The Ratchet Effect Re-examined: A Learning Perspective

V Bhaskar*

University College London

October 18, 2013

Abstract

We examine the ratchet effect in a situation where both principal and agent are uncertain about the difficulty of the job, and must learn this over time. Since the agent can always increase his future continuation value by shirking, this must be deterred by higher powered incentives today. However, with a continuum of effort levels, high powered incentives provide an incentive for the agent to over-work. This implies an impossibility result: *no interior effort level is implementable* today. We explore the role of rent sharing in solving the problem, and also the role of commitment and limited liability.

Keywords: ratchet effect, moral hazard, learning. JEL codes: D83, D86.

*Department of Economics, University College London, Gower Street, London WC1E 6BT, UK. Email: v.bhaskar@ucl.ac.uk

1 Introduction

The ratchet effect is one of earliest problems noted by modern incentive theory, being prominent in discussions of Soviet planning (Berliner, 1954; Bergson, 1976). If the factory met or exceeded its plan target, the target for subsequent years was increased, thereby reducing current effort incentives for the factory manager (Weitzman, 1980). The problem also arises in capitalist firms, as Milgrom and Roberts (1990) note. When a firm installs new equipment, both firms 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 studies that document the importance of "output restriction".[1] The ratchet effect is also prominent in the marketing literature. Salesmen are often paid bonuses that depend on the excess of sales over a quota, that is usually 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 the dynamic mechanism design problem without commitment, on how to induce the agent to reveal her private information. This work includes Lazear (1986), Gibbons (1986), Freixas, Rochet & Tirole (1985), and Laffont & Tirole (1986). 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 (1986) prove a general result, that one cannot induce full separation given a continuum of types. Perhaps the most comprehensive discussion of the problem is in Laffont and Tirole (1994), who consider both the case of binary types and of a continuum of types, and once we present our findings, it will be useful to relate these

[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.

to their discussion. [2]

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, a situation which has not been, to our knowledge, formally analyzed. Indeed, in Bolton and Dewatripont (2005) dynamic information revelation without commitment is seen as synonymous with the ratchet effect. This identification is, in our view, overstated. 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. Milgrom and Roberts (1990) present an illuminating (albeit somewhat informal) discussion of this problem. They assume a linear technology and normally distributed shocks, and argue that the ratchet effect implies that incentives are more high powered at the beginning, and the effort induced increases over time. While long term commitments alleviate the problem, it may be hard to stick to these commitments since they are likely to be inefficient ex post. [3]While their discussion is extremely insightful, they do not make explicit their assumptions.

Our analysis can be seen as a part of the emerging literature on dynamic moral hazard with learning/experimentation that includes Bergemann and Hege (1998, 2005), Manso (2011), Horner and Samuelson (2009), and Kwon (2011). It also relates to Holmstrom's (1982) career concern model, and the literature that develops this further. Continuous time formulations of these problems include De Marzo and Sannikov (2011), Cisternas (2012) and Prat and Jovanovic (2013). We discuss this literature in more detail in section 5.

We study optimal contracting where the firm (the principal) and the agent (the worker) are ex ante symmetrically informed and uncertain about the difficulty of the job, and learn about 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. We allow for a rather general information structure, and study the dynamic

[2] Also, Ickes and Samuelson (1987) explore the role of job transfers in mitigating the ratchet effect. Kanemoto and Macleod (1992) show that if the private information pertains to worker ability rather than job characteristics, then the ratchet effect is does not apply.

[3] They discuss at length the case of Lincoln Electric, which developed a reputation for never revising the piece rate downwards in the light of information – see also Harvard Business Case study.

moral hazard problem that arises in this learning environment. 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 learnt regarding the job.

In our context, the ratchet effect arises from the possibility that the agent can manipulate the beliefs of the principal, by shirking. Suppose that both principal and agent have a common prior λ that the job is good or easy, and bad with complementary probability. Consider a contract where the principal seeks to induce some non zero effort level e^*. If the agent chooses e^*, and some output y^k is realized, both principal and agent will update their beliefs regarding the job to $\mu^k_{e^*}$, and the second period contract minimizes expected wage payments subject the incentive and individual rationality constraints defined by $\mu^k_{e^*}$. Suppose now that the agent deviates to $e < e^*$. Since effort is privately chosen, the principal's belief is still $\mu^k_{e^*}$ after y^k, while the agent's belief is μ^k_e. Under fairly general conditions, the agent's continuation value *strictly increases* if he shirks a little relative to e^*. The intuition comes from the fact that the individual rationality constraint always binds given belief $\mu^k_{e^*}$. If the agent is more pessimistic about the job (i.e. $\mu^k_e < \mu^k_{e^*}$), then the constraint is violated, while if the agent is more optimistic (i.e. $\mu^k_e > \mu^k_{e^*}$), then the individual rationality constraint holds strictly, and the agent makes a surplus above his reservation utility. Now, when the IR constraint is violated, the agent will simply refuse the contract and earn his reservation utility, and therefore suffers no loss. Since the agent accepts the payoff gains but can refuse the payoff losses, he will benefit as long as there is some signal y^k such that $\mu^k_e > \mu^k_{e^*}$, i.e. where he is more optimistic regarding the job than the principal thinks that he is.

We show that there always exists some signal y^k such that $\mu^k_e > \mu^k_{e^*}$ as long as $e < e^*$. This follows from the martingale property of beliefs. The expectation of the agent's posterior, over all signal realizations, must equal his prior, λ, regardless of whether the agent performs the experiment e^* or the experiment e. Since good signals have higher probability under e^* than under e, this equality of expectations can only be satisfied if there are some signals such that $\mu^k_e > \mu^k_{e^*}$.

Since the agent's second period continuation value is higher when he deviates to low effort in period one, as compared to the case where he does not deviate, this implies that the incentive constraint in the first period must be modified. That is, the principal must provide greater incentives for high effort than she would need to do in a static context, where there was no second period. This argument is quite general, and holds as long as the signal structure satisfies the property that a signal that is informative of high effort is also informative of the job being easy. Thus current period incentives have to be sufficiently high powered so as to deter all such downward deviations. This is the point made by Milgrom and Roberts (1990) – our contribution, so far, is to show that this argument generalizes with rather mild assumptions on the information structure. In particular, it is possible that the agent is more pessimistic than the principal after some realizations of the output – the argument still applies.

The main contribution of the paper is to show that by deterring downward deviations, the principal makes upward deviations profitable. Let $W(e, e^*)$ denote the agent's continuation value at $t = 2$, net of his outside option, when the principal seeks to induce e^*, and he chooses $e < e^*$. The preceding arguments imply that $W(e, e^*) > 0$ for $e < e^*$. $W(e^*, e^*) = 0$ since the agent gets his reservation utlity on the equilibrium path. Also, $W(e, e^*) \geq 0$ for $e > e^*$ since the agent cannot get less than his outside option. We see therefore that there is fundamental convexity of the continuation value function – take any convex combination $\lambda e + (1 - \lambda)e' = e^*$, and we note that $\lambda W(e, e^*) + (1 - \lambda)W(e', e^*) > 0 = W(e^*, e^*)$. Since the continuation value function is strictly convex at e^*, the first order condition for implementing e^* can never be satisfied, as long as $e^* \in (0, 1)$.

One solution to this problem is to the leaving of rents to the agent. This makes the continuation value smooth at e^*. If the rents are large enough, and the cost of effort function is sufficiently convex, this alleviates the implementablity problem. We also explore limited liability models, and show that the ratchet effect is a less robust phenomenon in such a context.

2 The model

Our model combines moral hazard with uncertainty regarding job difficulty. Specifically, the job is either good (easy) or bad (hard), i.e. the job type is $\alpha \in \{G, B\}$. Let $\lambda \in (0, 1)$ denote the common prior that $\alpha = G$. The agent chooses effort $e \in \{0, 1\}$. Let $y \in Y = \{y_1, y_2, .., y_n\}$ denote the signal that is realized following effort choice. This depends, stochastically, on both the type and the effort chosen. Let $p^k_{e\alpha}$ be the probability of signal y^k given effort e and type $\alpha \in \{G, B\}$. Thus for each signal y^k, we have a 4-tuple $(p^k_{0B}, p^k_{1B}, p^k_{0G}, p^k_{1G})$. With a slight abuse of notation, we may also define $p^k_{1\mu}$ (resp. $p^k_{0\mu}$) to be the probability of signal k when effort level 1 (resp. 0) is chosen, given that μ is the probability that the agent is type G.

We shall distinguish two types of likelihood ratio, the likelihood ratio on efforts for a given type (or belief over types) and the likelihood ratio over types for a given effort choice. The former is relevant for providing effort incentives, while the latter determines Bayesian learning. Let $\ell^k_\alpha = \frac{p^k_{1\alpha}}{p^k_{0\alpha}}$ be the likelihood ratio for signal k for type α. Generalizing this, $\ell^k_\mu = \frac{\mu p^k_{1G} + (1-\mu) p^k_{1B}}{\mu p^k_{0G} + (1-\mu) p^k_{0B}}$ denote the likelihood ratio for signal k when μ is the probability that the agent is type G. Let $\ell^k_e = \frac{p^k_{eG}}{p^k_{eB}}$ be the likelihood ratio for signal k for effort level e.

Our main assumption, that is maintained throughout this paper, is as follows:

A1 All probabilities belong to $(0, 1)$. For some y^k, $p^k_{1G} \neq p^k_{0B}$ i.e. there exists some informative signal. For any informative signal y^k, p^k_{1B} and p^k_{0G} lie in the interior of the interval spanned by p^k_{1G} and p^k_{0B}, i.e. $p^k_{1B}, p^k_{0G} \in (\min\{p^k_{1G}, p^k_{0B}\}, \max\{p^k_{1G}, p^k_{0B}\})$.

To provide some intuition for this assumption, let Y^H be the set of high signals, where $p^k_{1G} > p^k_{0B}$. Then this assumption implies that if $y^k \in Y^H, \ell^k_\alpha > 1$ for $\alpha \in \{G, B\}$ and $\ell^k_e > 1$ for $e \in \{0, 1\}$.That is, if a signal is more likely when a given type of agent chooses high effort, it is also more likely for a given effort level when the job is the good type. This implies that signals that are indicative of high effort are also indicative of the agent being the good type. Similarly, let Y^L be the set of low signals, where p^k_{1G} $< p^k_{0B}$. The assumption implies that if $y^k \in Y^L, \ell^k_\alpha < 1$ for $\alpha \in \{G, B\}$ and $\ell^k_e < 1$ for $e \in \{0, 1\}$, so that a low signal indicates low ability as well as low effort. Finally, we may have some uninformative signals when $p^k_{1G} = p^k_{0B}$, where all likelihood ratios are one, but since there is at least one informative signals, both Y^H and Y^L are non-empty. Let Y^U denote the set of uninformative signals, and let $\Pr(Y^U)$ denote the probability that an uninformative signal is realized – this does not depend upon effort choice or ability.

We shall assume that the agent's payoff in any period is given by $u(w) - c(e)$ where $u(.)$ is strictly concave, and unbounded, while $c(.)$ is increasing.

We begin our analysis by focusing on the principal's cost minimization problem. That is, we assume that the principal seeks to induce high effort in every period, and solve for the sequentially optimal dynamic contract that minimizes expected wage costs. Specifically, 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. We do not have to deal with out of equilibrium beliefs, since there are no observable deviations. 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).

3 Binary effort

The focus of our paper is on the case where the principal may make commitments within the period, but cannot commit to future contracts. We shall also focus on the case the principal seeks to induce high effort with probability one. There is an initial common prior probability $\lambda \in (0, 1)$ that the job is good.

Suppose that the principal and agent interact for two periods. The agent discounts future payoffs at rate $\delta \in (0, 1]$, while principal discounts at rate $\beta \in [0, 1]$. We shall assume that neither the principal nor the agent can commit in period one regarding the contract in period two. One interpretation of the model is that there are two short term principals, one arriving in period one and the second arriving in period two, after consumption has taken place in period one. The principal in period two observes the public signal (output) in period one. This implies that wages paid have to satisfy incentive compatibility and individual rationality period by period.

Suppose that the principal wants to induce $e = 1$ in both periods. Period 2 contracts are straightforward. Given that $e = 1$ is chosen, the principal's beliefs about the agent's beliefs are degenerate, and are given by μ_1^k after signal y^k. Thus the period two contract after signal y^k is given by $w(\mu_1^k) \in \mathbb{R}^{|Y|}$. Let $w_j(\mu_1^k)$ denote the wage paid under the optimal second period contract after second period signal realization y_j given that the principal has belief μ_1^k. We now show that the agent can always increase his continuation value by deviating to $e = 0$ in the first period. Suppose that the agent deviates in period one and chooses $e = 0$. His period two beliefs are now different from the principal's beliefs about the agent's beliefs. In particular, there is at least one signal realization such that he becomes more optimistic about his ability. Since the agent suffers no penalty when he becomes more pessimistic – he quits and gets his outside utility, which is the same as he would get when he chooses high effort, this implies that his continuation value increases.

The fact that the agent always becomes more optimistic at some signal realization after deviating is a consequence of the martingale property of beliefs. For any effort level e that the agent chooses, the expectation of his posterior must equal his prior, λ. Thus his expected beliefs under $e = 0$ must equal his expected beliefs under $e = 1$. Since $e = 1$ makes signals in Y^H more likely than when $e = 0$ is chosen the equality of expectations can only be satisfied if there is some signal realization y such that $\mu_0^k > \mu_1^k$, where μ_e^k is the posterior probability that the agent is the good type given signal realization y^k and effort choice e.

Lemma 1 *There exists some k such that $\mu_0^k > \mu_1^k$.*

Proof. From the martingale property of beliefs, $\mathbf{E}(\mu_1^k|e = 1) = \mathbf{E}(\mu_0^k|e = 0) = \lambda$, i.e.

$$\sum_{k=1}^{n} p_{0\mu}^k \mu_0^k = \sum_{k=1}^{n} p_{1\mu}^k \mu_1^k.$$

This can be written as

$$\sum_{k=1}^{n} p_{0\mu}^k (\mu_0^k - \mu_1^k) = \sum_{k=1}^{n} (p_{1\mu}^k - p_{0\mu}^k) \mu_1^k.$$

Since $\sum_{k=1}^{n} (p_{1\mu}^k - p_{0\mu}^k) = 0$ (being the difference between two probability distributions), $\sum_{k=1}^{n} (p_{1\mu}^k - p_{0\mu}^k)\lambda = 0$, so that

$$\sum_{k=1}^{n} p_{0\mu}^k (\mu_0^k - \mu_1^k) = \sum_{k=1}^{n} (p_{1\mu}^k - p_{0\mu}^k)(\mu_1^k - \lambda).$$

Under assumption A1, for any k, $(p_{1\mu}^k - p_{0\mu}^k)$ 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 $\sum_{k=1}^{n} p_{0\mu}^k (\mu_0^k - \mu_1^k) > 0$, 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 $\mu_0^k > \mu_1^k$. ■

We have therefore shown that the expectation of the "false belief" held by the principal, μ_1^k, that is induced when the agent performs the experiment $e = 0$, is strictly smaller than the expectation of the true belief μ_0^k. Thus there must be some signal realization for which $\mu_0^k > \mu_1^k$. This immediately proves that the agent can increase his continuation value by deviating to low effort.

The following examples illustrate our arguments. Let output be binary, so that $y \in \{y^H, y^L\}$. Tables 1-3 give examples of information structures that satisfy our assumptions, where the entries show the probability of the signal y^H, and $0 < q < p < 1$. In all these examples, high output is most likely when $\alpha = G$ and $e = 1$, and least likely when $\alpha = B$ and $e = 0$. Let $\lambda = 0.5$, although any interior value will suffice.

	$e = 1$	$e = 0$
G	p	p
B	p	q

Example 1

Consider example 1, where high effort only makes a difference to output when the job is bad. Thus $e = 0$ is a more informative experiment than $e = 1$. Indeed, $e = 1$ is uninformative about the realization of α, and if the principal induces high effort at $t = 1$, his posterior will equal the prior after either signal realization. Since $e = 0$ is informative, the agent becomes more optimistic after a success and more pessimistic after a failure. That is $\mu_0^H = \frac{p}{p+q} > \mu_1^H = \frac{1}{2}$ and $\mu_0^L = \frac{q}{p+q} < \mu_1^L = \frac{1}{2}$. The agent will quit after observing y^H. After y^H, he stays on the job and earns a surplus. In this case, he chooses $e = 0$, since his greater optimism implies that the incentive constraint is violated.

	$e = 1$	$e = 0$
G	p	q
B	q	q

Example 2

Example 2 is the polar opposite of the first example, since effort only makes a difference when the job is good, and $e = 0$ is uninformative. If he deviates to $e = 0$, he becomes more pessimistic than the principal after a success, and more optimistic after a failure. That is $\mu_1^H = \frac{p}{p+q} > \mu_0^H = \frac{1}{2}$ and $\mu_1^L = \frac{q}{p+q} < \mu_0^L = \frac{1}{2}$. Thus he quits after a success and earns a surplus after a failure. With greater optimism, he has more incentives to exert effort, and his incentive constraint is slack.[4]

	$e=1$	$e=0$
G	p	$\frac{p+q}{2}$
B	$\frac{p+q}{2}$	q

Example 3

In example 3 high effort raises the probability of success by the same magnitude, regardless of the nature of the job. In this case, it may be verified that $\mu_0^H > \mu_1^H$ and $\mu_0^L > \mu_1^L$, so that the agent is more optimistic than the principal after *either* output realization. Thus the agent always earns a surplus after deviating to $e = 0$. Since his incentive constraint is satisfied under his beliefs, it is optimal for him to exert high effort at $t = 2$. This example seems slightly paradoxical, since $\mu_0^k > \mu_1^k$ for all output realizations y^k. We label this phenomenon *uniform optimism* – as we shall see, many of our subsequent characterization results depend upon whether or not uniform optimism holds. Notice that uniform optimism does not violate the martingale property of beliefs – in our example, y^H has lower probability under $e = 0$ than under $e = 1$, and thus one can have the equality of $\mathbf{E}(\mu_1^k|e = 1)$ and $\mathbf{E}(\mu_0^k|e = 0)$, even though $\mu_0^k > \mu_1^k$ for every value of k.

3.1 A continuum of possible effort levels

We now assume that effort e must be chosen from $[0, 1]$. For simplicity, we focus on the two period case – extension to multiple periods is straight-forward. 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. Assume that cost of effort, $c(e)$ is strictly increasing, strictly convex and differentiable. Assume that the probability of signal y^k, conditional on type $\alpha \in \{G, B\}$ and $e \in [0, 1]$ is linear in

[4] In examples 1 and 2, a weaker version of A1 holds since p_{1B}^H and p_{0G}^H are not necessarily strictly less than p_{1G}^H or strictly greater than p_{0B}^H. As remark **??** makes clear, the agent necessarily gets a surplus when he is more optimistic, but he could get his reservation utility when he is more pessimistic, e.g. by choosing $e = 1$ in example 1, and $e = 0$ in example 2.

effort, and equal to[5]

$$p^k_{e\alpha} = ep^k_{1\alpha} + (1-e)p^k_{0\alpha}.$$

The revenue from inducing e at belief μ,

$$R(e,\mu) = \sum_k p^k_{e\mu} y^k,$$

is linear in e. We assume it is strictly increasing in e, since otherwise, inducing $e = 0$ is optimal. Note that the cross-partial $\frac{\partial R(e,\mu)}{\partial e \partial \mu}$ maybe positive or negative, so that it may be less profitable or more profitable to induce higher effort at optimistic beliefs, as in examples 2 and 3 respectively.

Definition 2 *Effort $\hat{e}$ is* implementable *if there exists some contract $(w^k)_{k=1}^{|Y|}$ such that $\hat{e}$ is optimal for the agent. A contract $(\hat{e}, (\hat{w}^k)_{k=1}^{|Y|})$ is optimal if it maximizes the principal's profits over all $(e, (w^k)_{k=1}^{|Y|})$ such that e is implementable.*

Claim 3 *In the final period, for any public belief μ, every effort $e \in [0,1]$ is implementable. The profit maximizing level of e, $\hat{e}(\mu)$ and the corresponding optimal wages, $\hat{w}_j(\mu)$, satisfy the first order conditions for the principal's maximization problem. Let $(w^k)_{k=1}^{|Y|}$ be an arbitrary contract at $t = 2$. There is a unique effort level that maximizes the agent's utility given this contract.***??**

We omit the proof the first part of this claim, since it is straightforward, and almost identical to the argument in Hart and Holmstrom (1987). To prove the second part, note that the agent's payoff from choosing e equals $\mathbf{E}(u(w|e) - c(e)$. Since the $p^k_{e\mu}$ is linear in e, the first term is linear in e and $c(e)$ is strictly concave, so there is a unique solution to the agent's maximization problem.

Let $V(\pi,\mu)$ denote the agent's rent at $t = 2$ given that a public belief μ and the agent's private belief π. $V(\pi,\mu) \geq 0$ since the agent can always quit when he gets less than his outside option. $V(\pi,\mu) = 0$ when $\pi = \mu$ since the agent's participation constraint binds under the optimal contract if he has the same beliefs.

Let us now turn to the first period, and suppose that the principal seeks to implement effort level e^*. Let $W(e, e^*)$ denote the expected continuation value of the agent

[5] 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 that our parameterization is equivalent to theirs, define a new effort variable, $c(e)$. This gives linear costs and a concave weighting function.

in the second period, net of his outside option, when he deviates to e. $W(e^*, e^*) = 0$, since the agent has no private information when he chooses the equilibrium level of effort.

Lemma 4 *If $\hat{e}(\mu) \neq 0$, then $V(\nu, \mu) > 0$ if $\nu > \mu$.*

Proof. From claim **??**, the optimal contract in the final period, $\hat{w}_j(\mu)$, must satisfy the first order conditions for $\hat{e}(\mu)$ to be optimal, i.e. [6]

$$\left.\frac{\partial \mathbf{E}(u(\hat{w}(\mu))|e, \mu))}{\partial e}\right|_{\hat{e}(\mu)} = c'(e)|_{\hat{e}(\mu)} > 0. \tag{1}$$

This can be re-written as

$$\sum_j (p^j_{1\mu} - p^j_{0\mu}) u\left(\hat{w}_j(\mu)\right) = c'(e)|_{\hat{e}(\mu)} > 0.$$

But this implies that

$$\sum_j (p^j_{\hat{e}(\mu)G} - p^j_{\hat{e}(\mu)B}) u\left(\hat{w}_j(\mu)\right) > 0.$$

■

Let $\hat{e}(\pi)$ denote the optimal effort of the agent under belief π. Thus,

$$V(\pi, \mu) = (\pi - \mu) \sum_j (p^j_{\hat{e}()G} - p^j_{\hat{e}(\mu)B}) u\left(\hat{w}_j(\mu)\right) > 0.$$

We now derive the conditions that a period one contract must satisfy if some effort level $e^* \in (0, 1)$ is to be implementable. Since e^* is in the interior, we must ensure that the agent does not want to deviate either downwards or upwards.

Let Y^-denote the set of signal realizations such that $\mu^k_0 > \mu^k_1$, i.e. $Y^- = \{y^k \in Y : \ell^k_0 > \ell^k_1\}$. Lemma 1 has established that Y^- is non-empty. Let Y^+ denote the set of signal realizations such that $\mu^k_0 < \mu^k_1$, i.e. $Y^+ = \{y^k \in Y : \ell^k_0 < \ell^k_1\}$.

Lemma 5 *Let $e, \tilde{e} \in [0, 1]$, and $e < \tilde{e}$. Then $\mu^k_e > \mu^k_{\tilde{e}}$ if and only if $y^k \in Y^-$. $\mu^k_e < \mu^k_{\tilde{e}}$ if and only if $y^k \in Y^+$.*

[6] If $e^*(\mu) = 1$, then equation (1) applies to the left hand derivative of $c(e)$ at 1.

Suppose that the principal seeks to implement effort level e^* at $t = 1$. This lemma implies that if the agent deviates downwards, to $e < e^*$, then he becomes more optimistic than the principal after signals in Y^-. If he deviates upwards, so that $e > e^*$, then he becomes more optimistic than the principals after signals in Y^+. Of course, under uniform optimism, Y^+ is empty.

Let $W(e, e^*)$ denote the second continuation value of the agent (over and above his reservation utility $\bar{u}$) when the principal seeks to implement e^*, and when the agent chooses e instead. If signal y^k is realized, then the agent's continuation value conditional on the signal is $V(\mu_e^k, \mu_{e^*}^k)$ as long as $\mu_e^k \geq \mu_{e^*}^k$. If $\mu_e^k < \mu_{e^*}^k$, then his continuation value is zero since he can always quit the job. Thus $W(e, e^*)$ is bounded below by zero. Clearly, $W(e^*, e^*) = 0$ since $V(\mu_{e^*}^k, \mu_{e^*}^k) = W(e, e^*)$ can be written as

$$W(e, e^*) = \begin{cases} \sum_{y^k \in \bar{Y}} p_{e\lambda}^k V(\mu_e^k, \mu_{e^*}^k) \text{ if } e < e^* \\ \sum_{y^k \in Y^+} p_{e\lambda}^k V(\mu_e^k, \mu_{e^*}^k) \text{ if } e > e^*. \end{cases}$$

We now evaluate left-hand and right hand derivatives of $W(e, e^*)$ at $e = e^*$, that enter the first order conditions for the implementability of e^*.The left hand derivative of $W(e, e^*)$ with respect to e, evaluated at $e = e^*$, is given by

$$\left.\frac{\partial W^-(e, e^*)}{\partial e}\right|_{e=e^*} = \sum_{y^k \in Y^-} \frac{\partial p_{e\lambda}^k}{\partial e} V(\mu_{e^*}^k, \mu_{e^*}^k) + \sum_{y^k \in Y^-} p_{e^*\lambda}^k \frac{dV(\mu_e^k, \mu_{e^*}^k)}{de}.$$

Since $V(\mu_{e^*}^k, \mu_{e^*}^k) = 0$, this simplifies to

$$\left.\frac{\partial W^-(e, e^*)}{\partial e}\right|_{e=e^*} = \sum_{y^k \in \bar{Y}} p_{e^*\lambda}^k \left.\frac{\partial V(\mu_e^k, \mu_{e^*}^k)}{\partial \mu_e^k}\right|_{\mu_e^k = \mu_{e^*}^k} \left.\frac{\partial \mu_e^k}{\partial e}\right|_{e=e^*} \tag{2}$$

We now show that each term under the summation sign is strictly negative. The first term, $p_{e^*\lambda}^k$, is positive. To evaluate the second term, let us consider the derivative of $V(\nu, \mu)$ with respect to ν. When $\nu > \mu$, $V(\nu, \mu)$ is given by

$$V(\nu, \mu) = \mathbf{E}(u(\hat{w}(\mu))|\tilde{e}(\nu), \nu) - c(\tilde{e}(\nu)) - \bar{u}, \tag{3}$$

where $\tilde{e}(\nu)$ denotes the payoff maximizing effort choice at belief ν.The derivative of this expression with respect to ν equals

$$\frac{dV(\nu,\mu)}{d\nu}=\frac{\partial\mathbf{E}(u(\hat{w}(\mu))|\tilde{e}(\nu),\nu)}{\partial\nu}+\left[\frac{\partial\mathbf{E}(u(\hat{w}(\mu))|\tilde{e}(\nu),\nu)}{\partial e}-c'(\tilde{e}(\nu))\right]\frac{\partial\tilde{e}(\nu)}{\partial\nu}. \quad (4)$$

By the envelope theorem, the second term is zero, and so

$$\frac{dV(\nu,\mu)}{d\nu}=\frac{\partial\mathbf{E}(u(\hat{w}(\mu))|\tilde{e}(\nu),\nu)}{\partial\nu}. \quad (5)$$

Evaluating this derivative at $\nu=\mu$, we get

$$\left.\frac{dV(\nu,\mu)}{d\nu}\right|_{\nu=\mu}=\left.\frac{\partial\mathbf{E}(u(\hat{w}(\mu))|\hat{e}(\mu),\nu)}{\partial\nu}\right|_{\nu=\mu}=\sum_j(p^j_{\hat{e}(\mu)G}-p^j_{\hat{e}(\mu)B})u\left(\hat{w}_j(\mu)\right).$$

This implies

$$\left.\frac{dV(\nu,\mu)}{d\nu}\right|_{\nu=\mu}\left.\frac{\partial\mathbf{E}(u(\hat{w}(\mu))|\hat{e}(\mu),\nu)}{\partial\nu}\right|_{\nu=\mu}=\sum_j(p^j_{\hat{e}(\mu)G}-p^j_{\hat{e}(\mu)B})u\left(\hat{w}_j(\mu)\right)>0.$$

Consider now the second term under the summation size in 2, the derivative of beliefs:

$$\frac{\partial\mu^k_e}{\partial e}=\frac{\lambda(1-\lambda)\left[p^k_{1G}p^k_{0B}-p^k_{1B}p^k_{0G}\right]}{\left[\lambda p^k_{eG}+(1-\lambda)p^k_{eB}\right]^2}.$$

The numerator is negative if $p^k_{1G}p^k_{0B}<p^k_{1B}\ p^k_{0G}$, i.e. if $\ell^k_1<\ell^k_0$. Thus the second term is negative, since $y^k\in Y^-$.

Since the left hand derivative of $W(e,e^*)$ at $e=e^*$ in equation (2) is the sum strictly negative terms, we conclude that the left hand derivative is strictly less than zero.

our main result is an impossibility theorem.

Theorem 6 *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 set of implementable efforts at $t=1$ is $\{0,1\}$.*

We now provide the intuition underlying the impossibility result. Let $e^*\in(0,1)$ be the effort level that the principal seeks to induce at $t=1$. Fig. 1 graphs the expected continuation value function of the agent, as a function of his effort level e, $W(e, e^*)$. $W(e^*, e^*) = 0$, since the agent has no private information when he chooses the equilibrium level of effort. Since the principal induces positive effort at $t = 2$, he must provide incentives for effort, i.e. the agent must be rewarded for outputs that are informative of high effort. Lemma 1 then implies that $W(e, e^*) > 0$ if $e < e^*$ – if deviating from $e = 1$ to $e = 0$ yields positive rents, then so does any downward deviation from e^*. Furthermore, the left hand derivative of $W(e, e^*)$, evaluated at $e = e^*$, is negative. On the other hand, for $e > e^*$, $W(e, e^*) \geq 0$ since the agent cannot get below his outside option. Thus at e^*, there is a kink the continuation value function, with the left-hand derivative being strictly negative and the right hand derivative non-negative. If Y^+ is empty, so that the signal structure satisfies uniform optimism, then the right hand derivative is zero. If Y^+ is non-empty, it can be shown that the right hand derivative is strictly positive, using the same arguments as above.

Let $(w^k)_{k=1}^{|Y|}$ denote an arbitrary contract that seeks to implement e^*. Let $\mathbf{E}(u(w)|e, \lambda))$ denote the expected utility of wages, where this expectation is taken with respect to the probability distribution over Y induced by e, and given the prior beliefs λ. The probability distribution over Y given e and λ is a differentiable function of e, and so is $c(e)$. Thus the first order conditions for e^* to be optimal for the agent are:

$$\left.\frac{\partial \mathbf{E}(u(w)|e, \lambda)}{\partial e}\right|_{e^*} - c'(e)|_{e^*} + \delta \left.\frac{\partial W^-(e, e^*)}{\partial e}\right|_{e=e^*} \geq 0.$$

$$\left.\frac{\partial \mathbf{E}(u(w)|e, \lambda)}{\partial e}\right|_{e^*} - c'(e)|_{e^*} + \delta \left.\frac{\partial W^+(e, e^*)}{\partial e}\right|_{e=e^*} \leq 0.$$

Since $\left.\frac{\partial W^-}{\partial e}\right|_{e=e^*} < \left.\frac{\partial W^+}{\partial e}\right|_{e=e^*}$, the two conditions cannot be simultaneously satisfied.

Our result is quite striking: no interior effort level can be implemented in the first period. That is, 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 a little shirking reduces the agent's current payoff. However, this implies that the agent can also 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. This point has not recognized in literature on ratchet effect, which has assumed that high powered incentives can solve the problem.

3.2 Risk Neutrality

Our results do not depend upon the agent being risk averse, but they do depend upon the absence of long term commitments. Suppose that the agent is risk neutral, but that 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, $F(\mu)$. This must satisfy the individual rationality constraint:

$$\max_e \left[\mathbf{E}(y|e,\mu) - c(e)\right] - F(\mu) \geq \bar{u}.$$

The optimal contract maximizes $F(\mu)$ subject to this constraint, so that

$$F(\mu) = \max_e \left[\mathbf{E}(y|e,\mu) - c(e)\right] - \bar{u}.$$

Let $\hat{e}(\mu)$ denote the value of e that maximizes $\left[\mathbf{E}(y|e,\mu) - c(e)\right]$. Thus the principal charges the agent a fee $F(\mu)$, that is increasing in μ under our assumptions.

Suppose that the agent is offered a contract $F(\mu)$, but has belief $\nu > \mu$. Then he will accept the contract and his payoff will be

$$\mathbf{E}(y|\hat{e}(\nu),\nu) - c(\hat{e}(\nu)) - F(\mu) = F(\nu) - F(\mu) + \bar{u}.$$

Thus

$$V(\nu,\mu) = F(\nu) - F(\mu) > 0 \text{ if } \nu > \mu.$$

In particular, the derivative is given by

$$\left.\frac{dV(\nu,\mu)}{d\nu}\right|_{\nu=\mu} = \sum_j (p^j_{\hat{e}(\mu)G} - p^j_{\hat{e}(\mu)B})y^k.$$

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 $\left.\frac{dV(\nu,\mu)}{d\nu}\right|_{\nu=\mu} > 0$. 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 $\bar{u}$ today by increasing his effort level beyond e^*, and quitting the job tomorrow, when signals in Y^- 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. The total return from this project is

$$\max \left\{ \sum_k p_{e\lambda}^k \left\{ y^k + \delta \left[\mathbf{E}(y|\hat{e}(\mu_e^k), \mu_e^k) - c(\hat{e}(\mu_e^k)) \right] \right\} \right\}.$$

Thus the agent would be willing to buy this project for this sum, minus $(1+\delta)\bar{u}$.

3.3 Implementing Random effort

One may ask if interior effort levels are implementable with positive probability. That is, can the principal design a contract where the agent randomizes over effort levels, using some mixed strategy σ, such that the maximal effort level is some $e \in (0,1)$. Let $S(\sigma)$ denote the support of σ. The following proposition gives a partial negative answer.

Proposition 7 *Suppose that there is uniform optimism. $e = 0$ is implemenatable. Let σ be a probability distribution over effort levels where* $\sup S(\sigma) \in (0,1)$. *Then σ is not implementable.*

The appendix provides the proof of this proposition. The underlying intuition is rather similar to that for our main result. If σ is optimal, then $e^* = \sup S(\sigma)$ must be optimal (even if e^* does not itself belong to $S(\sigma)$). Under uniform optimism, $W(e^*, e^*) = \bar{u}$, since the agent has the most pessimistic beliefs when he chooses e^*. Since $W(e, e^*) > \bar{u}$ when $e < e^*$, current wage payments must make it unprofitable for the agent to reduce effort below e^*. But this means that the agent can increase his current payoff by raising effort above e^*, and thus σ cannot be optimal.

3.4 Rent sharing

We have assumed so far that the firm holds the worker down to her outside option. However, the literature on worker compensation argues that firms often feel that they have to treat their workers fairly (e.g., Bewley, 1999), and this involves a degree of rent sharing. Leaving the worker some rent often improves incentive problems – the efficiency wage model is an example. This is the case in our context as well. Rent sharing may also arise from the worker having some bargaining power, whereby he gets a utility strictly higher than his outside option/

Rent sharing may make the non-implementability problem less severe. Suppose that the firm has a policy of offering a contract where it offers the worker an expected overall payoff of $\bar{u} + \Delta$, rather than her reservation utility $\bar{u}$. Thus the second period contract, $\hat{w}(\mu)$, ensures an expected payoff $\bar{u} + \Delta$ after every belief μ, on the equilibrium path.

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 $\bar{u}$ and $\bar{u} + \Delta$, he will stay on the job. When ν is slightly below μ, $V(\nu, \mu)$ is given by

$$V(\nu, \mu) = \mathbf{E}(u(\hat{w}(\mu))|\tilde{e}(\nu), \nu) - c(\tilde{e}(\nu)) - \bar{u}, \tag{6}$$

as long as this is positive ($\tilde{e}(\nu)$ denotes the payoff maximizing effort choice at belief ν). Since the agent will no longer quit if his private belief ν is just a little below the public belief μ, $W(e, e^*)$ decreases as e is increased beyond e^*, and is smooth at e^*. Fig 2a illustrates this for the case of uniform optimism. Now the non differentiability is now at some $\tilde{e} >> e^*$. In the case where there is not uniform optimism, there is an additional effect, since downward deviations below e^* are also less profitable. There are some signal realizations where the agent becomes more pessimistic after shirking, and since he stays on the job he incurs the cost of earning a rent lower than Δ. Thus rent sharing also makes $W(e, e^*)$ less steep around $e = e^*$. Fig 2.b illustrates the case without uniform optimism.

In either case, $W(e, e^*)$ is decreasing and differentiable at $e = e^*$. This implies that the first order conditions for implementing e^* are satisfied if

$$\left.\frac{\partial \mathbf{E}(u(w)|e, \lambda)}{\partial e}\right|_{e^*} - c'(e)|_{e^*} + \delta \left.\frac{\partial W(e, e^*)}{\partial e}\right|_{e=e^*} = 0.$$

Now the non differentiability is now at some $\tilde{e} >> e^*$. Indeed the value function $W(e, e^*)$ is necessarily convex as the diagrams illustrate. One has to therefore verify that choosing e^* is globally optimal. Large upward deviations may be unprofitable if the cost of effort function $c(e)$ is sufficiently convex.

3.5 Rents via Private Information

We now assume that the agent's reservation utility in the final period, v, is a random variable that has distribution function F.

Suppose that the principal has belief μ. Then his problem is to choose $\bar{v}, \hat{e}$ and $\left(w^k\right)_{k=1}^K$ to maximize

$$F(\bar{v})\left[\mathbf{E}((y-w)\,|\hat{e}, \mu)\right],$$

subject to the constraints

$$\mathbf{E}(u(w)|\hat{e}, \mu) - c(\hat{e}) \geq \bar{v},$$

$$\hat{e} \in \arg\max_e \mathbf{E}(u(w)|e, \mu) - c(e).$$

Since the first order approach is valid in this case, we may replace the incentive constraint by

$$\left[\frac{\partial \mathbf{E}(u(w)|e, \mu)}{\partial e} - c'(e)\right]\Bigg|_{e=\hat{e}} = 0.$$

We may break down this problem in two steps. First, for a given reservation 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.

Suppose now that the agent has belief π that may be different from μ. His payoff from accepting the contract equals

$$V(\pi,\mu) = \max_e \mathbf{E}(u(w)|e,\pi) - c(e).$$

Thus he will accept the contract if $V(\pi,\mu) \geq v$. His expected payoff, given optimal acceptance, is given by

$$\tilde{V}(\pi,\mu) = F\left(V(\pi,\mu)\right)V(\pi,\mu) + \int_{V(\pi,\mu)} vf(v)dv.$$

Suppose that $V(\pi,\mu)$ is differentiable in π. If F is continuous at $V(\pi,\mu)$, then the left and right hand derivatives of $\tilde{V}(\pi,\mu)$ are equal and the derivative of his expected payoff, with respect to π is given by

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

We now turn to the differentiability of $V(\pi,\mu)$, the value function conditional on the worker accepting the job. Recall that $w(\mu)$ denotes the optimal contract under public belief μ. Let

$$\hat{V}(\pi,\mu,e) = \sum e(p^k_{\pi 1} - p^k_{\pi 0})u(w^k(\mu)) - c(e)$$

denote the payoff as a function of effort. Let $\hat{e}(\pi)$ denotes the utility maxizing choice. The derivative of the value function is therefore $V(\pi,\mu) = \hat{V}(\pi,\mu,\hat{e}(\pi))$.

The derivative of V with respect to π equals

$$\frac{\partial V(\pi,\mu)}{\partial \pi} = \frac{\partial \hat{V}(\pi,\mu,\hat{e}(\nu))}{\partial \pi} + \frac{\partial \tilde{V}(\pi,\mu,e)}{\partial e}\frac{\partial \hat{e}}{\partial \pi}.$$

By the envelope theorem, the second term equals zero. Thus

$$\begin{aligned}
\frac{\partial V(\nu,\mu)}{\partial \nu} &= \frac{\partial \tilde{V}(\nu,\mu,\hat{e}(\nu))}{\partial \nu} \\
&= \sum\left\{\hat{e}(\pi)(p^k_{1G} - p^k_{1G}) + (1-\hat{e}(\pi))u(w^k(\mu))\right\} - c(\hat{e}(\pi)).
\end{aligned}$$

Evaluating this expression at $\nu = \mu$, we get

$$\left.\frac{\partial V(\nu,\mu)}{\partial \nu}\right|_{\nu=\mu} = \left[(e^*(\mu)(p_{1G} - p_{1B}) + (1 - e^*(\mu))(p_{0G} - p_{0B})\right] w^*(\mu).$$

3.5.1 first period

Let $W(e, e^*)$ denote the first period continuation value of the agent when he deviates to e, given that P induces e^*.This can be written as

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

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

4 Limited Liability

Consider the following model. The agent has limited liability, so that payments cannot be below some lower bound, that we may set to zero, i.e. $w^k \geq \dot{0} \forall k$. At $t = 2$, the has a finite action set and must choose $e \in \{0, 1\}$, with costs $c(1)$ and $c(0)$ respectively, where we may normalize $c(0)$ to zero. At $t = 1$, he has a continuum action set and must choose $e \in [0, 1]$, and $c(e)$ is strictly convex. We assume that the principal seeks to induce $e = 1$ at $t = 2$ at all beliefs. Our focus is on the conditions under which an interior effort level is implementable at $t = 1$.

Let μ denote the public belief at $t = 2$. The IC at $t = 2$ is given by

$$\mu \sum_k (p^k_{1G} - p^k_{0G}) u(w_k) + (1 - \mu) \sum_k (p^k_{1B} - p^k_{0B}) u(w_k) \geq c(1). \tag{7}$$

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

$$\pi \sum_k (p^k_{1G} - p^k_{0G}) u(w_k) + (1 - \pi) \sum_k (p^k_{1B} - p^k_{0B}) u(w_k) \geq c(1). \tag{8}$$

Define $\Delta(\mu)$ by

$$\Delta(\mu) = \sum_k \left[(p_{1G}^k + p_{0B}^k) - (p_{1B}^k + p_{0G}^k) \right] u(w_k). \tag{9}$$

Our analysis depends upon the sign of $\Delta(\mu)$, i.e. on whether it is strictly positive, strictly negagtive or zero. Consider first the case where $\Delta(\mu) > 0$. Then if $\pi > \mu$, it is strictly optimal to choose $e = 1$, and if $\pi < \mu$, it is strictly optimal to pick $e = 0$. The agent's expected utility may be written as a function of beliefs (π, μ) and is given by

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

This can be re-written as

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

Thus the derivative of V with respect to π equals

$$\frac{\partial V(\pi, \mu)}{\partial \pi} = \begin{cases} \sum_k \left[p_{1G}^k - p_{1B}^k \right] u\left(w_k\right) \text{ if } \pi > \mu \\ \sum_k \left[p_{0G}^k - p_{0B}^k \right] u\left(w_k\right) \text{ if } \pi < \mu. \end{cases}$$

V is therefore a piecewise linear function of π, with slope $V'^{-}(\mu)$ at $\pi \leq \mu$ and $V'^{+}(\mu)$ at $\pi > \mu$. The difference,

$$V'^{+}(\mu) - V'^{-}(\mu) = \sum_k \left[(p_{1G}^k + p_{0B}^k) - (p_{1B}^k + p_{0G}^k) \right] u(w_k) = \Delta(\mu) > 0. \tag{10}$$

On the other hand, if $\Delta(\mu) < 0$, the 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 $V'^{+}(\mu) - V'^{-}(\mu) = -\Delta(\mu) > 0$. Finally, if $\Delta\left(\mu\right) = 0$, both actions are optimal at every belief and $V(\pi, \mu)$ is linear in π. We conclude therefore that

$$V'^{+}(\mu) - V'^{-}(\mu) = \left|\Delta(\mu)\right|.$$

We now turn to the left-hand and right hand derivatives of $W(e, e^*)$ at $e = e^*$. These generate the first order conditions for the implementability of e^*. Let $Y^{-} = \{y^k \in Y : \mu_0^k > \mu_1^k\}$ and $Y^{+} = \{y^k \in Y : \mu_0^k < \mu_1^k\}$. If $y^k \in Y^{-}$, the agent's beliefs at

y^k are more pessimistic when he increases effort, i.e. $\frac{\partial \pi_e^k}{\partial e} < 0$. If $y^k \in Y^+$, $\frac{\partial \pi_e^k}{\partial e} > 0$.

The agent's expected value function when he chooses effort e at $t = 1$, given that the principal seeks to induce e^*, is

$$W(e, e^*) = \sum_{y^k \in Y} p_{e\lambda}^k V(\pi_e^k, \mu_{e^*}^k).$$

The left hand derivative of $W(e, e^*)$ with respect to e, evaluated at $e = e^*$, is given by

$$\begin{aligned} \left.\frac{\partial W^-(e, e^*)}{\partial e}\right|_{e=e^*} &= \sum_{y^k \in Y} \left.\frac{\partial p_{e\lambda}^k}{\partial e}\right|_{e=e^*} V(\pi_{e^*}^k, \mu_{e^*}^k) + \sum_{y^k \in Y^-} p_{e^*\lambda}^k V'^+(\mu_{e^*}^k) \left.\frac{\partial \pi_e^k}{\partial e}\right|_{e=e^*} \\ &\quad + \sum_{y^k \in Y^+} p_{e^*\lambda}^k V'^-(\mu_{e^*}^k) \left.\frac{\partial \pi_e^k}{\partial e}\right|_{e=e^*}. \end{aligned}$$

The right hand derivative is

$$\begin{aligned} \left.\frac{\partial W^+(e, e^*)}{\partial e}\right|_{e=e^*} &= \sum_{y^k \in Y} \left.\frac{\partial p_{e\lambda}^k}{\partial e}\right|_{e=e^*} V(\pi_{e^*}^k, \mu_{e^*}^k) + \sum_{y^k \in Y^-} V'^-(\mu_{e^*}^k) p_{e^*\lambda}^k \left.\frac{\partial \pi_e^k}{\partial e}\right|_{e=e^*} \\ &\quad + \sum_{y^k \in Y^+} p_{e^*\lambda}^k V'^+(\mu_{e^*}^k) \left.\frac{\partial \pi_e^k}{\partial e}\right|_{e=e^*}. \end{aligned}$$

Thus the difference between the right-hand and left-hand derivatives is

$$\begin{aligned} \left.\frac{\partial W^+(e, e^*)}{\partial e}\right|_{e=e^*} - \left.\frac{\partial W^-(e, e^*)}{\partial e}\right|_{e=e^*} &= \sum_{y^k \in Y^-} - \left|\Delta(\mu_{e^*}^k)\right| p_{e^*\lambda}^k \left.\frac{\partial \pi_e^k}{\partial e}\right|_{e=e^*} \\ &\quad + \sum_{y^k \in Y^+} \left|\Delta(\mu_{e^*}^k)\right| p_{e^*\lambda}^k \left.\frac{\partial \pi_e^k}{\partial e}\right|_{e=e^*}. \end{aligned}$$

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

It would seem that for generic information structures, $\Delta(\mu) = 0$ since this implies a linear restriction on the underlying probabilities. However, $\Delta(\mu)$ is the inner product of two vectors, one of which depends only on probabilities, while the other is the utility vector $\mathbf{u}(\boldsymbol{\mu})$ that depends upon the optimal contract at belief μ. To get some insight, let us consider the case of binary signals. Let the probability of y^H be given by the following table.

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

Binary signals: $\Pr(y^H)$

Assume $p > q$. By assumption A1, $\theta, \gamma \in (0, p - q)$. Clearly, if the IR constraint is irrelevant and the limited liability constraint binds, $w_L = 0$, and let us normalize utilities so that $u_L = u(0) = 0$. Let u_H denote $u(w_H)$, and let $c(0) = 0$. Thus the incentive constraint given μ is

$$[\mu\theta + (1 - \mu)\gamma]\, u_H = c(1).$$

Or

$$u_H = \frac{c(1)}{\mu\theta + (1 - \mu)\gamma}.$$

Thus $\mathbf{u} = (\frac{c(1)}{\mu\theta+(1-\mu)\gamma}, 0)$. The probability vector is

$$\mathbf{p}_{1G} + \mathbf{p}_{0B} - \mathbf{p}_{1B} - \mathbf{p}_{0G} = \begin{pmatrix} \theta - \gamma \\ \gamma - \theta \end{pmatrix}.$$

The inner product of the two yields

$$\Delta(\mu) = \frac{(\theta - \gamma)c(1)}{\mu\theta + (1 - \mu)\gamma}.$$

We conclude that unless $\theta = \gamma$, $|\Delta(\mu)| > 0$ for every value of μ. If $\theta = \gamma$, $|\Delta(\mu)| = 0$ for any μ, and thus in this case, W is differentiable in e.

We now show that the argument extends when there are many signals. Suppose that there are K signals, and let y^H denote the signal with the smallest likelihood ratio at belief μ, i.e the signal for which $\frac{p^k_{\mu 0}}{p^k_{\mu 1}}$ is minimal (if there are multiple signals with the maximal likelihood ratio, we can pool them and let y^H denote their union). Consider first the case where the agent is risk neutral, so that $u(w) = w$. Consider the contract that pays zero wages if $y \neq y^H$, and w_H after y^H, where

$$w_H = \max\left\{\frac{c(1)}{p^H_{\mu 1} - p^H_{\mu 0}}, \frac{c(1)+\bar{u}}{p^H_{\mu 1}}\right\}.$$

That is the w_H is set to satisfy both incentive and participation constraints. This is clearly an optimal contract under risk neutrality.

Consider first the case where the incentive constraint binds. Here the analysis is as for the binary signal case, where now the signals are partitioned into two sets. Consider next the case where the participation constraint binds. But this is exactly analagous to the case of the previous section, and thus the agent's value function is again locally convex.

The Lagrangian for this problem is

$$\max_{w_1,..,w_K,\rho,\beta} \mathcal{L}(\mu) = -\sum_k p^k_{1\mu} w_k + \rho\left(\sum_k p^k_{1\mu} u(w_k) - c(1) - \bar{u}\right) + \beta\left(\sum_k \left(p^k_{1\mu} - p^k_{0\mu}\right) u(w_k) - c(1)\right).$$

subject to the non-negativity constraints $w_k \geq 0, k = 1, 2, , K$.

The derivative of the Lagrangian with respect to w_k is

$$\frac{\partial \mathcal{L}(\mu)}{\partial w_k} = p^k_{1\mu}\left[-1 + \rho u'(w_k) + \beta\left(1 - \frac{p^k_{0\mu}}{p^k_{1\mu}}\right) u'(w_k)\right]. \tag{11}$$

Suppose that at the optimum, $\rho = 0$ so that the individual rationality constraint does not bind. Since β must be strictly positive, if $w_k > 0$,

$$\frac{1}{u'(w_k)} = \beta\left(1 - \frac{p^k_{0\mu}}{p^k_{1\mu}}\right). \tag{12}$$

Consider $y^k \in Y^L$, where the likelihood ratio $\frac{p^k_{0\mu}}{p^k_{1\mu}} > 1$. Since $u'(.)$ cannot be negative, the above condition cannot be satisfied for such a signal, implying that $w_k = 0$. In other words, wages are set at the minimum possible level for signals in Y^L. On the other hand, for signals in Y^H, wages can be zero or strictly positive, but are ordered in terms of the likelihood ratio, with wages being strictly positive at least for the most favourable signal.

Consider the case where the agent is risk neutral. Since $u'(.)$ is constant, the first order condition **??** can be satisfied at most for one likelihood ratio. Thus wages are positive only for the signal with the lowest value of $\frac{p_{0\mu}^k}{p_{1\mu}^k}$. This reduces to the binary signal case.

5 Related Literature

In its most general setting, we have focused on an agent and a principal who are both learning about the state of the world. The state is chosen by nature once and for all. The agent chooses an action, which may be thought of as effort or an experiment. The agent's action and the state jointly determine a public signal, but the agent's action is private, and not observed by principal. Phrased in these terms, this problem is being addressed by a growing literature on private experimentation with public signals in an agency setting, and by the career concerns model.

The first paper to deal with this general setting is Holmstrom's (1999) career concerns model. This early paper set out a model where the non-observability of effort gives rise to possible private information between employers and firms. By assuming a linear technology and normally distributed noise, Holmstrom was able to finesse many of the difficulties that arise due to the privateness of effort choice. In particular, optimal effort only depends upon calendar time, and not upon previous outputs or previously chosen efforts. Thus the agent's optimal continuation strategy does not depend upon his private information. There is a substantial literature that has developed on Holmstrom's career concerns model, retaining the key assumptions of a linear technology and normally distributed noise. Gibbons and Murphy (1992) allow for linear contracts in the context of Holmstrom's model, and show that explicit incentive become more important over time, as uncertainty is reduced. Meyer and Vickers (1997) study the interaction between implicit incentives arising from career concerns and explicit incentives. Dewatripont et al. (1999a) analyze the implications of alternative information structures and technologies, including a multiplicative technology. Their companion paper (1999b) considers career concerns in public organizations. Prat and Jovanovic (2011) analyze long term contracts with full commitment in a setting similar to Holmstrom's. Their main finding is that as information accumulates, the contracting problem becomes easier. De Marzo and Sannikov (2011) analyze a continuous time contracting problem where the state follows a Brownian motion.

More recently, a series of papers on venture capital consider information structures and technologies that are different from Holmstrom's. Bergemann and Hege (1998, 2005) consider the problem of venture capital financing, where output is binary and depends upon the quality of the project as well as the effort of the agent. Horner and Samuelson (2009) analyze the same question under different economic assumptions (where the project rather than capital is the scarce factor). Manso (2011) considers the problem of motivating innovation, in a context where the outcome variable is binary. Kwon (2011) analyzes a limited liability moral hazard problem, where the probability of success depends upon an unobserved state variable that is partially persistent.

The agency models discussed in this new literature mainly assume either limited liability or that the agent has all the bargaining power. Thus the ratchet effect does not arise, at least to the same extent, in these models.

6 Appendix

Claim 8 *If* $e \in S(\sigma)$, *then* $e \leq \hat{e}$.

Proof. Since there is uniform optimism, $W(e) \leq 0$ for $e > \hat{e}$. Since $\hat{e}$ yields a higher current payoff than any $e > \hat{e}$, e cannot be optimal. ■

Claim 9 $\sup S(\sigma) = \hat{e}$.

Proof. Suppose $\sup S(\sigma) = \tilde{e} < \hat{e}$. Then $W(\tilde{e}) = 0$, since there is uniform optimism. But then $\hat{e}$ yields a higher total payoff than $\tilde{e}$. ■

Suppose that $\tilde{e}, \hat{e} \in S(\sigma)$. Since $W(\hat{e}) = 0$, indifference requires

$$W(\tilde{e}) = [\mathbf{E}(y|\hat{e}, \lambda) - c(\hat{e})] - [\mathbf{E}(y|\tilde{e}, \lambda) - c(\tilde{e})] .$$

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

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

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

[6] Dewatripont, M., I. Jewitt and J. Tirole, 1999b, The economics of career concerns, part II: Application to missions and accountability of government agencies, *Rev. Econ. Studies* 66, 199-217.

[7] Edwards, R., 1979, Contested Terrain,

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

[9] Fudenberg, D., B. Holmstrom and P. Milgrom, 1990, Short term contracts and long-term agency relationships, *J. Econ. Theory* 51, 1-31.

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

[11] 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.

[12] Holmstrom, B., 1979, Moral hazard and observability, *Bell J. Econ.* 10, 74-91.

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

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

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

[16] Kanemoto, Y., and W. B. Macleod, 1992, The ratchet effect and the market for second-hand workers, *J. Labor Econ.*, 10(1),85-98.

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

[18] Laffont, J-J., and J. Tirole, 1990, The dynamics of incentive contracts, *Econometrica* 58, 1279-1319.

[19] Laffont, J-J., and J. Tirole, 1993. A theory of incentives inprocurement and regulation. MIT Press.

[20] Lambert, R., 1983, Long term contracts and moral hazard, *Bell J. Econ.* 14, 255-275.

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

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

[23] Mathewson, 1931, Restriction of output among unorganized workers,

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

[25] Rogerson, W., 1985, Long term contracts, *Econometrica* 58, 597-619.

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