}))\right| p^k_{e^*\lambda}\left.\frac{\partial \pi^k_e}{\partial e}\right|_{e=e^*} + \sum_{y^k\in Y^U} \left|\rho.u(\mu^k_{e^*})\right| p^k_{e^*\lambda}\left.\frac{\partial \pi^k_e}{\partial e}\right|_{e=e^*}. \tag{14}$$

Since $\frac{\partial \pi^k_e}{\partial e} < 0$ if $y^k \in Y^D$ and $\frac{\partial \pi^k_e}{\partial e} > 0$ if $y^k \in Y^U$, we conclude that each term in each of the above summations is positive. Thus $\left.\frac{\partial W^{+}(e,e^*)}{\partial e}\right|_{e=e^*} > \left.\frac{\partial W^{-}(e,e^*)}{\partial e}\right|_{e=e^*}$, with the inequality being strict unless $\rho.u(\mu^k_{e^*}) = 0$ for all $\mu^k_{e^*}$.

Notice that this argument does not depend upon the principal inducing $e = 1$ at $t = 2$ after every signal realization. It suffices that there exists at least one signal $y^k \in Y^U \cup Y^D$ such that at the belief $\mu^k_{e^*}$, inducing $e = 1$ is optimal. This is sufficient to ensure the inequality $\left.\frac{\partial W^{+}(e,e^*)}{\partial e}\right|_{e=e^*} > \left.\frac{\partial W^{-}(e,e^*)}{\partial e}\right|_{e=e^*}$ as long as $\rho.u(\mu^k_{e^*}) \neq 0$.

We now show that for generic information structures, $\rho.u(\mu) \neq 0$. Consider first binary signals, with the information structure set out in Table 1. Under limited liability, if the IR constraint is irrelevant and the limited liability constraint binds, $u^L = \underline{u}$, and the incentive constraint given belief μ implies

$$\left[\mu\theta + (1-\mu)\gamma\right] u^H = c_1 - c_0,$$

so that $u = (\underline{u} + \frac{c_1 - c_0}{\mu\theta + (1-\mu)\gamma}, \underline{u})$. Since $\rho = (\theta - \gamma, \gamma - \theta)$, the inner product is

$$\rho.u(\mu) = \frac{(\gamma - \theta)(c_1 - c_0)}{\mu\theta + (1-\mu)\gamma}.$$

We conclude that unless $\theta = \gamma$, $\rho.u(\mu) \neq 0$ for every μ.

Consider next the case where the agent is risk neutral and there are many signals. By lemma 12 in the appendix, utility payments exceed $\underline{u}$ only after the signal or signals that have the lowest likelihood ratio, $\frac{p^k_{0\mu}}{p^k_{1\mu}}$. Thus the principal makes a binary partition of the set of signals, and pays $\underline{u}$ after all except the highest signal. This reduces to the case of binary signals, and therefore generically, $\rho.u(\mu) \neq 0$ for every μ. The proof for risk-aversion and many signals follows the statement of the theorem.

Theorem 10 *Consider a limited liability model where the agent has a binary effort choice in the second period, and must choose from* $[0, 1]$ *in the first period. Suppose that the principal finds it optimal to induce high effort in the second period after at least one signal realization in* $Y^U \cup Y^D$. *Then no interior effort level is implementable in the first period for generic information structures.*

Proof. Assume that the agent is risk averse, so that the inverse utility function $w(u)$ is strictly convex. By lemma 12 in the appendix, for any k,

$$w'(u^k) = \max\left\{ w'(\underline{u}), \beta\left(\frac{p^k_{1\mu} - p^k_{0\mu}}{p^k_{\hat{e}\mu}}\right)\right\},$$

where β is the Lagrange multiplier on the incentive constraint. Let $\xi : \mathbb{R}^+ \to \mathbb{R}$ be the map from $w'(u)$ to u.[15] Since ξ is strictly increasing, and arranging signals in the order of the likelihood ratio, the cost minimizing utility vector u takes the form

$$u = \left(\underline{u}, .., \underline{u}, \xi\left(\beta\frac{p^k_{1\mu} - p^k_{0\mu}}{p^k_{\hat{e}\mu}}\right), .., \xi\left(\beta\frac{p^K_{1\mu} - p^K_{0\mu}}{p^k_{\hat{e}\mu}}\right)\right).$$

Let p denote the information structure, i.e. the vector $(p^k_{e\omega})$ where $k \in \{1, 2., , .K\}, e \in \{0, 1\}, \omega \in \{G, B\}$. This lies in the subset of the $\left(\Delta^{K-1}\right)^4$ that satisfies assumption 1 (Δ^{K-1} is the $K-1$ dimensional simplex). The equation $\rho.u = 0$ can only be satisfied on a set of points p that are of Lebesgue measure zero, and thus, $\rho.u \neq 0$ for almost all information structures p. ■

The impossibility result generalizes to the case of a finite set of efforts at $t = 2$, with $c(e)$ increasing. Assume that the principal induces some effort $\hat{e} > 0$ (the smallest effort) after some first period signal y^k. The incentive constraint must bind; under

[15] The map ξ is standard in contract theory with risk aversion – see e.g. Jewitt (1988).

the optimal contract, at the belief $\mu^k_{e^*}$, the agent must be indifferent between $\hat{e}$ and his next best choice. If the agent's beliefs π^k_e differ, then his optimal action when he is more optimistic will differ from his optimal choice when pessimistic. This ensures that the continuation value function will be kinked as a function of his first period effort, and prevents the first order conditions from being satisfied. On the other hand, if the agent only has continuous choices at $t = 2$, then the standard envelope theorem applies, and his continuation value function is smooth.

In the light of our findings, we may now discuss the recent literature on agency models with experimentation. Bergemann and Hege (1998, 2005), Horner and Samuelson (2009), Manso (2011) and Kwon (2011) analyze models where agent has a discrete set of actions (usually binary). De Marzo and Sannikov (2011), Cisternas (2012) and Jovanovic and Prat (2013) analyze continuous time models, where the agent has continuation action choices, and where participation constraints are irrelevant. Either of these formulations do not encounter the problems that arise here, since it is the combination of continuous and discrete actions that give rise to the difficulty.

We now show that the impossibility result can be overcome if the principal commits to the two-period contract in the first period. As in section 3.3, suppose that the principal offers a contract $u := (u_1, u_2)$ where $u_1 \in \mathbb{R}^K$ and $u_2 : Y \to \mathbb{R}^K$. Let $\mathbf{e} \in [0,1] \times \{0,1\}^K$ denote a profile of efforts, where $e^* \in [0,1]$ is the first period effort level, with $e^j \in \{0,1\}$ denoting an effort level after signal y^j at $t = 2$.

Theorem 11 *If the principal can commit to a two period contract at* $t = 1$, *then any* $\mathbf{e} \in [0,1] \times \{0,1\}^K$ *is implementable.*

Proof. We construct second period payments so that the agent's second period continuation value after signal realization y^j, $V(\pi, \mu^j_{e^*})$ is differentiable (and linear) in π for all relevant values of π. To reduce notation, let μ to denote $\mu^j_{e^*}$, the principal's belief after y^j. If $e^j = 0$, the principal can set u_{2j} to be constant and in this case the agent's continuation value, $V(\pi, \mu)$ does not vary with π after signal y^j. Suppose that $e^j = 1$. Let $\tilde{u}$ be second period payments such that $\tilde{u}.(p_{1\mu} - p_{0\mu}) = c_1 - c_0$. Suppose that $\rho.\tilde{u} > 0$, so that the incentive constraint would be violated at $\pi < \mu$ if the principal offered $\tilde{u}$. Let $\pi^j_{\min} = \min_{e\in[0,1]}\{\pi^j_e\}$ denote the most pessimistic belief that the agent could have after y^j. Now if the principal offers second period payments $\hat{u}$ such that $\hat{u}.(p_{1\pi^j_{\min}} - p_{0\pi^j_{\min}}) = c_1 - c_0$, then the agent's incentive constraint is always satisfied after signal y^j, no matter what his first period effort (when his beliefs are more optimistic than μ, it always optimal to choose $e = 1$), and he will always choose $e = 1$. Thus $V(\pi, \mu)$ is smooth and linear in π. This can be done for every signal y^k after which the principal seeks to induce $e^k = 1$, thereby ensuring that the agent's continuation value function $W(e, e^*)$ is smooth, and the first order conditions for implementing e^* can be satisfied. Since $V(\pi, \mu)$ is linear, the agent's overall payoff is strictly concave as function of his first period effort. The argument for the case where $\rho.\tilde{u} < 0$ is similar – the principal now ensures that the incentive constraint is satisfied at $\pi^j_{\max} = \max_{e \in [0,1]}\{\pi^j_e\}$. ■

Two remarks are in order here. First, the principal cannot implement e^* in the interior unless there is some slack in the incentive constraint for implementing $e^j = 1$ at the equilibrium belief $\mu^j_{e^*}$ – if the constraint binds, the agent's first order condition for e^* will be violated, as in the no-commitment case. However, the principal does not necessarily have to ensure that the agent's incentive constraint is satisfied at all possible beliefs, as in the proof, and it may be sufficient to deter smaller deviations, if the cost of effort is sufficiently convex. Second, the commitment contract is not renegotiation-proof if the agent is risk-averse. After the agent has chosen his first period effort, it is inefficient to make him bear extra risk, to satisfy an incentive constraint that is now redundant, given that he has not deviated at $t = 1$. Thus the impossibility result would recur if the contract had to be renegotiation-proof.

5 Conclusions

Dynamic agency problems with learning where the agent has continuous as well as discrete choices give rise to serious difficulties for the principal. Sequential rationality implies that the principal implements his desired discrete choice in the final period at least cost. This ensures that the agent is indifferent between the principal's desired action and his next best alternative. Any perturbation in the agent's beliefs will affect the payoffs from these two actions differently, and generically, the agent will choose one action when he is more optimistic, and a different one when he is more pessimistic. The consequent convex kink in his continuation value function makes implementing interior efforts in the initial period impossible, since one cannot guard against both upward and downward deviations. While the present paper has restricted attention to situation where these constraints are imposed by a profit-maximizing principal, these could also arise in competitive markets. Also, we have focused the analysis on the case where the uncertainty pertains to the nature on the principal's project, rather than the talent of the agent. However, we believe that the conceptual problem identified here may also arise in versions of the careers concerns model, where learning affects the outside option of the agent.

One response to the non-implementability result is to make choices continuous from the point of view of the initial period. For example, making the costs of different actions stochastic could solve the problem provided that the uncertainty is sufficiently large. In some contexts, large uncertainty may well be reasonable, but this may not appeal in all situations. A second response is to argue that these arguments illustrate the importance of the employer's ability to commit. As we have seen, the costs of not being able to commit can be large.

References

[1] Bergemann, D., and U. Hege, 1998, Venture capital financing, moral hazard and learning, *J. Banking and Finance*, 22, 703-735.

[2] Bergemann, D., and U. Hege, 2005, The financing of innovation: Learning and stopping, *Rand J. Econ.*, 78, 737-787.

[3] Bewley, T.F., 1999, *Why Wages don't Fall in a Recession,* Cambridge MA, Harvard University Press.

[4] Bhaskar, V., and G. Mailath, 2014, The curse of long horizons, in preparation.

[5] Carmichael, H.L., and W.B. MacLeod, 2000, Worker cooperation and the ratchet effect, *J. Labor Econ.*, 18(1),1-19.

[6] Cisternas, G., 2013, Two-sided learning and moral hazard, mimeo, MIT Sloan.

[7] De Marzo, P., and Y. Sannikov, 2011, Learning, termination and payout policy in dynamic incentive contracts, mimeo.

[8] Dewatripont, M., I. Jewitt and J. Tirole, 1999, The economics of career concerns, part I: Comparing information structures, *Rev. Econ. Studies* 66, 183-198.

[9] Edwards, R., 1979, *Contested Terrain: The Transformation of the Workplace in the Twentieth Century*, Basic Books.

[10] Freixas, X., R. Guesnerie and J. Tirole, 1985, Planning under incomplete information and the ratchet effect, *Rev. Econ. Studies* 52 (2), 173-191.

[11] Gibbons, R., 1986, Piece-rate incentive schemes, *J. Labor Econ.*, 5(4),413-429.

[12] Gibbons, R., and K. G. Murphy, 1992, Optimal incentive contracts in the presence of career concerns: Theory and evidence, *J. Polit. Economy* 100, 468-505.

[13] Ickes, B. W., and L. Samuelson, 1987, Job transfers and incentives in complex organizations: thwarting the ratchet effect, *Rand J. Econ.* 275-286.

[14] Holmstrom, B., 1982, 1999, Managerial incentive problems: A dynamic perspective, *Rev. Econ. Studies* 66, 169-182.

[15] Holmstrom, B., and O. Hart, 1987, The theory of contracts, in *Advances in Economic Theory, Fifth World Congress*, Cambridge University Press, Cambridge.

[16] Horner, J., and L. Samuelson, 2009, Incentives for experimenting agents, mimeo.

[17] Innes, R. D., 1990, Limited liability and incentive contracting with ex-ante action choices, *Journal of Economic Theory,* 52 (1), 45-67.

[18] Jewitt, I., 1988, Justifying the first-order approach to principal-agent problems. *Econometrica*, 58, 1177-1190.

[19] Jewitt, I., O. Kadan, and J.M. Swinkels, Moral hazard with bounded payments, *Journal of Economic Theory*, 143(1), 59-82.

[20] Jovanovic, B., and J. Prat, 2013, Dynamic contracts when agent's quality is unknown, *Theoretical Economics.*

[21] Kwon, S., 2011, Dynamic moral hazard with persistent states, mimeo, UCL.

[22] Laffont, J-J., and J. Tirole, 1988, The dynamics of incentive contracts, *Econometrica* 58, 1279-1319.

[23] Laffont, J-J., and J. Tirole, 1993. *A Theory of Incentives in Procurement and Regulation.* MIT Press.

[24] Lambert, R. A., 1983, Long-term contracts and moral hazard. *The Bell Journal of Economics,* 441-452.

[25] Lazear, E., 1986, Salaries and piece rates, *J. Business.*, 59(3),405-431.

[26] Malcomson, J., 2013, Relational incentive contracts with persistent private information, mimeo, Oxford.

[27] Manso, G., 2011, Motivating innovation, *J. Finance,* 66, 1823-1869.

[28] Mathewson, S.B., 1931, *Restriction of Output among Unorganized Workers*, Southern Illinois University Press.

[29] Meyer, M., and J. Vickers, 1997, Performance comparisons and dynamic incentives, *J. Polit. Econ.* 105.

[30] Milgrom, P., and J. Roberts, 1992, *Economics, Organization and Management*, Prentice Hall.

[31] Milgrom, P., and I. Segal, 2002, Envelope theorems for arbitrary choice sets, *Econometrica* 70 (2) , 583–601.

[32] Rogerson, W., 1985, Repeated moral hazard, *Econometrica*, 53(1), 69-76.

[33] Roy, D., 1952, Output restriction and goldbricking in a machine shop, *Amer. J. Sociology,* 57 (5), 427-442.

6 Appendix

We first prove the following lemma, that is standard (see e.g. Jewitt, Kadan and Swinkels, 2008), but set out here for completeness.

Lemma 12 *Consider the static contracting problem under a binding limited liability constraint, and the cost-minimizing contract that implements* $\hat{e} > 0$, *where the set of possible efforts is either* $[0,1]$ *or* $\{0,1\}$. *If the agent is risk-neutral, the principal makes a binary partition of the signal space; he pays* $u^K > \underline{u}$ *after the signal (or signals) with the highest likelihood ratio* $\frac{p_{1\mu}^k - p_{0\mu}^k}{p_{\hat{e}\mu}^k}$, *and* $\underline{u}$ *after all other signals. If the agent is risk averse, the principal pays* $\underline{u}$ *for all signals* y^k *with likelihood ratio* $\frac{p_{1\mu}^k - p_{0\mu}^k}{p_{\hat{e}\mu}^k}$ *below a critical threshold, and pays* $u^k > \underline{u}$ *for signals with a likelihood ratio above this threshold, where* u^k *is increasing in* $\frac{p_{1\mu}^k - p_{0\mu}^k}{p_{\hat{e}\mu}^k}$.

Proof. Since limited liability constraints bind, the participation constraints do not, and the Lagrangian for the principal's cost minimization problem when e is chosen from $[0,1]$ is

$$\mathcal{L} = -\sum_k p_{\hat{e}\mu}^k w(u^k) + \beta \left(\sum_k \left(p_{1\mu}^k - p_{0\mu}^k\right) u^k - c_e(\hat{e}) \right),$$

where β the multiplier on the incentive constraint. The principal maximizes with respect to $\left(u^k\right)_{k=1}^K$ and β, subject to the limited-liability constraints $u^k \geq \underline{u} \forall k$. In the case of discrete effort, the expression is identical, except that $c_e(\hat{e})$ is replaced by $c_1 - c_0$. In either case, the derivative of the Lagrangian with respect to u^k is

$$\frac{\partial \mathcal{L}(\mu)}{\partial u^k} = -p_{\hat{e}\mu}^k w'(u^k) + \beta \left(p_{1\mu}^k - p_{0\mu}^k\right). \tag{15}$$

$\beta > 0$ since the incentive constraint binds. So the derivative is negative if $\left(p_{1\mu}^k - p_{0\mu}^k\right) \leq 0$ (i.e. if $y^k \notin Y^H$), which implies $u^k = \underline{u}$. If the agent is risk neutral, $w'(.)$ is constant, and the derivative can equal zero only for one value of the likelihood ratio, $\frac{p_{1\mu}^k - p_{0\mu}^k}{p_{\hat{e}\mu}^k}$, the largest one among all Y. So if the agent is risk neutral, the principal makes a binary partition of the state space, and pays the agent $\underline{u}$ after all signals except those with the highest likelihood ratio.

When the agent is risk averse, it is still the case that $u^k = \underline{u}$ if $y^k \notin Y^H$. For signals in Y^H, the first order condition (15) implies

$$w'(u^k) = \max\left\{w'(\underline{u}), \beta\left(\frac{p^k_{1\mu} - p^k_{0\mu}}{p^k_{\hat{e}\mu}}\right)\right\}.$$

That is the agent is paid an amount exceeding $\underline{u}$ after signals that have a sufficiently high likelihood ratio, $\left(\frac{p^k_{1\mu}-p^k_{0\mu}}{p^k_{\hat{e}\mu}}\right)$. ■

To prove theorem 9**,** we prove the following lemma.

Lemma 13 *Consider a static contracting problem, where Y is binary as in table 1, and where limited liability constraints bind so that u^L =$\underline{u}$. Let the cost of effort $c(e,\phi) = \frac{1}{2}\phi e^2$, and let $\hat{e}(\mu,\phi)$ denote the profit maximizing effort induced by the principal. $\hat{e}(\mu,\phi)$ and $c_e(\hat{e}(\mu,\phi))$ are decreasing in ϕ.*

Proof. The principal's maximization problem is:

$$\max_{e,u^H} p^H_{e\mu}\left[\left(y^H - y^L\right) - \left[w(u^H) - w(\underline{u}))\right]\right] + y^L - w(\underline{u}),$$

subject to the incentive constraint

$$\left(p^H_{1\mu} - p^H_{0\mu}\right)\left(u^H - \underline{u}\right) = c_e(e,\phi).$$

Define $y := y^H - y^L + w(\underline{u}))$. Substituting the incentive constraint, the principal's problem is:

$$\max_e p^H_{e\mu}\left(y - w\left(\frac{c_e(e,\phi)}{\left(p^H_{1\mu} - p^H_{0\mu}\right)} + \underline{u}\right)\right).$$

The first order condition for optimal effort is

$$\left(p^H_{1\mu} - p^H_{0\mu}\right)\left(y - w\left(\frac{c_e(e,\phi)}{\left(p^H_{1\mu} - p^H_{0\mu}\right)} + \underline{u}\right)\right) - w'\left(\frac{c_e(e,\phi)}{\left(p^H_{1\mu} - p^H_{0\mu}\right)} + \underline{u}\right)\frac{c_{ee}(e,\phi)}{\left(p^H_{1\mu} - p^H_{0\mu}\right)}\left(p^H_{0\mu} + e\left(p^H_{1\mu} - p^H_{0\mu}\right)\right) = 0$$

This can be re-written as

$$\left(p^H_{1\mu} - p^H_{0\mu}\right) w\left(\frac{c_e(e,\phi)}{\left(p^H_{1\mu} - p^H_{0\mu}\right)} + \underline{u}\right) + w'\left(\frac{c_e(e,\phi)}{\left(p^H_{1\mu} - p^H_{0\mu}\right)} + \underline{u}\right) c_{ee}(e,\phi)\left(\frac{p^H_{0\mu}}{\left(p^H_{1\mu} - p^H_{0\mu}\right)} + e\right) = K,$$

where K is a constant. Totally differentiating with respect to ϕ, the following expression equals zero:

$$w'(.)\left(c_{e\phi} + c_{ee}\frac{de}{d\phi}\right) + w'(.)\left(c_{ee\phi} + c_{eee}\frac{de}{d\phi}\right)\left(\frac{p^H_{0\mu}}{(p^H_{1\mu} - p^H_{0\mu})} + e\right) + w'(.)c_{ee}\frac{de}{d\phi}$$
$$+w''(.)\left(c_{e\phi} + c_{ee}\frac{de}{d\phi}\right)\frac{1}{(p^H_{1\mu} - p^H_{0\mu})}c_{ee}(e,\phi)\left(\frac{p^H_{0\mu}}{(p^H_{1\mu} - p^H_{0\mu})} + e\right).$$

Rearranging the above, and letting $r_A(u^H) = \frac{w''(u^H)}{w'(u^H)}$ (r_A is the Arrow-Pratt coefficient of absolute risk aversion), we get

$$\frac{d\hat{e}}{d\phi} = -\frac{c_{e\phi} + c_{ee\phi}\left[\frac{p^H_{0\mu}}{p^H_{1\mu} - p^H_{0\mu}} + \hat{e}\right] + r_A(u^H)\frac{c_{e\phi}c_{ee}}{(p^H_{1\mu} - p_{0\mu})}\left(\frac{p^H_{0\mu}}{(p^H_{1\mu} - p^H_{0\mu})} + e\right)}{2c_{ee} + c_{eee}\left[\frac{p^H_{0\mu}}{p^H_{1\mu} - p^H_{0\mu}} + \hat{e}\right] + r_A(u^H)\frac{(c_{ee})^2}{(p^H_{1\mu} - p^H_{0\mu})}\left(\frac{p^H_{0\mu}}{(p^H_{1\mu} - p^H_{0\mu})} + e\right)}$$

Note that

$$\frac{dc_e(\hat{e}(\mu,\phi))}{d\phi} = c_{e\phi} + c_{ee}\frac{d\hat{e}}{d\phi}. \tag{16}$$

Since $c_{e\phi} > 0$ and $c_{ee} > 0$, it suffices to show that $\frac{dc_e(\hat{e}(\mu,\phi))}{d\phi} \geq 0$ since this implies $\frac{d\hat{e}}{d\phi} < 0$. Expanding 16, $\frac{dc_e(\hat{e}(\mu,\phi))}{d\phi}$ can be written as the ratio of two terms, with numerator

$$\text{NUM} = c_{ee}c_{e\phi} + \left(c_{eee}c_{e\phi - c_{ee}c_{ee\phi}}\right)\left[\frac{p^H_{0\mu}}{p^H_{1\mu} - p^H_{0\mu}} + \hat{e}\right].$$

The denominator is

$$\text{DEN} = 2c_{ee} + c_{eee}\left[\frac{p^H_{0\mu}}{p^H_{1\mu} - p^H_{0\mu}} + \hat{e}\right] + r_A(u^H)\frac{(c_{ee})^2}{(p^H_{1\mu} - p^H_{0\mu})}\left(\frac{p^H_{0\mu}}{(p^H_{1\mu} - p^H_{0\mu})} + e\right). \tag{17}$$

When $c(e,\phi) = \frac{1}{2}\phi e^2$, $c_{ee} = \phi, c_{e\phi} = e, c_{eee} = 0, c_{ee\phi} = 1$, so that

$$\frac{dc_e(\hat{e})}{d\phi} = -\frac{p^H_{0\mu}}{2(p^H_{1\mu} - p^H_{0\mu}) + r_A(.)\phi\left[\frac{p^H_{0\mu}}{p^H_{1\mu} - p^H_{0\mu}} + \hat{e}\right]} < 0.$$

■

Proof. of theorem 9**:** The overall payoff in the first period is $v(e, e^*, u) = p_{e\mu}.u - c(e,\phi) + W(e, e^*; \phi)$. We show that this is strictly concave in e when $c(e,\phi) = \frac{1}{2}\phi e^2$

and ϕ is sufficiently large. The second derivative is

$$\frac{\partial^2 v}{\partial e^2} = -\phi + \frac{\partial^2 W}{\partial e^2}.$$

Note that $\frac{\partial^2 W}{\partial e^2}$ is a weighted average of the terms $\frac{\partial^2 \hat{V}(\pi_e^k, \mu_{e*}^k)}{\partial \left(\pi_e^k\right)^2}$, as established in equation 6 in the text. In the light of Lemma 12, we may restrict attention to binary signals, and in this case,

$$\frac{\partial^2 \hat{V}(\pi, \mu)}{\partial \pi^2} = \left(\frac{(\theta - \gamma) c_e(\hat{e}(\mu))}{(p_{1\mu}^H - p_{0\mu}^H)} \right)^2 \frac{1}{\phi}. \tag{18}$$

This is decreasing in ϕ, since lemma 13 establishes that the first term is decreasing. Thus v is strictly concave in e if ϕ is large enough. ■