Optimal Design of Peer Review and Self-Assessment Schemes*

Sandeep Baliga
Kellogg Graduate School of Management, Northwestern University

Tomas Sjöström
Department of Economics, Pennsylvania State University

September 26, 1999

Abstract

A principal must decide whether or not to implement a project which originated with one of her employees. The employees have private information about the quality of the project. A successfully implemented project raises the inventor's chance of promotion, at his peer's expense, but a failed project may ruin the inventor's career. If the inventor is already ahead in his career, then he may be tempted to suppress his own ideas in order not to risk a big failure. If he is not ahead, then he is instead tempted to exaggerate the quality of his ideas in order to get ahead. The peer may either try to promote the inventor's bad ideas to see him fail, or to denigrate promising ideas to stop the inventor from getting ahead. Within the class of incentive compatible and renegotiation-proof mechanisms, *self-assessment* (without any peer reports) is optimal. Truthtelling can be guaranteed in different ways. For example, to avoid the exaggeration effect, the inventor can be promised some chance of promotion even if his project is cancelled, or he can be paid a relatively high wage when he is *not* promoted. We show how the optimal method depends on the parameters.

1 Introduction

Career concerns create incentives for agents to misrepresent the quality of the work of their colleagues as well as their own work. In this paper, we identify four such effects. First, the incentive to promote a colleague's *bad* projects to see him fail. Second, the incentive to denigrate a colleague's *good* projects to prevent him from getting ahead.

*This is a revised version of our working paper "Not Invented Here" (Baliga and Sjöström [1]). We are grateful to Lars Stole and two anonymous referees for comments that significantly improved the paper. We also thank Bengt Holmström, Eric Maskin, and seminar participants at Chicago, Columbia, Harvard-MIT, Wisconsin (Madison), Maryland, Northwestern, Pennsylvania, Penn State, Pittsburgh, Stanford, SITE and the 1999 Decentralization Conference at NYU. Any errors are our responsibility.

Third, the incentive for workers who are not close to a promotion to take risks and promote their own work no matter what the perceived quality is, as they have little to lose. A fourth effect is the incentive for workers who are close to a promotion to suppress their own ideas in order not to have a big failure that ruins their reputation. We refer to these four effects as, respectively, *false praise*, *denigration*, *exaggeration*, and *false modesty*. We investigate the problem of designing optimal contracts in the presence of these effects. The relative importance of these effects depends on how far ahead in their careers the agents are. An agent who is close to being promoted will be tempted to denigrate the work of others and to be falsely modest about his own work. But an agent who has a long way to promotion is tempted to be enthusiastic about all projects, both his own (exaggeration) and those of his colleagues (false praise). Formally, this difference shows up in the fact that different truth telling constraints bind in the different cases.[1]

In our model there are two agents. One of them (the *inventor*) has developed a blueprint for a project. Both agents (but not the principal) observe the same signal about the quality of the blueprint, and the principal has to decide whether or not to implement the project. The quality of the project is correlated with the talent of the inventor and becomes known if and only if the project is implemented. The agents' careers are at stake: at the end of the game the principal must promote one of the agents, and she would prefer to promote the most talented one. Career concerns give the agents a reason to misrepresent their information. We compare different information gathering systems: (i) *Self-assessment*: only the inventor makes an assessment of the project. (ii) *Peer review*: only the agent who did not develop the project (the *peer*) makes an assessment of it. (iii) *Multiple reports*: both agents make assessments.

If the principal can commit to a mechanism, then it is optimal to ask for multiple reports. If both agents agree that the project is good, it is implemented. If they agree it is bad, it is not implemented. If the project is implemented its quality is observed by the principal, but if a project is not implemented she will never learn its true quality. Therefore, to encourage the agents to tell the truth, the principal should commit to implementing the project whenever the agents disagree. This gives her information that helps relax the truthtelling constraints, and is costless because in equilibrium the agents will not disagree. Since the principal's policy following a disagreement is designed solely to relax the truth-telling constraints, out of equilibrium the wrong agent may be promoted even when doing so is very costly. Such a

[1] This point is illustrated by Gamache and Kuhn's [7] empirical investigation of creativity and innovation in organizations. They claim that "{A} reason creativity dies is peer pressure. Individuals begin to think, If *he* fails he'll look bad and I'll get ahead. The message then is, Don't make a mistake your peers could exploit. Interestingly, years ago an executive with General Mills showed us...research on risk taking in the organization....The researchers found that ..high risk taking described entry level employees. They had little .. personal stakes in the organization and, consequently, felt they had little or nothing to lose. [T]he lowest risk taking was an arena peopled by middle management. These folks had a large career investment in the organization, essentially no security, lots of peer group pressure and competition and a lot to lose."

mechanism is not *renegotiation-proof.* Since it is hard to see how renegotiation of bad outcomes can be ruled out, we will impose a renegotiation-proofness constraint. A renegotiation-proof mechanism is robust in the sense that very costly policies will not be imposed if the agents "tremble" and disagree due to, for example, observational mistakes. This requirement seems to make the mechanism more realistic. Also, since the renegotiation-proofness constraint restricts the principal's ability to freely choose her response when reports disagree, one expects that renegotiation will reduce the value of collecting multiple reports. This is indeed what happens.

We show that self-assessment is optimal in the class of incentive compatible and renegotiation-proof mechanisms. In the optimal self-assessment contract, the project is implemented if and only if the inventor sends a positive report. If the project is implemented, the principal discovers the inventor's talent, and promotes the inventor if and only if he is revealed to be a good type. This part of the contract agrees with the first best[2] outcome. However, other aspects of the contract are distorted (compared to the first best). We say the inventor is *behind* in his career if in the first best he is not promoted when his project is cancelled, and he is *ahead* in his career if in the first best he is promoted even when his project is cancelled. If the principal tried to implement the first best outcome, then an inventor who is *behind* in his career would exaggerate to make sure his project is implemented, but an inventor who is *ahead* in his career (and so expects a promotion unless he suffers a big failure) would indulge in "false modesty". Therefore, to make sure the inventor reports his signal honestly, the self-assessment mechanism introduces certain distortions. If, conditional on the project being cancelled, the agents have approximately the same qualifications, the cheapest way to guarantee truthtelling is to distort the *promotion policy.* An inventor who is behind is promoted with strictly positive probability when the project is cancelled to avoid exaggeration. An inventor who is ahead is promoted with probability strictly less than one when the project is cancelled to avoid false modesty. Thus, along the equilibrium path the "wrong" agent is sometimes deliberatively promoted. But this outcome will not be renegotiated, as the principal would have to compensate the agent who gives up his promotion, and this would cost more than promoting the wrong agent as long as the agents' qualifications are fairly similar. However, suppose instead that after the project is cancelled one agent is *much* more qualified than the other: if the inventor is much more qualified than the peer then we say the inventor is *well ahead,* if the inventor is much less qualified than the peer then the inventor is *well behind.* In these cases, the cost of promoting the wrong agent would be significant, and the principal will not distort his promotion policy. Instead, the *wage policy* is distorted: an inventor who is well behind will be paid an extra high wage when he is not promoted in order to avoid exaggeration; an inventor who is well ahead and who has a successful project implemented is paid an extra high wage in order to avoid false modesty. Since the "right" agent is always promoted in these cases, this contract is

[2]The first best is what the principal would choose if she could observe the same signals as the agents.

certainly renegotiation proof.[3]

What we call *denigration* is in the sociology literature sometimes referred to as the *Not Invented Here (NIH) syndrome*. Coleman [5], page 443, characterizes the NIH-syndrome as "a lack of motivation, interest, and effort concerning ideas that originated outside a group, either elsewhere in the firm or in another firm. The group's investigation into an idea that originated elsewhere seems often to result in only a catalogue of reasons why the idea will not be useful." Coleman conjectures that the incentive to denigrate other people's work arises because "If an idea is clearly another's, an actor appears to have an interest in seeing the idea fail. This interest appears to arise because the success or failure of others' ideas provides a benchmark for evaluating one's own performance. By demonstrating the defects in another's idea, one justifies not having had the idea oneself; by allowing the idea's potential to be realized, one would be relatively worse off, because that would raise the standard for evaluation of one's own work."[4] Coleman adds that "The NIH syndrome is the opposite of what generally occurs when an innovator is given control of the development of his innovation. With that control he has a strong interest in seeing the idea successfully carried through to implementation." The last sentence seems to suggest a potential problem of exaggeration. However, while the literature generally worries about denigration by peers, exaggeration by inventors does not seem to be a big issue. Our results are consistent with this. Intuitively, exaggeration should be a less serious problem than denigration in our model. It is a risky strategy to exaggerate the quality of your own work, since at best you can convince the principal to implement a less promising project which is likely to fail. It is safer to denigrate the work of others, since if a promising project is stopped due to an unfair peer report, the principal will never learn the project's true quality. Thus, denigration may be harder for the principal to detect than exaggeration. The optimal renegotiation-proof contract in our model is in fact a policy of self-assessment. A policy of peer review is not optimal.

In the economics literature, Holmström [8] presented the first analysis of relative performance evaluation in a team (see also Levitt [12]). Holmström [9] analyses the impact of career concerns on incentives to work. Our focus is instead on adverse selection and on the optimal methods of gathering information. A number of papers consider a supervisor's evaluation of a subordinate (Prendergast and Topel [18] and Tirole [20]). They focus on rather different issues than our paper such as collusion and the effect of favoritism on optimal performance evaluation. Baker, Gibbons and Murphy [2] present several models on the use of subjective performance measures in optimal incentive schemes and present results on the substitutability and complementarity of (objective) explicit and (subjective) implicit incentive schemes. The fact that incentive schemes and promotion policies can cause agents to behave destruc-

[3]Wages will not be renegotiated even when they differ from first best, since monetary payments amount to a zero-sum game.

[4]In the sciences, it is often claimed that senior scientists block the ideas of younger colleagues. "A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die" (Max Planck [16]).

tively to further their own careers at the expense of others has been pointed out by Lazear [11], Milgrom [15], and Itoh [10], although in this literature there is no adverse selection and no predictions about optimal information systems. Rotemberg and Saloner [19] analyze a different kind of conflict within the firm: the production and sales departments disagree about business strategy, and try to present arguments that damage the other side's position. Two interesting papers that are somewhat closer to our setup are Carmichael [4], who shows the optimality of tenure contracts, and Levitt and Snyder [13], who analyze a principal-agent model where the agent is protected by limited liability and receives a signal as well as exerts effort. In the latter model the agent is tempted to exaggerate in order to get the high wage he earns if the project is successful, so he has to be compensated for honestly reporting bad signals. An interesting analysis of incentives to be aggressive and conservative in self-assessment is Prendergast and Stole [17]. None of these papers analyze the principal's choice between self-assessment, peer review and multiple reports, which is the focus of our paper.

The rest of the paper is organized as follows. In Section 2 we present the model and discuss the concept of renegotiation-proofness. In Section 3 we derive the optimal contract. It is useful to start by assuming the principal can commit not to renegotiate (the commitment case). In Section 3.2 we derive the optimal commitment contract when the inventor is behind, and we show that this contract is in fact renegotiation-proof. In Section 3.3 we derive the optimal commitment contract when the inventor is ahead. However, this contract is *not* renegotiation proof. The optimal renegotiation-proof contract for the case when the inventor is ahead is derived in Section 3.3.2. In Section 4.1, we show that, whether the inventor is behind or ahead, the optimal renegotiation-proof contract can always be replicated by a policy of self-assessment, where only the inventor sends a message. Section 4.2 discusses why self-assessment is better than peer reporting, and how the optimal self-assessment contract differs from the first best. Section 5 contains some extensions.

2 The Model

2.1 The time line

There are two agents and a principal. All are risk neutral, but there is *limited liability* for the agents: all wages must be non-negative. Each agent can be a *Good* or *Bad* type, denoted $T \in \{G, B\}$. An agent's type is not observable to anyone, including himself. The agents' types are uncorrelated random variables. Let λ_i be the prior probability that agent i is a good type. [5]

Agent 1 has developed a blueprint (or "project"). He is therefore called the *inventor* and agent 2 is called the *peer*. There is no moral hazard: the existence and quality of a project cannot be changed by any action taken by the agents (see

[5] Perhaps the agents looked identical at the time they were hired. After the hiring the principal has revised her estimates to λ_1 and λ_2 respectively.

Section 5 for a discussion of this assumption). A project can be of *Good* or *Bad* quality, denoted $v \in \{G, B\}$. Good projects are *successful* if implemented, and bad projects *unsuccessful.* A successful project is worth $G > 0$ to the principal, while an unsuccessful project is worth $B < 0$. The quality of the project is perfectly correlated with the inventor's type: good agents produce good projects, bad agents produce bad projects. The important point here is that the quality of the project is valuable information for the promotion decision; it is only for convenience that we assume the project outcome reveals the inventor's type *perfectly* (good if successful, bad otherwise). Imperfect but positive correlation between types and projects would not change our basic results. [6]

Persons with specialized knowledge, such as the agents, can decide if the project is promising or not, but the principal lacks the knowledge to make this judgement. Formally, each agent (but not the principal) observes a signal σ which is imperfectly correlated with the quality of the project. The signal is either good ($\sigma = g$) or bad ($\sigma = b$).[7] Both agents observe *the same signal*: if the project looks promising to one agent, it looks promising to the other agent too. The case where the agents might honestly disagree in their evaluations is more complicated and we postpone this to future work. Under our assumption, two truthful reports do not contain more information than one truthful report. However, by collecting two reports the principal can relax the truth telling constraints. In particular, she knows that some agent is lying if they disagree (although she does not know which one). However, we will show that when renegotiation is possible, the principal prefers to collect only one report. Moreover, this should always be the inventor's own evaluation.

Incentives to misrepresent signals arise because of career concerns. The agents are competing for a desirable promotion.[8] One and only one agent must be promoted.[9] Thus, we rule out the possibility of promoting *both* agents (it would be too expensive to create a second managerial position), as well as the possibility of promoting *no*

[6] Suppose the project outcome is an *imperfect* signal of the inventor's type, but still provides information that can change the promotion decision. If agent 1 is behind in his career, he has a reason to exaggerate the quality of his idea in order to get a boost from a successful project, just as discussed in this paper (and his peer has a reason to deprecate it). Similarly, an inventor who is ahead has a reason to indulge in false modesty (and his peer is tempted to give false praise). Of course, if the project quality is such an imperfect signal that the outcome of the project cannot actually change the promotion decision, then there are no strategic incentives for anybody to misrepresent any information. This trivial situation is ruled out in our model.

[7] The "bad signal" may simply be the absence of any information which supports the project, while the "good signal" might be the existence of such supporting material.

[8] If the agents did not care about being promoted, the mechanism could specify a constant (zero) wage for all messages and, as payoffs would be independent of messages, agents would tell the truth. To avoid this trivial case, we assume the promotion is desirable to the agents.

[9] One can allow for the possibility of not promoting any agent in the following way. Suppose not promoting any agent is so costly for the principal that she always promotes one agent in equilibrium. However, she can threaten to promote neither agent *out of equilibrium.* With this possibility, the full commitment contracts will be trivial and stipulate no promotion for either agent if messages disagree. If believed, this threat will certainly enforce a truthful Nash equilibrium. However, such a threat may not be credible if it is very costly not to promote anybody. Therefore, as long as we impose the constraint of renegotiation proofness, the results in this paper go through.

agent (the managerial position must be filled). The value to the principal of promoting a bad type is normalized to zero, the value of promoting a good type is $\Delta > 0$.

The non-pecuniary value an agent derives from a promotion is denoted $R > 0$.[10] Thus, if the agent is promoted and gets paid w his total payoff is $w + R$. Limited liability requires that $w \geq 0$. If limited liability were not imposed, the incentive to misrepresent information could be negated entirely by having a promoted agent pay the principal R. Since we assume limited liability, and since R is non-pecuniary, charging a fee from the promoted agent in this way is impossible. Limited liability also rules out punishing agents with fines when they disagree with each other, or when their predictions about the project turn out to be wrong.

The principal wants to elicit the agents' signals in order to inform the project implementation and promotion decisions. Since the agents observe the same signal, it is possible to use one agent's announcement to cross-check the other's. A version of the revelation principle implies that we can study direct revelation mechanisms where the agents are given the incentive to announce their information honestly.

After the project has been generated and signals observed by the agents, the game goes as follows:

Time $t = 1$. Each agent i sends a *message* $m_i \in \{g, b\}$, interpreted as a report on the project's quality.

Time $t = 2$. The principal receives the instruction "implement the project" or the instruction "don't implement the project" from the mechanism. The instruction "implement the project" is received with probability $h(m)$ where $m = (m_1 m_2)$ is the pair of messages sent at time $t = 1$. The principal does not observe the messages, only the instruction whether or not to implement.

Time $t = 3$. The outcome of the project is realized and becomes public knowledge. The set of possible outcomes of a project is

$$Y \equiv \{G, B, \emptyset\}$$

with generic element y. Here G (resp. B) denotes the outcome: the inventor's project was implemented at time 2 and successful (resp. unsuccessful), and $\emptyset$ denotes the outcome: the inventor's project was not implemented. The principal immediately observes the project's true quality if she implements it. If she does not implement it, she never observes the true quality.

Time $t = 4$. The principal receives the instruction "promote agent 1" or the instruction "promote agent 2" from the mechanism, together with an instruction about

[10] Thus, R is an "intrinsic reward" from the higher position. An alternative monetary interpretation of R can be given. Calvo and Wellisz [3] and others have argued that at low levels of a hierarchical firm, employees can be supervised by managers to make sure that they work. The (optimal) degree of supervision of high-level employees is smaller, but instead they receive (efficiency) wages which significantly exceed their reservation wages. In this case, $R > 0$ may represent the discounted expected value of a future increase in monetary (extrinsic) rewards. Now the firm could conceivably make the agent "pay for his promotion" using some part of the future efficiency wage. However, if the manager does not get to keep his efficiency wage he will shirk, so even though R represents a monetary value the firm will in fact never ask the manager to pay for his promotion in this way. In this sense, it does not matter if R is monetary or non-monetary.

what wages to pay. The instruction "promote agent 1" is received with probability $\theta^y(m)$ if messages $m = (m_1 m_2)$ were sent at $t = 1$ and $y \in Y$ was the outcome at $t = 3$. The wages can depend on the messages, the outcome of the project, and whether or not the agent is promoted.

The outcome of the project, the output (instructions) of the mechanism and the principal's actions are all verifiable to an outside party (a court), so the principal cannot *unilaterally* renege on the contract. However, we will consider the possibility of *Pareto-improving* renegotiation.

2.2 Conditional probabilities

The signal σ is accurate with probability q, where $\frac{1}{2} < q < 1$. Let $p(\sigma)$ denote the probability that the project of is of good quality, conditional on the signal $\sigma \in \{g, b\}$. By Bayes' rule,

$$
\begin{aligned}
p(g) &= \frac{\lambda_1 q}{\lambda_1 q + (1-\lambda_1)(1-q)} > \lambda_1 \\
p(b) &= \frac{\lambda_1 (1-q)}{\lambda_1 (1-q) + (1-\lambda_1) q} < \lambda_1
\end{aligned}
$$

Conditional on a good signal, the project is assumed to be profitable:

$$p(g)G + (1 - p(g))B > 0 \tag{1}$$

Bad projects make losses (since $B < 0$), but they have an informational value as they reveal the agent's type. Thus, if the value of promoting the right agent is high compared to the cost of implementing a bad project, implementing projects with bad signals might be desirable just to obtain more information about agent 1's type. The optimal contract in this case would not be very interesting (see Section 2.3). Thus, we assume that expected losses conditional on a bad signal are big compared to the value of promoting the right agent:

$$p(b)(G + \Delta) + (1 - p(b))(B + \lambda_2 \Delta) \leq \Delta \max\{p(b), \lambda_2\} \tag{2}$$

This inequality, which holds if $p(b)$ is small and B highly negative, guarantees that the principal does not want to implement projects with bad signals (Lemma 1 part v). Finally, we assume:

$$R < \Delta \min\{\lambda_2, 1 - \lambda_2\} \tag{3}$$

On the right hand side of (3), $\Delta\lambda_2$ is the principal's gain from promoting agent 2 instead of agent 1 if agent 1 is known to be bad, and $\Delta(1 - \lambda_2)$ is the gain from promoting agent 1 instead of agent 2 if agent 1 is known to be good. Thus, (3) says that the value to an agent of being promoted, R, is small compared to the cost to the principal of promoting the wrong agent in those cases where agent 1's type is known. This guarantees that at the optimum, the promotion decision is sensitive to the outcome of the project (Lemma 1 part iv). If (3) is violated, then the principal is better off removing the competition among the agents by always promoting the same agent, and paying zero wages (see Section 5.2).

Thus, our maintained assumption throughout the paper is:

Assumption 1 (1), (2) and (3) hold.

2.3 First best

Suppose the principal can costlessly observe the same signal as the agents. Then there is no need to pay any wages above the minimum of zero. Assuming (1) holds, the project should certainly be implemented when the signal is good. The inventor is promoted if the project succeeds, and the peer is promoted if it fails.

Now suppose the signal is bad. First, suppose the project is cancelled. Following the cancellation, the inventor is a better candidate for a promotion if and only if $p(b) > \lambda_2$. The principal will promote the agent who is the better candidate, and make an expected profit conditional on the bad signal of $\Delta \max\{p(b), \lambda_2\}$. Suppose instead the project is implemented despite the bad signal. If the project is successful, the principal learns that the inventor is good and promotes him, while if it fails she promotes the peer who is good with probability λ_2; the expected profit conditional on a bad signal is therefore

$$p(b)(G + \Delta) + (1 - p(b))(B + \lambda_2 \Delta)$$

Thus, (2) implies that the first best policy is to implement the project if and only if the signal is good. However, this gives rise to high-powered incentives for the agents to manipulate the principal's information. If the principal cannot observe the signal, the first best is not incentive compatible. In Section 3 we find the optimal contract subject to incentive compatibility constraints.

If (2) is violated, then *always* implementing the project is first best. In this case the first best outcome is incentive compatible: there is no need for the principal to observe the signal, or even for the agents to send messages, since the principal would just disregard the signal anyway. We rule this out by assuming that (2) holds.

2.4 Renegotiation

Since the agents have the same information, in equilibrium they will send the same report. If agents disagree, then the best way to encourage truth-telling is to promote the agent whose report coincides with the realized project quality. This policy minimizes the gain from lying, since lying is less likely to lead to a promotion. It may lead to the less talented agent getting promoted, but this does not cost the principal anything as long as disagreements only happen out of equilibrium. Realistically, however, there may be some probability that reports differ because agents honestly disagree about the project (or make mistakes). If this happens with positive probability, it will impose more discipline on the principal's policy: promoting the wrong agent in case of disagreement will no longer have zero cost. An alternative way of imposing such discipline, which we follow in this paper, is to insist on *renegotiation-proofness*. The contract is renegotiation-proof if at no stage of the game would the principal and both agents prefer to replace the old contract with a new one. The constraint of renegotiation-proofness ensures that the outcome is Pareto-efficient off as well as on the equilibrium path. Therefore, the principal cannot use promotion policies that it would be very costly for her to actually carry out.

In our model, the only times when the principal can have any incentive to propose a new contract are at $t = 2$ and at $t = 4$. Consider first time $t = 2$. Under our assumptions, the principal prefers a contract where in equilibrium the project is implemented if and only if the signal is good. With such a contract, if the principal gets the instruction “don’t implement the project” she infers that the signal is bad and then she has no reason to try to implement the project. Similarly, if the instruction is “implement the project” she infers that the signal was good, so she has no reason not to implement. Therefore, no matter what the principal hears there will be no temptation to renegotiate at $t = 2$. Notice that at time $t = 2$ the principal only receives the instruction on whether or not to implement, while the actual messages are kept secret from her.[11] This simplifies the analysis, for if she could see the actual messages at time $t = 2$, we would have to worry about renegotiation when $m_1 \neq m_2$. It turns out that in the optimal self-assessment mechanism we will identify in Section 4, there is no need for secret messages. Thus, ruling out secret messages would in fact not change our main results, but it would make the analysis of general message games more tedious.

Now consider time $t = 4$. Here the situation is different. As argued above, if the principal can commit not to renegotiate, then the optimal contract may specify that after some histories the principal promotes an agent who she knows is not the best person for the job. Such contracts are not renegotiation-proof if at $t = 4$ there exists a contract with a different promotion policy, which the agents would accept if offered to them, and which would make the principal better off. In order to introduce the constraint of renegotiation-proofness at time $t = 4$ we need to consider the principal’s beliefs at the time of the renegotiation, since she may infer something about the agents’ types from the history of play. The analysis is simplified by the assumption that if the project was implemented, the principal knows agent 1’s type *for sure*. We also assume that the principal always believes agent 2 is good with probability λ_2, independently of what has happened before time $t = 4$, because nothing that happens reveals any information about agent 2’s type. It only remains to discuss the principal’s beliefs about agent 1 when the project was not implemented. As long as the principal’s observations are consistent with equilibrium play, beliefs are assigned by Bayes’ rule. If the principal’s observations are inconsistent with equilibrium, Bayes’ rule is not applicable. Fortunately, it turns out that the only thing we need to assume in this case is that the principal thinks agent 1 is good with *at least* probability

[11] Suppose the two agents secretly type in their messages into a computer, which then makes public recommendations to the principal about project implementation, wages and promotions, as we have described. Any attempt by the principal to hack into the computer program in order to discover the agents’ messages causes the program to delete itself and all the messages.

$p(b)$. This is a reasonable lower bound, because the worst possibility for agent 1 is that the signal was bad. None of our results need a more sophisticated analysis of out-of-equilibrium beliefs. In particular, using the terminology of Maskin and Tirole [14], the concepts of weak and strong renegotiation-proofness coincide in this model.

It may appear that if renegotiation is possible, then the principal must always promote the agent who is the best candidate for the job. But this is not so. Suppose the contract is such that along the equilibrium path, the reports suggest that the project is bad, so the principal cancels it. She now believes the inventor is good with probability $p(b)$, and the peer with probability λ_2. There are two cases. (a) Suppose the inventor is *behind:* $\lambda_2 > p(b)$. Then the peer is the best choice for a promotion. If the mechanism recommends that both agents get a zero wage, and that the inventor is promoted, will the principal renegotiate and promote the peer instead ? The inventor will not accept the new contract unless he is compensated for losing the promotion; this compensation costs the principal R dollars. The peer would be willing to pay R dollars to get promoted, but limited liability rules this out. Thus, renegotiation increases the principal's wage cost by R. The expected gain from promoting the peer instead of the inventor is $(\lambda_2 - p(b))\Delta$, so if $(\lambda_2 - p(b))\Delta < R$ the contract will not be renegotiated. However, if $(\lambda_2 - p(b))\Delta > R$, renegotiation pays off; in this case we say the inventor is *well behind*. (b) Now suppose the inventor is *ahead:* $p(b) > \lambda_2$. Then the inventor is the best choice for a promotion. Suppose the mechanism recommends that both agents get a zero wage, and that the peer is promoted. Will the principal renegotiate and promote the inventor instead ? She would have to compensate the peer with R dollars, and she can take no money from the inventor by limited liability. Therefore, symmetrically to the previous case, if $(p(b) - \lambda_2)\Delta < R$ no renegotiation occurs. However, if $(p(b) - \lambda_2)\Delta > R$, renegotiation pays off; in this case we say the inventor is *well ahead*.

3 The Optimal Contract

3.1 Incentive constraints

It is useful to start out by describing the optimal *commitment contract*, i.e., the optimal contract if the principal can commit not to renegotiate. By the revelation principle, we may assume the agents play a Nash equilibrium where both tell the truth.[12] There are two truthtelling or IC constraints for each agent i, one for each signal $\sigma \in \{g, b\}$. (Recall that both agents see the *same* signal σ). Denote by $\text{IC}_i(\sigma)$ the constraint that agent i should tell the truth after seeing σ. Following messages $m = (m_1 m_2)$, the project is implemented with probability $h(m)$. Let $w_i^y(m)$ denote agent i's expected *wage* and $u_i^y(m)$ his expected *payoff* if the messages are m and the outcome of the project is $y \in \{G, B, \emptyset\}$. The payoff is the sum of the expected wage

[12] Just as in other multi-agent revelation games, in addition to the truthtelling equilibrium there may be other non-truthtelling ones. However, when we consider optimal renegotiation-proof mechanisms, only one agent (the inventor) needs to send a message. This eliminates the problem of multiple equilibria.

and the value of being promoted times the probability that a promotion occurs. For agent 1:

$$u_1^y(m) = w_1^y(m) + \theta^y(m)R$$

and similarly for agent 2. The limited liability constraints specify that all wages are non-negative.

Suppose agent 2 always tells the truth. Agent 1's expected payoff when he sees $\sigma = g$ and truthfully announces $m_1 = g$ is:

$$h(gg)\left(p(g)u_1^G(gg) + (1 - p(g))u_1^B(gg)\right) + (1 - h(gg))u_1^\emptyset(gg) \tag{4}$$

If he instead untruthfully announces $m_1 = b$, he expects to get

$$h(bg)\left(p(g)u_1^G(bg) + (1 - p(g))u_1^B(bg)\right) + (1 - h(bg))u_1^\emptyset(bg) \tag{5}$$

Using (4), (5) and the definition of conditional probabilities, the $\mathrm{IC}_1(g)$ constraint can be written as

$$\begin{aligned} & h(gg)\left(\lambda_1 q u_1^G(gg) + (1-\lambda_1)(1-q)u_1^B(gg)\right) + (1-h(gg))\left(\lambda_1 q + (1-\lambda_1)(1-q)\right)u_1^\emptyset(gg) \\ \geq\ & h(bg)\left(\lambda_1 q u_1^G(bg) + (1-\lambda_1)(1-q)u_1^B(bg)\right) + (1-h(bg))(\lambda_1 q + (1-\lambda_1)(1-q))u_1^\emptyset(bg) \end{aligned}$$

Similarly, the $\mathrm{IC}_1(b)$ constraint is

$$\begin{aligned} & h(bb)\left(\lambda_1(1-q)u_1^G(bb) + (1-\lambda_1)qu_1^B(bb)\right) + (1-h(bb))(\lambda_1(1-q) + (1-\lambda_1)q)u_1^\emptyset(bb) \\ \geq\ & h(gb)\left(\lambda_1(1-q)u_1^G(gb) + (1-\lambda_1)qu_1^B(gb)\right) + (1-h(gb))(\lambda_1(1-q) + (1-\lambda_1)q)u_1^\emptyset(gb) \end{aligned}$$

Similarly, we obtain two IC constraints for agent 2. The following Lemma is proved in the Appendix.

Lemma 1 *If the principal can commit not to renegotiate, then the following is optimal: (i) Pay a zero wage to both agents whenever they disagree ($m_1 \neq m_2$). (ii) Pay a zero wage to both agents whenever the project fails. (iii) Implement the project whenever the agents disagree: $h(gb) = h(bg) = 1$. (iv) Set $\theta^B(gg) = 0$ and $\theta^G(gg) = 1$. (v) Set $h(gg) = 1$ and $h(bb) = 0$.*

What happens in case of disagreement matters only through the right hand side of the IC constraints. Therefore, the principal will want to minimize wage payments (set them to zero) when there is disagreement, in order to reduce the wage payments that are made in equilibrium when agents tell the truth (part (i) of the lemma). Agents and the principal are risk-neutral and care only about expected wages. Therefore, it suffices to pay the inventor when his project succeeds (part (ii) of the lemma). What happens in case of disagreement does not influence the principal's expected payoff directly, because disagreement only happens out of equilibrium. Therefore, implementing the project when agents disagree is costless, but helps relax incentive constraints by allowing the principal to cross-check agents' messages against the realization of the project quality. Part (iii) follows from this observation.

The inequality (3) is used to prove part (iv) of the lemma. Suppose the inventor is promoted with positive probability when the project is bad. Then, if the principal reduces this probability by ε and compensates the inventor by increasing the expected wage by εR, she gains $\varepsilon\lambda_2\Delta$ by more often promoting a potentially good agent rather than one known to be bad, and she does not violate any IC constraints. The inequality (3) guarantees that the gain exceeds the increase in wages, so the inventor should never be promoted when the project is bad. Similarly, the principal would rather promote an inventor known for sure to be good over a peer who is good with probability $\lambda_2 < 1$. Finally, the inequalities (1) and (2) imply that projects with good signals make expected profits, and projects with bad signals are too costly to be worth implementing: this explains part (v).

Using Lemma 1 we can simplify the problem. Consider the principal's payoff. With probability $\lambda_1 q$ both the project and the signal is good. In this case, assuming the agents tell the truth, the project is successfully implemented (by Lemma 1 part (v)), agent 1 is promoted (by Lemma 1 part (iv)), and the principal's payoff is

$$G + \Delta - w_1^G(gg) - w_2^G(gg)$$

The principal's payoff for other cases is similarly computed using Lemma 1. Overall, the principal's expected payoff is

$$\begin{aligned} &\lambda_1 q \left(G + \Delta - w_1^G(gg) - w_2^G(gg)\right) + (1-\lambda_1)(1-q)\left(B + \lambda_2\Delta\right) \\ &+\lambda_1(1-q)\left(\theta^{\emptyset}(bb) + (1-\theta^{\emptyset}(bb))\lambda_2\right)\Delta + (1-\lambda_1)q(1-\theta^{\emptyset}(bb))\lambda_2\Delta \\ &-(\lambda_1(1-q)+(1-\lambda_1)q)\left(w_1^{\emptyset}(bb) + w_2^{\emptyset}(bb)\right) \end{aligned} \tag{6}$$

She maximizes this expression subject to the IC constraints, which using Lemma 1 we can simplify as follows:

$\mathrm{IC}_1(g)$:

$$\lambda_1 q \left(w_1^G(gg) + R\right) \geq \lambda_1 q \theta^G(bg) R + (1-\lambda_1)(1-q)\theta^B(bg) R \tag{7}$$

$\mathrm{IC}_1(b)$:

$$\begin{aligned} &(\lambda_1(1-q)+(1-\lambda_1)q)\left(w_1^{\emptyset}(bb) + \theta^{\emptyset}(bb)R\right) \\ \geq\ &\lambda_1(1-q)\theta^G(gb)R + (1-\lambda_1)q\theta^B(gb)R \end{aligned} \tag{8}$$

$\mathrm{IC}_2(g)$:

$$\lambda_1 q w_2^G(gg) + (1-\lambda_1)(1-q)R \geq \lambda_1 q(1-\theta^G(gb))R + (1-\lambda_1)(1-q)(1-\theta^B(gb))R \tag{9}$$

$\mathrm{IC}_2(b)$:

$$\begin{aligned} &(\lambda_1(1-q)+(1-\lambda_1)q)\left(w_2^{\emptyset}(bb) + (1-\theta^{\emptyset}(bb))R\right) \\ \geq\ &\lambda_1(1-q)(1-\theta^G(bg))R + (1-\lambda_1)q(1-\theta^B(bg))R \end{aligned} \tag{10}$$

The limited liability constraints are:

$$w_1^G(gg), w_2^G(gg), w_1^\emptyset(bb), w_2^\emptyset(bb) \geq 0 \tag{11}$$

To explain the left hand side of $\text{IC}_1(g)$, for example, use Lemma 1 to set $h(gg) = 1$, $u_1^G(gg) = w_1^G(gg) + \theta^G(gg)R = w_1^G(gg) + R$, and $u_1^B(gg) = w_1^B(gg) + \theta^B(gg)R = 0$ in the left hand side of the $\text{IC}_1(g)$ constraint stated before Lemma 1. For $m_1 \neq m_2$, we get $u_1^G(bg) = \theta^G(bg)R$, $u_2^G(bg) = (1 - \theta^G(bg))R$, $u_1^\emptyset(bg) = \theta^\emptyset(bg)R$ etc.

The details of the solution to the principal's problem depend on which incentive constraints are binding. We will split the discussion into two cases: $p(b) < \lambda_2$ and $p(b) > \lambda_2$.

3.2 The case where the inventor is behind

3.2.1 Optimal commitment contract

The inventor is *behind* if $p(b) < \lambda_2$. Then the peer is the best candidate for promotion conditional on a bad signal, so the inventor will be tempted to *exaggerate* the quality of the project. The peer is tempted to *denigrate* the project. The binding truth telling constraints therefore are $\text{IC}_1(b)$, avoiding *exaggeration*, and $\text{IC}_2(g)$, avoiding *denigration* (this is proved in the appendix).

Obviously, the principal particularly has to worry about what to do when the inventor says his project is good and the peer says it is bad. If the messages conflict in this way, the contract should minimize the incentive to lie by giving the promotion to the agent whose report agrees with the actual project quality. Thus, the inventor is promoted if and only if the project succeeds. With such a policy, denigration does not increase the peer's probability of promotion. However, the inventor can expect $p(b)R > 0$ by saying he has a good signal when in fact he has a bad one. To prevent exaggeration from occurring, in equilibrium the contract must give the inventor at least that much when he admits that his project is bad. This can potentially be done by promoting the inventor with positive probability when reports are bad. This does not raise the principal's wage cost, but it does mean that the wrong agent is promoted in equilibrium. If the reputation of the inventor given a bad signal, $p(b)$, is not much worse than the reputation of the peer, λ_2, then this policy is optimal. However, if $p(b)$ is much smaller than λ_2, then it is too costly for the principal to distort the promotion policy in order to reduce wage payments. Promoting the inventor instead of the peer has an expected cost of $(\lambda_2 - p(b))\Delta$, and the promotion is worth R to the inventor. If the inventor is *well behind*, so that $R < (\lambda_2 - p(b))\Delta$, then it is cheaper for the principal to simply pay R dollars to the inventor and promote the peer with probability one when reports are bad. Thus, the optimal policy depends on whether the inventor is well behind or not. The proof of the following proposition can be found in the Appendix.

Proposition 2 *If the principal can commit not to renegotiate and the inventor is behind, then the following is optimal. When the agents' reports contradict each other,* *promote the inventor if and only if his project is good ($\theta^G(m) = 1$ and $\theta^B(m) = 0$ if $m_1 \neq m_2$). If the inventor is not well behind, promote him with positive probability when both agents report the project is bad ($\theta^{\emptyset}(bb) = p(b) > 0$), and all wages are zero. If the inventor is well behind, never promote him when both reports are bad ($\theta^{\emptyset}(bb) = 0$), and all wages should be zero except $w_1^{\emptyset}(bb)$.*

The salient features of the optimal contract when the inventor is behind are that (1) the inventor should be compensated (either via a monetary payment or a probability of promotion) for admitting that his own project is bad, and (2) when the project is implemented, the inventor is promoted if and only if his project is good, whatever the message profile.

3.2.2 Renegotiation

We now show that the contract described in Proposition 2, which is optimal when the principal can commit, is actually renegotiation-proof.

Proposition 3 *Suppose the inventor is behind. The optimal contract from Proposition 2 is renegotiation-proof.*

Proof. It suffices to consider time $t = 4$. First, suppose the events the principal has observed are consistent with the equilibrium. From Lemma 1 part (iv), if the project is implemented, the inventor is promoted if and only if his project succeeds. Thus, the right person is promoted and no renegotiation takes place.

Now suppose the project is cancelled along the equilibrium path. From Proposition 2, if the inventor is behind and $R < \Delta(\lambda_2 - p(b))$, then the "right" worker is always promoted, so there is no reason to renegotiate. If instead $R > \Delta(\lambda_2 - p(b))$, the principal sometimes promotes the inventor ($\theta^{\emptyset}(bb) > 0$) even though he prefers to promote the peer (as $p(b) < \lambda_2$). Nevertheless, the contract is renegotiation-proof. Suppose the inventor is about to be promoted at $t = 4$, and the principal proposes a new contract where the peer is promoted instead. The inventor will insist on a compensation of at least R to give up the promotion, while the benefit from promoting the inventor instead of the peer is only $\Delta(\lambda_2 - p(b)) < R$. Thus, renegotiation is too expensive to the principal. Therefore, the optimal contract is always renegotiation-proof along the equilibrium path.

There are two out-of-equilibrium message profiles, $m = bg$ and $m = gb$. Following these messages the project is implemented (Lemma 1, part (iii)), and as shown in the proof of Proposition 2, the inventor is promoted if and only if his project is good. Clearly there is no incentive to renegotiate this outcome. ■

It may be remarked that the condition of being well behind plays two roles. First, as explained in Section 2.4, it is not renegotiation proof to promote the inventor when he is well behind and the project is bad. The reason is that the principal would prefer to make a monetary payment of R dollars to the inventor and promote the peer instead, since promoting the inventor would carry a loss of $\Delta(\lambda_2 - p(b)) > R$.

But secondly, Proposition 2 shows that even when the principal can commit not to renegotiate, he will not promote the inventor when he is well behind and the project is bad. The intuition is given before Proposition 2. In other words, the principal does not want to promote the inventor precisely when it would not be renegotiation proof to do so.

3.3 The case where the inventor is ahead

3.3.1 Optimal commitment contract

The inventor is *ahead* if $p(b) > \lambda_2$. In this case he is preferred for promotion even after a bad signal, but would not remain so after an unsuccessful project (recall that a bad signal does not necessarily mean that the inventor's type is bad, but a failed project does). Thus, the inventor likes the status quo and has an incentive to *underestimate* the quality of his own project (playing it safe rather than suffering a costly failure, which we call *false modesty*). The peer on the other hand thinks he can only get promoted if the inventor suffers an unsuccessful project (a bad signal is not enough), and a necessary condition for this is that a project is implemented. Thus, the peer is tempted to *overestimate* the project's quality. The binding truth-telling constraints are $\text{IC}_1(g)$ (no *false modesty*), and $\text{IC}_2(b)$ (no *false praise*). Here, the principal particularly has to worry about the promotion decision when the inventor says his project is bad while the peer report is good, i.e. $m = bg$. If messages conflict in this way, the principal minimizes the incentive to lie by promoting the agent whose message best corresponds to the project quality. Thus, she promotes the inventor with a higher probability if the project is unsuccessful than if it is successful.

With this policy, the peer gets a positive expected payoff from falsely praising the inventor when the signal is bad, as he gets promoted with some probability when the inventor's project fails. To prevent false praise, in equilibrium the peer must be compensated when he sends a negative report. This can potentially be done by promoting the peer with positive probability when reports are bad. If the reputation of the inventor following a bad signal, $p(b)$, is not much better than the prior reputation of the peer, λ_2, this method of inducing truth-telling is optimal. However, if $p(b)$ is much bigger than λ_2, then it is very costly to promote the peer following the bad signal. Recall that the inventor is *well ahead* if $\Delta\left(p(b) - \lambda_2\right) > R$. By a familiar argument, if the inventor is well ahead, then he should be promoted with probability one when reports are bad. False praise must be prevented, but when the inventor is well ahead it is cheaper to do this by a monetary payment to the peer rather than a promotion. The proof of the following proposition is in the Appendix.

Proposition 4 *Suppose the principal can commit not to renegotiate and the inventor is ahead. Then, the following policy is optimal. If $m = bg$ and the project* fails, *promote the* inventor *with positive probability ($\theta^B(bg) > 0$). If $m = bg$ and the project* succeeds, *promote the* peer *with positive probability ($\theta^G(bg) < 1$). If the inventor is well ahead, then promote him when $m = bb$ (i.e., $\theta^\emptyset(bb) = 1$), and all wages are zero except $w_2^\emptyset(bb)$ and possibly $w_1^G(gg)$. If the inventor is* not *well ahead,* *then promote the peer with positive probability when both agents report bad signals* ($\theta^{\emptyset}(bb) = 1 - p(b)(1 - \theta^G(bg))$)*, and all wages are zero except possibly* $w_1^G(gg)$.

The salient features of the optimal contract when the inventor is ahead are that (1) the inventor is sometimes given a monetary bonus $w_1^G(gg) > 0$ beyond the benefit of promotion when the project is implemented and is successful, (2) the peer is rewarded (either through a monetary payment or some probability of promotion) when both agents report the project is bad, and (3) for some message profiles, the inventor can get promoted even when his project is a *failure,* but he is sometimes not promoted when it is *successful*. Notice that point (1) is due to the need to prevent false modesty, and point (2) is due to the need to prevent false praise.

3.3.2 Renegotiation

We showed in Section 3.2.2 that renegotiation does not compromise the optimal commitment contract when the inventor is behind. We now show that when the inventor is ahead, the situation is different and the optimal commitment contract is not renegotiation-proof.

It suffices to consider time $t = 4$. First, suppose the events the principal has observed are consistent with the equilibrium. If both agents report good signals, the project is implemented, and the inventor is promoted if and only if his project succeeds so no renegotiation takes place. But suppose both agents report bad signals so the project is cancelled. From Proposition 4, if the inventor is well ahead then the "right" worker is always promoted, so there is no reason to renegotiate. In the remaining case, where $0 < \Delta(p(b) - \lambda_2) < R$, the peer is sometimes promoted ($\theta^{\emptyset}(bb) < 1$) even though the inventor is a better candidate. We claim there is still no incentive to renegotiate. Indeed, the peer will insist on a monetary compensation of R to give up the promotion, while the benefit from promoting the inventor instead of the peer is only $\Delta(p(b) - \lambda_2) < R$. Thus, renegotiation is too expensive for the principal and the optimal contract is renegotiation proof along the equilibrium path. Notice that the role played by the well ahead inequality in this argument is symmetric to the role played by well behind inequality when the inventor is behind.

Next, suppose the principal's observations are inconsistent with equilibrium. There are two cases where it may happen. (i) After a disagreement among the agents the principal may receive the instruction to promote the inventor even though his project fails: $\theta^B(gb) \neq 0$ or $\theta^B(bg) \neq 0$. Now, the principal knows the inventor is low quality after the unsuccessful project, and she can convince the inventor to decline the promotion by paying him R. By (3) the principal gains at least $\lambda_2\Delta - R > 0$. Therefore, the principal has an incentive to renegotiate the contract in this way. (ii) The principal may receive the instruction not to promote the inventor even though his project succeeds: $\theta^G(gb) \neq 1$ or $\theta^G(bg) \neq 1$. Analogously with the previous case, the principal gains at least $(1 - \lambda_2)\Delta - R > 0$ by renegotiating and promoting the inventor.

Renegotiation-proofness fails *out of equilibrium* when the inventor is ahead because, in order to provide incentives to tell the truth, the peer is promoted if he was the only one who supported a successful project, and the inventor is promoted if he was the one who did *not* support an *unsuccessful* project. However, at time $t = 4$ the principal prefers to renegotiate the contract rather than promote the "wrong" agent. Renegotiation-proofness implies that the inventor must be promoted whenever the project succeeds, but if the project fails he must not be promoted. This constraint strictly lowers the principal's expected payoff.

We now consider the optimal renegotiation-proof contract for the case where the inventor is ahead. Renegotiation-proofness imposes the constraint: for all m (even *out-of-equilibrium* m)

$$\theta^G(m) = 1 \quad \text{and} \quad \theta^B(m) = 0 \tag{12}$$

Now, we can no longer argue as we did in Lemma 1 parts (iii) and (v), that it is optimal for the principal to implement the project whenever the agents disagree because, when (12) is imposed, the principal is forced to promote the inventor if and only if his project is good. This will increase wage payments compared to the optimal commitment contract. In fact, it turns out that the principal prefers not to implement the project following messages $m = bg$. She still wants to implement when both reports are good ($h(gg) = 1$) since she cares sufficiently about having a successful project and promoting the "right" agent (see the Appendix). Moreover, it is clearly optimal to set $h(bb) = 0$, since the principal's incentive to cancel the project is reinforced by renegotiation.

Turning now to the principal's policy when the project is not implemented, recall that she always thinks agent 1 is good with probability *at least* $p(b)$. If the principal has been asked to promote agent 2, the gain from promoting 1 instead of 2 is at least $(p(b) - \lambda_2)\Delta$, and the cost is never greater than an extra payment of R to agent 2 (the principal can pay off agent 2 with the same salary as he would get according to the original contract, plus a compensation of R for not being promoted). If the inventor is well ahead, then $(p(b) - \lambda_2)\Delta > R$ and the contract is not renegotiation-proof. In this case we must impose the constraint

$$\theta^{\emptyset}(m) = 1 \quad \text{for all } m \tag{13}$$

That is, agent 1 is promoted whenever the project is not implemented. If the inventor is ahead but not well ahead, then (13) can be violated.

To find the optimal renegotiation-proof contract when the inventor is ahead, we modify the first program of Section 3 by *not* imposing

$$h(gg) = h(gb) = h(bg) = 1 \tag{14}$$

Instead we impose (12) and, if the inventor is well ahead, also (13).

Lemma 5 *Suppose the inventor is ahead but not well ahead. The following is the solution if the program of Section 3.1 is modified by not imposing (14), but instead* *imposing (12). Set $h(gg) = h(gb) = 1$ and $h(bg) = 0$. All wages except possibly $w_1^G(gg)$ are zero. If*

$$\frac{qp(b)}{(1-q)p(g)} R + (\ \lambda_2 - p(b))\,\Delta < 0 \tag{15}$$

then $\theta^{\emptyset}(bb) = \theta^{\emptyset}(bg) = 1$ and

$$w_1^G(gg) = \frac{1-p(g)}{p(g)} R$$

If

$$\frac{qp(b)}{(1-q)p(g)} R + (\ \lambda_2 - p(b))\,\Delta \geq 0 \tag{16}$$

then

$$\theta^{\emptyset}(bb) = \theta^{\emptyset}(bg) = p(g) \tag{17}$$

and $w_1^G(gg) = 0$.

Lemma 6 *Suppose the inventor is well ahead. When the program of Section 3 is modified by not imposing (14), but instead imposing (12) and (13), the solution is as follows. Set $h(gg) = h(gb) = 1$, $h(bg) = 0$. All wages are zero, except*

$$w_1^G(gg) = \frac{1-p(g)}{p(g)} R$$

Proofs of these lemmas are in the appendix.

Proposition 7 *The contracts characterized in Lemmas 5 and 6 are optimal within the set of renegotiation-proof contracts (for the cases where the inventor is ahead but not well ahead, or well ahead, respectively).*

Proof. We have already argued that a renegotiation-proof contract must satisfy the constraints of the programs analyzed in Lemmas 5 and 6. Thus, it suffices to show that the contracts found in Lemmas 5 and 6 are in fact renegotiation-proof. This is certainly true in the case of Lemma 6 because the right agent is always promoted by construction. That is, agent 1 (who is well ahead) is promoted except when his project has failed.

In the case of Lemma 5 the only problematic aspect is equation (17). In equilibrium the project is implemented if and only if the signal is good. Thus, if the principal receives the instruction "don't implement", he thinks the signal was bad, and the inventor is good with probability $p(b)$. In this case, the principal prefers to promote agent 1 (by definition of being ahead). According to (17), agent 2 is nevertheless promoted with probability $1 - p(g) > 0$. By renegotiating and promoting agent 1 instead of agent 2, the principal would gain $(p(b) - \lambda_2)\,\Delta > 0$. However, the principal would have to pay R to agent 2 to make him willing to give up the promotion. Since $R > (p(b) - \lambda_2)\,\Delta$ when the inventor is not well ahead, renegotiation does not pay. Therefore, the contract is renegotiation-proof. ■

When the inventor is ahead the optimal *renegotiation-proof* contract involves *not implementing* the project when only the peer supports it: $h(bg) = 0$. The point is that if the project is implemented, the principal discovers the inventor's true type, but this is not necessarily advantageous for a principal who cannot commit, and who is forced to obey (12).(Dewatripont and Maskin [6] discuss the fact that having more information is not always advantageous for the principal if she cannot commit.) By cancelling the project when the inventor says it is bad and the peer disagrees, the principal gains flexibility in using promotions as a reward. However, this policy in effect gives the inventor veto power over the implementation of the project. Therefore, as we show in the next section, this policy can be replicated by a self-assessment mechanism where only the inventor reports the quality of the project.

4 Self-Assessment

4.1 Self-assessment is optimal

The optimal renegotiation-proof contracts we identified above involve *multiple reports*. It turns out however that the outcomes can be replicated by a simpler self-assessment mechanism. It works as follows. Let the inventor announce his signal, but ask for no message from the peer. If the inventor supports his own project, then implement it, and promote the inventor if and only if the project is successful. If the project is successfully implemented and the inventor is ahead, then pay him $R(1-p(g))/p(g)$, except when he is not well ahead and (16) holds, in which case pay him zero. If the inventor does not support his own project, then don't implement it, and use the following promotion policy:

Case 1. Suppose agent 1 is behind. If agent 1 is well behind then promote agent 2 and pay agent 1 the salary $p(b)R$; but if agent 1 is not well behind then promote agent 1 with probability $p(b)$ and pay him zero.

Case 2. Suppose agent 1 is ahead. Then, if agent 1 is not well ahead and (16) holds, promote him with probability $p(g)$. Otherwise, promote him for sure.

Except as mentioned, all wages are zero.

It is easy to check that the inventor will tell the truth, and the outcome mimics the optimal renegotiation-proof contract. Moreover, the binding constraint when the inventor is behind is that he should not say his project is *good* when he receives a *bad* signal, i.e., *exaggeration* must be prevented. The binding constraint when the inventor is ahead is that he should not say his project is *bad* when he receives a *good* signal, i.e., *false modesty* must be prevented (cf. the low risk-taking of middle management and high risk-taking of entry level employees identified in Gamache and Kuhn [7]). Notice that the outcome of the self-assessment mechanism replicates the optimal *commitment* contract when the inventor is behind, but not when he is ahead. Another way to express this is that if the principal can commit, then multiple reports are useful if the inventor is ahead, but unnecessary if the inventor is behind.

This self-assessment procedure is renegotiation-proof. A project is implemented if and only if the inventor reports a good signal, and the principal does in fact want to go ahead with the project if and only if a good signal was received. When the project is implemented, the correct agent is promoted. When the project is not implemented, the promotion policy is the same as the optimal renegotiation-proof contracts derived above, so by the same reasoning the self-assessment mechanism is renegotiation-proof. Finally, the fact that the promotion policies and wages when the inventor is behind are the same as under the optimal full commitment contract (from Proposition 2), and when the inventor is ahead the same as under the optimal renegotiation-proof contract (from Proposition 7), means that the self-assessment mechanism is always optimal subject to renegotiation constraints.

Self-assessment *strictly dominates* peer review. Consider any mechanism where only the peer evaluates the project, and the project is implemented if and only if the peer report is good. Suppose, moreover, that in equilibrium the promotion policy replicates the policy from the optimal contracts of Proposition 2. First, consider the case where the inventor is a behind. Suppose the peer receives a good signal. If he sends a truthful report, then he is promoted only with probability $1 - p(g)$. But by announcing the project is bad, the peer will be promoted with probability one or $1 - p(b)$, depending on parameters. Therefore, the NIH syndrome arises, and the $\mathrm{IC}_2(g)$ constraint is violated unless the peer receives a monetary compensation for sending a good report. However, this compensation will be more expensive than the optimal contract (this is obvious when the inventor is not well behind, since in this case the optimal contract pays no wages, and when he is well behind it can be easily checked). Thus, the optimal contract cannot be replicated using peer review. Now suppose the inventor is ahead. Suppose the peer receives a bad signal. Then, following a bad report, the peer will be promoted with probability zero or $1 - p(g)$, depending on the parameters. By instead sending a good report, he is promoted with probability $1-p(b)$. Therefore, the false praise effect arises, and the $\mathrm{IC}_2(b)$ constraint is violated unless the peer receives a monetary compensation for sending a bad report. Again, it is readily checked that this is more expensive than the optimal contract so the optimal renegotiation-proof contract cannot be replicated by peer review.

One can gain some insight into why self-assessment dominates peer review by considering the agents' incentives to manipulate the *first best* outcome (described in Section 2.3). With self-assessment, an inventor who is behind will pretend the signal is good when it is bad. But this benefits the inventor only in the rather unlikely event that the project is successful even though the signal was bad, so the expected gain from manipulation is only $\lambda_1(1-q)R$. With peer reporting the peer will say the signal is bad when it is good. This unfair report causes a promising project to be stopped, and the peer is promoted. If the peer truthfully reports that the project is promising, the project is implemented and the peer is only promoted when the project fails, so the expected gain to the peer from this manipulation is $\lambda_1 qR > \lambda_1(1 - q)R$. The fact that the peer has more to gain from manipulating the first best suggests why denigration by a peer is a more difficult problem than exaggeration by an inventor. If the inventor convinces the principal to implement the project by exaggerating its quality, then it is very likely that the project is unsuccessful, so the inventor's expected gain is small. On the other hand, if a promising project is stopped due to an unfair peer report, the principal will never learn the project's true quality. Therefore, the incentives for the peer to manipulate are much greater. Consequently, guaranteeing truthtelling in a peer reporting mechanism will be more expensive for the principal. [13]

4.2 Discussion

The optimal self-assessment mechanism has many nice properties. First, it is standard in mechanism design to ask if a simple indirect mechanism is optimal (for example, simple nonlinear tariffs implement the optimal contract in the principal-agent model with adverse selection). In our model, self-assessment is a simple, optimal mechanism. Second, message games of the "Maskin type", where agents report all the information they have about each other to a social planner, are often criticized as being unnatural and not respecting the agents' privacy. A self-assessment mechanism where an agent only reports information pertaining to himself avoids this criticism. Third, while multi-agent revelation games often are plagued by multiple Nash equilibria, a mechanism where only one agents sends a report avoids this problem. Fourth, there is no need for "secret messages": the principal may just as well observe the inventor's message directly.

Since the first best is not incentive compatible, the optimal self-assessment mechanism (as described in Section 4.1) introduces certain distortions. If after the project is cancelled the agents have approximately the same qualifications, then the *promotion policy* is made less high-powered than it would be in the first best. This is relatively cheap as long as the principal does not care to much about which agent is promoted. Thus, an inventor who is behind but not well behind is promoted with probability $p(b) > 0$ even when the project is cancelled, in order to prevent exaggeration. (He would have had no chance of promotion under the first best policy.) If the inventor is ahead but not well ahead and (16) holds, then in order to prevent false modesty he is promoted only with probability $p(g) < 1$ when the project is cancelled. (He would have been promoted for sure under the first best policy.) But if one agent is a much better candidate for promotion than the other one after the project is cancelled, then distorting the promotion policy is very costly. Instead, the principal prefers to distort the *wage policy*. Namely, if the inventor is well ahead,[14] then to prevent false modesty he is given an extra high wage if his projects succeeds. But if the inventor is well behind, then to prevent exaggeration he is given an extra high wage when his project is cancelled.

[13] When the inventor is ahead, his temptation is not to exaggerate but to suppress his own project, while the peer wants them all implemented. In this case, a simple calculation shows that the inventor's gain from manipulating the first best contract is $(1-\lambda_1)(1-q)R$, while the peer's gain is $(1-\lambda_1)qR > (1-\lambda_1)(1-q)R$.

[14] Or not well-ahead, but (16) holds.

5 Extensions

5.1 Manipulation of project quality

In this paper the generation of projects has been treated as an exogenous process. It may be objected that if the inventor expects to be promoted even without a successful project, then he may simply refuse to develop a project. Actually, many explanations could be given to justify the existence of a project. For example, suppose there are three kinds of projects: Bad, Good, and Brilliant. Brilliant projects succeed for sure, and everyone can recognize them. If the payoff from completing a Brilliant project is very high, an inventor who is ahead will be willing to develop a project. Our model applies whenever he failed to develop a Brilliant one, and it is commonly known that the blueprint is either Good or Bad. However, it would seem possible for the inventor to manipulate (degrade) a promising project in such a way that it emits a less promising signal.[15] The principal will be persuaded to cancel the degraded project and the inventor who is ahead can avoid the risk of a failure which ruins his career. It turns out, however, that the optimal self-assessment mechanism discussed in Section 4 is immune to such manipulation. Therefore, *the self-assessment mechanism is optimal even when the inventor can degrade the quality of the blueprint.*

To see this, first suppose the manipulation of the blueprint is such that it only changes the signal from g to b but does not change the probability that the project actually succeeds if it is implemented. Then, deliberately creating a bad signal and announcing it truthfully is equivalent to observing a good signal but lying about it, and we know that this is not advantageous for the inventor. Now suppose the manipulation is such that the actual probability of success changes. That is, the inventor can manipulate a promising project in such a way that not only does it emit a bad signal, but the probability of success conditional on implementation falls, say from $p(g)$ to $p(b)$. But it can be checked that in the optimal self-assessment mechanism the inventor never gains from having a failure rather than a success. Therefore, for any message, the inventor wants the probability of success to be as high as possible. So, if he cannot gain from manipulation that leaves the probability of success unchanged, then he cannot gain from manipulation which lowers the probability of success either. Therefore, the inventor cannot gain by degrading the quality of the blueprint.

5.2 On the role of inequality (3)

Consider the self-assessment mechanism of Section 4.1. If the inventor is well ahead, the principal induces truth telling by paying $R(1 - p(g))/p(g)$ when the project is implemented and succeeds, which happens with probability $\Pr(\sigma = g) \times p(g)$. If we use the first best outcome as described in Section 2.3 as a benchmark, the self-assessment

[15] If a "bad signal" is simply the absence of any information that supports the project, then the inventor may be able to suppress any supporting information.

mechanism implies a loss of

$$\Pr(\sigma = g)p(g)\frac{1-p(g)}{p(g)}R = (1-\lambda_1)(1-q)R \tag{18}$$

for the principal.

Inequality (3) implies $\lambda_2 > R/\Delta$. If this inequality is reversed, then the self-assessment mechanism of Section 4.1 is dominated by a policy where the inventor is *always* promoted, either one of the two agents is asked for a report, the project is implemented if and only if the report is good, and all wages are set equal to the minimum (zero). This new policy is certainly incentive compatible, and it is renegotiation-proof since the gain of $\Delta\lambda_2$ from promoting agent 2 when agent 1 has failed is smaller than the R which must be paid to agent 1 if he deprived of his promotion. The project is implemented and fails with probability $\Pr(\sigma = g) \times (1 - p(g))$, in which case the principal loses $\lambda_2\Delta$ by not promoting agent 2. The expected loss from this policy, compared to the first best, is

$$\Pr(\sigma = g)\,(1-p(g))\,\lambda_2\Delta = (1-\lambda_1)\,(1-q)\lambda_2\Delta$$

This cost is smaller than the cost given in (18) if and only if $\lambda_2 < R/\Delta$. Thus, $\lambda_2 > R/\Delta$ implies that the principal does not want to solve the problem of misrepresentation of information simply by always promoting agent 1. This inequality also implies that the second employee adds value to the firm: the principal prefers having two candidates for promotion to choose from instead of just one, even though wages can be zero when there is only one candidate.

The inequality (3) also implies $1 - \lambda_2 > R/\Delta$. By an argument similar to the one just given, it can be shown that this inequality implies that always promoting the *peer* does not dominates the self-assessment mechanism of Section 4.1.

6 Appendix

6.1 Optimal contracts with commitment

Proof of Lemma 1

(i) This follows from the fact that the disagreement payoffs $u_1^B(bg), u_1^G(bg)$ etc. only enter on the right hand side of the IC constraints.

(ii) The wages $w_1^G(gg)$ and $w_1^B(gg)$ enter in the principal's payoff and the $\mathrm{IC}_1(g)$ constraints through the term

$$p(g)w_1^G(gg) + (1-p(g))w_1^B(gg)$$

which is the expected wage to agent 1 when $\sigma = g$ and the project is implemented. Since both the principal and the agent only care about this expectation, it is without loss of generality to set $w_1^B(gg)$ as low as possible. A similar argument holds for $w_2^B(gg)$.

(iii) From (i) we can suppose all the disagreement wages are zero. Notice that $h(bg)$ only appears on the right-hand side of $\text{IC}_1(g)$ and $\text{IC}_2(b)$. It is intuitively clear that implementing the project after disagreement is optimal, as it conveys information about who was lying for free. Indeed, suppose a proposed contract has $h(bg) < 1$. Suppose the principal changes the contract in the following way: following the message bg, the project is implemented with probability one, and if the project turns out good, agent 1 is promoted with probability

$$h(bg)\theta^G(bg) + (1 - h(bg))\theta^{\emptyset}(bg)$$

(where $h(bg)$, $\theta^G(bg)$, $\theta^{\emptyset}(bg)$ are as specified in the original contract). If the project is bad, he promotes agent 1 with probability

$$h(bg)\theta^B(bg) + (1 - h(bg))\theta^{\emptyset}(bg)$$

This leaves the right-hand sides of $\text{IC}_1(g)$ and $\text{IC}_2(b)$ unchanged. The principal's welfare is unchanged since disagreements never happen in equilibrium. Hence, we can assume without loss of generality that $h(bg) = 1$ and by a similar argument, $h(gb) = 1$.

(iv) Suppose $\theta^B(gg) > 0$. Consider changing the contract by reducing $\theta^B(gg)$ by ϵ. Maintain the same expected payoffs (conditional on messages, outcomes and promotion decisions) for both agents, except that $u_2^B(gg)$ may have to be increased to respect the limited liability constraints (agent 2 is more often promoted if $\theta^B(gg)$ is reduced). Clearly, though, it will not be necessary to increase $u_2^B(gg)$ by more than ϵR. In the new contract the IC constraints are obviously still satisfied. The increase in expected wages is no greater than

$$\Pr(\sigma = g)(1 - p(g))\, h(gg)\epsilon R$$

while the principal gains

$$\Pr(\sigma = g)(1 - p(g))\, h(gg)\epsilon\lambda_2\Delta$$

because under the new contract he promotes agent 2 (who is good with probability λ_2) more often when agent 1 is *known* to be bad (after the outcome of the project was B). By (3), this improves the principal's payoff. The argument for $\theta^G(gg) = 1$ is similar.

(v) Consider $h(gg)$. By implementing the project when the signal is good, the principal gets more information because she can observe the outcome of the project. Since she can always disregard this information if she wants (as in part (iii) of this Lemma), she can design a policy with $h(gg) = 1$ which implies no greater wage payments than a policy with $h(gg) < 1$. As messages (gg) are received in equilibrium whenever $\sigma = g$, there is also a direct effect on the principal's revenue from increasing $h(gg)$. But this is positive, because $Gp(g)+B(1-p(g)) > 0$ by assumption. Therefore, $h(gg) = 1$ is optimal.

Consider $h(bb)$. Suppose a contract has $h(bb) = h^* > 0$. By the same argument as in part (iv), we can set $\theta^B(bb) = 0$ and $\theta^G(bb) = 1$. Consider a new contract where the project is never implemented following the message (bb). Now the principal implements fewer unsuccessful projects and her expected income increases by

$$-\Pr(\sigma = b)h^*\left(Gp(b) + B(1 - p(b))\right) > 0 \tag{19}$$

In the old contract agent 1 got promoted with probability $\theta^\emptyset(bb)$ when the project was not implemented and messages were bb. In the new contract replace $\theta^\emptyset(bb)$ by $\hat{\theta}^\emptyset(bb)$, where $\hat{\theta}^\emptyset(bb)$ is chosen so that after $m = (bb)$ agent 1 has the same chance of being promoted as in the old contract (that is, $\hat{\theta}^\emptyset(bb) = h^*p(b) + (1 - h^*)\theta^\emptyset(bb)$). Pay the agents the same (conditional on promotion or no promotion) as under the old contract. As the probability of promotion and the wages are the same, the expected payoffs are the same as under the old contract, so the IC constraints still hold, and so do the limited liability constraints. The principal does lose something from not being able to find out the true quality of the agent when $\sigma = b$: this loss is

$$-\ \Pr(\sigma = b)h^*\left(p(b)(1 - \hat{\theta}^\emptyset(bb))(1 - \lambda_2) + (1 - p(b))\hat{\theta}^\emptyset(bb)\lambda_2\right)\Delta \tag{20}$$

To see this, consider what happens under the old contract when $\sigma = b$, $m = (bb)$ and the principal implements the project. With probability $p(b)$ the inventor is a good type, and the principal finds it out and promotes him. On the other hand, in the new contract, the principal will promote a bad agent 2 with probability $(1 - \hat{\theta}^\emptyset(bb))(1 - \lambda_2)$ instead. With probability $1 - p(b)$, the inventor is a bad type, and under the old contract the principal finds it out and promotes agent 2 who is good with probability λ_2. In the new contract, the principal will promote the bad agent 1 instead with probability $\hat{\theta}^\emptyset(bb)$, thus reducing the probability of promoting a good agent by $\hat{\theta}^\emptyset(bb)\lambda_2$. Now (2) makes sure that the sum of (20) and (19) is positive, so the new contract dominates. ■

We shall use the results from Lemma 1 in the following propositions.

Proof of Proposition 2

Suppose the inventor is behind. There are two cases, (1) when the inventor is not well behind and (2) when the inventor is well behind.

Case 1: The inventor is not well behind:

$$p(b) < \lambda_2 < p(b) + \frac{R}{\Delta} \tag{21}$$

We claim the following is optimal: set all wages equal to zero and

$$\begin{aligned}
\theta^G(gb) &= \theta^G(bg) = 1 \\
\theta^B(gb) &= \theta^B(bg) = 0 \\
\theta^\emptyset(bb) &= p(b)
\end{aligned}$$

We prove the proposition via a series of claims.

Claim 1: At the optimum, it must be the case that $w_2^{\emptyset}(bb) = 0$.

Proof: Suppose $w_2^{\emptyset}(bb) > 0$. If $\theta^{\emptyset}(bb) = 0$ then $\mathrm{IC}_2(b)$ is not binding and $w_2^{\emptyset}(bb)$ should be reduced to zero. If $\theta^{\emptyset}(bb) > 0$, then the principal can reduce $w_2^{\emptyset}(bb)$ by ϵR, increase $w_1^{\emptyset}(bb)$ by ϵR and reduce $\theta^{\emptyset}(bb)$ by ϵ. This increases her payoff (by (21)) without violating any incentive constraints.

Claim 2: $\mathrm{IC}_1(b)$ binds at the optimum.

Proof: Suppose not. Then $w_1^{\emptyset}(bb) = 0$, or else the principal can lower $w_1^{\emptyset}(bb)$ without violating any constraints. Thus, $\theta^{\emptyset}(bb) > 0$ if $\mathrm{IC}_1(b)$ does not bind. But lowering $\theta^{\emptyset}(bb)$ raises the principal's profit by (21) without violating any incentive constraints.

Claim 3: $\mathrm{IC}_2(g)$ binds at the optimum.

Proof: Suppose not. Then we must have $w_2^G(gg) = 0$, and either $1 - \theta^G(gb)$ or $1 - \theta^B(gb)$ is strictly less than one. Hence, one of these variables can be increased without altering the principal's payoff and not violate $\mathrm{IC}_2(g)$. This relaxes $\mathrm{IC}_1(b)$, but then profit can be increased as in the proof of claim 2.

Claim 4: $\theta^B(gb) = 0$, $\theta^G(gb) = 1$ and $w_2^G(gg) = 0$.

Proof: By claims 2 and 3, $\mathrm{IC}_1(b)$ and $\mathrm{IC}_2(g)$ bind at the optimum. Therefore, at the optimum, the principal minimizes the sum of the right hand sides of $\mathrm{IC}_1(b)$ and $\mathrm{IC}_2(g)$ subject to the constraint that the right hand side of $\mathrm{IC}_2(g)$ must be at least $(1 - \lambda_1)(1 - q)R$. Then, as $q > 1 - q$, $1 - \theta^G(gb) = 0$ and $1 - \theta^B(gb) = 1$. Finally, $w_2^G(gg) = 0$ from $\mathrm{IC}_2(g)$.

Assume $\mathrm{IC}_1(g)$ and $\mathrm{IC}_2(b)$ are not binding (we will verify this later). Then, clearly $w_1^G(gg) = w_2^{\emptyset}(bb) = 0$ is optimal.

Claim 5: $w_1^{\emptyset}(bb) = 0$.

Suppose $w_1^{\emptyset}(bb) > 0$. $\mathrm{IC}_1(b)$ binding means $\theta^{\emptyset}(bb) < 1$. Lower $w_1^{\emptyset}(bb)$ by ϵR and increase $\theta^{\emptyset}(bb)$ by ϵ. This raises the principal's profit by

$$\Pr(\sigma = b)\epsilon\,(R - \Delta(\lambda_2 - p(b))) \geq 0$$

by (21) without violating any constraints. This proves the claim.

Claims 2 and 5 imply $\theta^{\emptyset}(bb) = p(b)$.

Finally, we can make sure $\mathrm{IC}_1(g)$ and $\mathrm{IC}_2(b)$ are satisfied by setting $\theta^B(bg) = 0$ and $\theta^G(bg) = 1$. This completes the proof for case 1.

Case 2: The inventor is well behind:

$$\lambda_2 > p(b) + \frac{R}{\Delta}. \tag{22}$$

We claim it is optimal to set all wages equal to zero except $w_1^{\emptyset}(bb) = p(b)R$, and

$$\begin{aligned} \theta^{\emptyset}(bb) &= \theta^B(bg) = \theta^B(gb) = 0 \\ \theta^G(gb) &= \theta^G(bg) = 1 \end{aligned}$$

Suppose $\theta^{\emptyset}(bb) > 0$. Then lower $\theta^{\emptyset}(bb)$ by ϵ and raise $w_1^{\emptyset}(bb)$ by $R\epsilon$. This changes the principal's payoff by

$$\Pr(\sigma = b)\left((\lambda_2 - p(b))\Delta - R\right)\epsilon > 0 \tag{23}$$

using (22), without violating any incentive constraints. Therefore, we must have $\theta^{\emptyset}(bb) = 0$ so agent 2 is always promoted if $m = (bb)$.

The gb-variables only appear in the $\text{IC}_2(g)$ and $\text{IC}_1(b)$ constraints. Notice that $\text{IC}_1(b)$ must hold with equality: otherwise just lower $w_1^{\emptyset}(bb)$. Therefore, as $q > 1-q$, it is optimal to set $\theta^B(gb) = 0$ as it reduces total expected wage payments. Then $\text{IC}_2(g)$ binds at the optimum: otherwise it must be the case that $w_2^G(gg) > 0$ but then $w_2^G(gg)$ can be reduced. Therefore, as $q > 1-q$, it is optimal to set $\theta^G(gb) = 1$ as it minimizes expected wage payments (the principal cares about the sum of the right hand sides of $\text{IC}_1(b)$ and $\text{IC}_2(g)$). But then, $\text{IC}_2(g)$ is satisfied with $w_2^G(gg) = 0$ so this is optimal. From $\text{IC}_1(b)$ we obtain $w_1^{\emptyset}(bb) = p(b)R$.

The bg-variables only appear on the right hand side of $\text{IC}_2(b)$ and $\text{IC}_1(g)$. This constraints are satisfied at minimum cost if $w_2^{\emptyset}(bb) = w_1^G(gg) = 0$, $\theta^B(bg) = 0$ and $\theta^G(bg) = 1$. This completes the proof for case 2. ■

Proof of Proposition 4.

Suppose the inventor is ahead. We split the proof into two cases: (1) the inventor is well ahead, and (2) the inventor is not well ahead.

Case 1: Suppose the inventor is well ahead:

$$\lambda_2 < p(b) - \frac{R}{\Delta}. \tag{24}$$

Then, we claim the following is optimal:

$$\theta^{\emptyset}(bb) = \theta^G(gb) = \theta^B(bg) = 1$$

$$\theta^B(gb) = 0$$

$$\theta^G(bg) = \max\left\{0, 1 - \frac{(1-\lambda_1)(1-q)}{\lambda_1 q}\right\}$$

Each agent is paid a zero wage, *except* that $w_2^{\emptyset}(bb) = p(b)(1 - \theta^G(bg))R$, and if $(1-\lambda_1)(1-q) \geq \lambda_1 q$ then

$$w_1^G(gg) = \left(\frac{(1-\lambda_1)(1-q)}{\lambda_1 q} - 1\right) R$$

To show this, first consider increasing $\theta^{\emptyset}(bb)$ by $\epsilon > 0$ and recall that $h(bb) = 0$. This reduces the payoff of agent 2 by $R\epsilon$ as he is promoted now only with probability $(1 - \theta^{\emptyset}(bb) - \epsilon)$. Raise $w_2^{\emptyset}(bb)$ by $R\epsilon$, thus compensating agent 2 for the reduced probability of promotion. By (24), this changes the principal's payoff by

$$\begin{aligned} & (\lambda_1(1-q)(1-\lambda_2) - (1-\lambda_1)q\lambda_2)\,\epsilon\Delta \quad - (\lambda_1(1-q) + (1-\lambda_1)q)\,R\epsilon \\ = \;& (\lambda_1(1-q) + (1-\lambda_1)q)\,[(p(b) - \lambda_2)\,\epsilon\Delta \quad - R\epsilon] \geq 0 \end{aligned}$$

without violating any of the other constraints. Thus, set $\theta^{\emptyset}(bb) = 1$ from now on.

The bg - variables only appear on the right hand side of $\text{IC}_1(g)$ and $\text{IC}_2(b)$ constraints. $\text{IC}_2(b)$ will hold with equality, otherwise $w_2^{\emptyset}(bb)$ can be reduced ($w_2^{\emptyset}(bb) > 0$

as $\theta^{\emptyset}(bb) = 1$). We claim also $\text{IC}_1(g)$ holds with equality. If not, then $w_1^G(gg) = 0$, and the principal will set the $bg-$ variables to minimize the right hand side of $\text{IC}_2(b)$ as it is the only binding constraint involving the bg-variables. This implies:

$$1 - \theta^G(bg) = 1 - \theta^B(bg) = 0 \tag{25}$$

But, now $\text{IC}_1(g)$ is

$$\lambda_1 q\ R \geq \lambda_1 q R + (1-\lambda_1)(1-q)R$$

which is violated. Thus, $\text{IC}_1(g)$ holds with equality. By inspection of the principal's expected wage payments, we see that she cares about the *sum* of the left hand side of the $\text{IC}_1(g)$ and $\text{IC}_2(b)$ constraints. As both these constraints are satisfied with equality, she should set the bg- variables to minimize the sum of the right hand side of the $\text{IC}_1(g)$ and $\text{IC}_2(b)$ constraints, with the restriction that the right hand side of $\text{IC}_1(g)$ must exceed $\lambda_1 qR$ for otherwise equality in $\text{IC}_1(g)$ is incompatible with (11).

We claim $\theta^B(bg) = 1$ is optimal. In other words, promote agent 1 when he is "more likely to have told the truth". For if $\theta^B(bg) < 1$ then raising $\theta^B(bg)$ lowers the sum of expected wage payments as the right hand side of $\text{IC}_1(g)$ increases more slowly than the right hand side of $\text{IC}_2(b)$ falls by $q > 1-q$. Also, by a similar argument, the principal should set $\theta^G(bg)$ as low as possible. However, the right hand side of $\text{IC}_1(g)$ must exceed $\lambda_1 qR$. This yields two cases: (a) if $(1-\lambda_1)(1-q) \geq \lambda_1 q$ then set $\theta^G(bg) = 0$; (b) if $(1-\lambda_1)(1-q) < \lambda_1 q$, then set $\theta^G(bg) = 1 - \frac{(1-\lambda_1)(1-q)}{\lambda_1 q}$. In case (a), $w_1^G(gg) + R = \frac{(1-\lambda_1)(1-q)}{\lambda_1 q} R \geq R$ and in case (b) $w_1^G(gg) = 0$, as there is equality in $\text{IC}_1(g)$. Finally, $w_2^{\emptyset}(bb) = \frac{\lambda_1(1-q)}{\lambda_1(1-q)+(1-\lambda_1)q}(1-\theta^G(bg))R$, as there is equality in $\text{IC}_2(b)$.

Finally, by setting $\theta^G(gb) = 1$, $\theta^B(gb) = 0$, $\text{IC}_1(b)$ and $\text{IC}_2(g)$ are satisfied with $w_1^{\emptyset}(bb) = w_2^G(gg) = 0$, and this is clearly optimal. This completes the proof for Case 1.

Case 2: The inventor is not well ahead:

$$p(b) - \frac{R}{\Delta} < \lambda_2 < p(b). \tag{26}$$

Then, we claim the following is optimal:

$$\theta^{\emptyset}(bb) = 1 - p(b)(1 - \theta^G(bg))$$

$$\begin{aligned} \theta^G(gb) &= 1 \\ \theta^B(gb) &= 0 \end{aligned}$$

$$\theta^G(bg) = \begin{cases} 0 & \text{if } \frac{(1-\lambda_1)(1-q)}{\lambda_1 q} \geq 1 \\ 1 - \frac{(1-\lambda_1)(1-q)}{\lambda_1 q} & \text{otherwise} \end{cases}$$

$$\theta^B(bg) = \begin{cases} \frac{\lambda_1 q}{(1-\lambda_1)(1-q)} & \text{if } \frac{(1-\lambda_1)(1-q)}{\lambda_1 q} \geq 1 \text{ and } p(b) - \lambda_2 - \frac{1-q}{q}\frac{R}{\Delta} < 0 \\ 1 & \text{otherwise} \end{cases}$$

$$w_2^{\emptyset}(bb) = w_1^{\emptyset}(bb) = w_2^G(gg) = 0$$

$$w_1^G(gg) = \begin{cases} \left(\frac{(1-\lambda_1)(1-q)}{\lambda_1 q} - 1\right) R & \text{if } \frac{(1-\lambda_1)(1-q)}{\lambda_1 q} \geq 1 \text{ and } p(b) - \lambda_2 - \frac{1-q}{q}\frac{R}{\Delta} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Claim 1: It is optimal to set $w_1^{\emptyset}(bb) = 0$.

Proof : Suppose $w_1^{\emptyset}(bb) > 0$. Then if $\theta^{\emptyset}(bb) < 1$ the principal can increase $\theta^{\emptyset}(bb)$ by $\epsilon > 0$, reduce $w_1^{\emptyset}(bb)$ by ϵR and increase $w_2^{\emptyset}(bb)$ by ϵR. By (26), this increases her payoff by

$$\Delta\left(\lambda_1(1-q) + (1-\lambda_1)q\right)\left(p(b) - \lambda_2\right)\epsilon > 0 \tag{27}$$

without violating any incentive or limited liability constraints. Thus we can set $\theta^{\emptyset}(bb) = 1$. But then $\text{IC}_1(b)$ is also slack, and $w_1^{\emptyset}(bb)$ can be lowered without violating any constraint.

Claim 2: $\text{IC}_2(b)$ binds at the optimum.

Proof : Suppose not. Then $w_2^{\emptyset}(bb) = 0$ (or else the principal can lower $w_2^{\emptyset}(bb)$ without violating any constraints) and if $\text{IC}_2(b)$ is not binding then $\theta^{\emptyset}(bb) < 1$. But we can raise $\theta^{\emptyset}(bb)$ by $\epsilon > 0$ which raises the principal's welfare by (27), which is positive by (26).

Claim 3: $\text{IC}_1(g)$ binds at the optimum.

Proof: Suppose not. Then $w_1^G(gg) = 0$ (or else profit can be increased by reducing $w_1^G(gg)$) and $\theta^G(bg) < 1$. Hence, $\theta^G(bg)$ can be increased without violating $\text{IC}_1(g)$ while relaxing $\text{IC}_2(b)$. But when $\text{IC}_2(b)$ is relaxed the principal can be made better off as shown in claim 2.

Assume from now on that $IC_1(b)$ and $IC_2(g)$ do not bind. (We will show later that this is indeed the case.) Under this assumption, it is clearly optimal to set $w_2^G(gg) = w_1^{\emptyset}(bb) = 0$

Claim 4: It is optimal to set $w_2^{\emptyset}(bb) = 0$.

Proof : Suppose $w_2^{\emptyset}(bb) > 0$. As $\text{IC}_2(b)$ binds from Claim 2, $w_2^{\emptyset}(bb) + (1 - \theta^{\emptyset}(bb))R \leq R$, hence $\theta^{\emptyset}(bb) > 0$. Now $\theta^{\emptyset}(bb)$ can be decreased by ϵ and $w_2^{\emptyset}(bb)$ decreased by ϵR. (Recall we are neglecting $\text{IC}_1(b)$ and $\text{IC}_2(g)$). This increases the principal's payoff by

$$\epsilon\Delta\left(\lambda_1(1-q) + (1-\lambda_1)q\right)\left(\lambda_2 - p(b) + \frac{R}{\Delta}\right)$$

which is positive by (26).

Claim 5: Either $\theta^G(bg) = 0$ or $\theta^B(bg) = 1$.

Proof : Increasing $\theta^B(bg)$ and decreasing $\theta^G(bg)$ in such a way that the right hand side of $\text{IC}_1(g)$ is constant, relaxes $\text{IC}_2(b)$ while leaving all other constraints unchanged. This proves the claim.

Now we know that

$$w_2^G(gg) = w_1^{\emptyset}(bb) = w_2^{\emptyset}(bb) = 0 \tag{28}$$

and

$$(1 - \theta^{\emptyset}(bb))R = p(b)(1 - \theta^G(bg))R + (1 - p(b))\left(1 - \theta^B(bg)\right)R \tag{29}$$

from $\text{IC}_2(b)$ and

$$w_1^G(gg) + R = \theta^G(bg)R + \frac{(1-\lambda_1)(1-q)}{\lambda_1 q}\theta^B(bg)R \tag{30}$$

from $\text{IC}_1(g)$. Using these expressions we can write the principal's payoff as a function of only $\theta^B(bg)$ and $\theta^G(bg)$. Changing $\theta^G(bg)$ by ϵ changes profits by

$$\Delta\lambda_1(1-q)\left(p(b) - \lambda_2 - \frac{q}{1-q}\frac{R}{\Delta}\right)\epsilon$$

and the expression in parenthesis is negative by (26). Hence, it is optimal to lower $\theta^G(bg)$ as much as possible subject to (30) and $w_1^G(gg) \geq 0$. There are two cases.

Case a: $\frac{(1-\lambda_1(1-q)}{\lambda_1 q} < 1$. Then claim 3 together with $w_1^G(gg) \geq 0$ implies $\theta^G(bg) > 0$ and hence $\theta^B(bg) = 1$ from claim 5. The lowest $\theta^G(bg)$ we can set is $\theta^G(bg) = \frac{\lambda_1 q - (1-\lambda_1)(1-q)}{\lambda_1 q}$, and then $w_1^G(gg) = 0$.

Case b: $\frac{(1-\lambda_1)(1-q)}{\lambda_1 q} \geq 1$. Then we can lower $\theta^G(bg)$ to zero without violating $w_1^G(gg) \geq 0$, so $\theta^G(bg) = 0$ is optimal. For $\theta^B(bg)$ there are two possibilities. If $p(b) - \lambda_2 - \frac{1-q}{q}\frac{R}{\Delta} \geq 0$ then an increase in $\theta^B(bg)$ by $\epsilon > 0$ increases the principal's profit by

$$(1-\lambda_1)q\Delta\left[p(b) - \lambda_2 - \frac{1-q}{q}\frac{R}{\Delta}\right]\epsilon$$

so $\theta^B(bg) = 1$ is optimal, with

$$w_1^G(gg) = \left(\frac{(1-\lambda_1)(1-q)}{\lambda_1 q} - 1\right)R$$

from (30). If $p(b) - \lambda_2 - \frac{1-q}{q}\frac{R}{\Delta} < 0$ then the principal's profit is decreasing in $\theta^B(bg)$ and the optimal $\theta^B(bg)$ is the lowest possible, subject to $w_1^G(gg) \geq 0$. Using (30) this gives $\theta^B(bg) = \frac{\lambda_1 q}{(1-\lambda_1)(1-q)}$ and $w_1^G(gg) = 0$.

Finally, it can be checked that by setting $\theta^G(gb) = 1$ and $\theta^B(gb) = 0$, the omitted constraints are automatically satisfied. In fact, $\text{IC}_2(g)$ is trivial and $\text{IC}_1(b)$ becomes $\theta^{\emptyset}(bb) \geq p(b)$. Using (29), this requires

$$\theta^{\emptyset}(bb) = p(b)\theta^G(bg) + (1-p(b))\,\theta^B(bg) \geq p(b) \tag{31}$$

and it is easy to check that $q > 1-q$ implies that (31) is satisfied in both case a and case b above. ■

6.2 Optimal renegotiation-proof contracts

Even when contracts are required to be renegotiation proof, an argument along the lines of Lemma 1 shows that $h(bb) = 0$. Suppose the program in Section 3 is modified by not imposing $h(gg) = h(gb) = h(bg) = 1$, but instead imposing (12). We can assume without loss of generality that $w_1^{\emptyset}(gg) = 0$. Otherwise, if $h(gg) > 0$, increase

$w_1^G(gg)$ to keep expected wage payments to agent 1 when $m = gg$ constant. This leaves the principal's payoff and the left hand side of $IC_1(g)$ unchanged. If $h(gg) = 0$, the principal is simply not implementing any projects and can set all wages equal to zero. A similar argument also establishes that we can set $w_2^{\emptyset}(gg) = 0$ without loss of generality.

Therefore, the program becomes: maximize the principal's expected payoff

$$\begin{aligned}
&h(gg)\left(\lambda_1 q\left(G+\Delta-w_1^G(gg)-w_2^G(gg)\right)+\ (1-\lambda_1)(1-q)\,(B+\lambda_2\Delta)\right)\\
&+(1-h(gg))\left(\lambda_1 q\left(\theta^{\emptyset}(gg)\Delta+(1-\theta^{\emptyset}(gg))\lambda_2\Delta\right)+(1-\lambda_1)(1-q)(1-\theta^{\emptyset}(gg))\lambda_2\Delta\right)\\
&\qquad+\lambda_1(1-q)\left(\theta^{\emptyset}(bb)\Delta+(1-\theta^{\emptyset}(bb))\lambda_2\Delta-w_1^{\emptyset}(bb)-w_2^{\emptyset}(bb)\right)\\
&\qquad+(1-\lambda_1)q\left((1-\theta^{\emptyset}(bb))\lambda_2\Delta-w_1^{\emptyset}(bb)-w_2^{\emptyset}(bb)\right)
\end{aligned}$$

subject to

$\mathrm{IC}_1(g)$:

$$\begin{aligned}
&h(gg)\ \lambda_1 q\left(w_1^G(gg)+R\right)\\
&+(1-h(gg))\,(\lambda_1 q+(1-\lambda_1)(1-q))\,\theta^{\emptyset}(gg)R\\
\geq\ &h(bg)\lambda_1 qR+(1-h(bg))\,(\lambda_1 q+(1-\lambda_1)(1-q))\,\theta^{\emptyset}(bg)R
\end{aligned}\tag{32}$$

$\mathrm{IC}_1(b)$:

$$\begin{aligned}
&(\lambda_1(1-q)+(1-\lambda_1)q)\left(w_1^{\emptyset}(bb)+\theta^{\emptyset}(bb)R\right)\\
\geq\ &h(gb)\lambda_1(1-q)R+(1-h(gb))\,(\lambda_1(1-q)+(1-\lambda_1)q)\,\theta^{\emptyset}(gb)R
\end{aligned}\tag{33}$$

$\mathrm{IC}_2(g)$:

$$\begin{aligned}
&h(gg)\left(\lambda_1 q w_2^G(gg)+(1-\lambda_1)(1-q)R\right)\\
&+(1-h(gg))\,(\lambda_1 q+(1-\lambda_1)(1-q))\,R(1-\theta^{\emptyset}(gg))\\
\geq\ &h(gb)(1-\lambda_1)(1-q)R+(1-h(gb))\,(\lambda_1 q+(1-\lambda_1)(1-q))\,R(1-\theta^{\emptyset}(gb))
\end{aligned}\tag{34}$$

$\mathrm{IC}_2(b)$:

$$\begin{aligned}
&(\lambda_1(1-q)+(1-\lambda_1)q)\left(w_2^{\emptyset}(bb)+(1-\theta^{\emptyset}(bb))R\right)\\
\geq\ &h(bg)(1-\lambda_1)qR+(1-h(bg))\,(\lambda_1(1-q)+q(1-\lambda_1))\,(1-\theta^{\emptyset}(bg))R
\end{aligned}\tag{35}$$

and the limited liability constraints:

$$w_1^G(gg), w_2^G(gg), w_1^{\emptyset}(bb), w_2^{\emptyset}(bb) \geq 0 \tag{36}$$

Proof of Lemma 5

We claim that if the inventor is ahead but not well ahead, then maximizing the principal's expected payoff subject to the IC constraints (32)-(35) and limited liability results in: $h(bg) = 0$, $h(gg) = h(gb) = 1$. If (15) holds then $\theta^{\emptyset}(bb) = \theta^{\emptyset}(bg) = 1$ and

$$w_1^G(gg) = \frac{1 - p(g)}{p(g)} R$$

Otherwise, $\theta^{\emptyset}(bb) = \theta^{\emptyset}(bg) = p(g)$ and $w_1^G(gg) = 0$. All other wages are zero.

We shall temporarily omit $\mathrm{IC}_1(b)$ and $\mathrm{IC}_2(g)$ from the program and show later that they are satisfied. Without these constraints $w_1^{\emptyset}(bb) = w_2^G(gg) = 0$ is certainly optimal.

Claim 1. It is optimal to set $h(gg) = 1$.

Proof. If $h(gg) < 1$ then increase $h(gg)$ by ϵ and, if possible, alter $w_1^G(gg)$ to keep the left hand side of $IC_1(g)$ constant. This will change the expected wage payments, which contains the component $\lambda_1 q h(gg) w_1^G(gg)$. Let ϕ denote the *change* in this component due to the changes in $h(gg)$ and $w_1^G(gg)$. There are two cases:

Case 1: If (i) $\theta^{\emptyset}(gg) \geq p(g)$ or (ii) if $\theta^{\emptyset}(gg) < p(g)$ and $w_1^G(gg) > 0$, the change in expected wages $\lambda_1 q h(gg) w_1^G(gg)$ needed to keep the left hand side of $IC_1(g)$ constant is

$$\begin{aligned} \phi &= \epsilon \left(\lambda_1 q + (1-\lambda_1)(1-q)\right) \theta^{\emptyset}(gg) R - \epsilon \lambda_1 q R \\ &= \epsilon (1-\lambda_1)(1-q) \theta^{\emptyset}(gg) R - \epsilon \lambda_1 q (1 - \theta^{\emptyset}(gg)) R \end{aligned} \tag{37}$$

In case (i), this expression is non-negative. In case (ii), it is negative, so $w_1^G(gg)$ may have to be reduced, but this is possible since $w_1^G(gg) > 0$. In any case, the principal's expected payoff changes by

$$\begin{aligned} &\epsilon \left(\lambda_1 q \left(G + \Delta\right) + \ (1-\lambda_1)(1-q) \left(B + \lambda_2 \Delta\right)\right) - \epsilon \lambda_1 q \theta^{\emptyset}(gg) \Delta \\ &-\epsilon \left(\lambda_1 q + (1-\lambda_1)(1-q)\right) (1 - \theta^{\emptyset}(gg)) \lambda_2 \Delta - \phi \end{aligned} \tag{38}$$

Substituting from (37), (38) becomes

$$\begin{aligned} & \epsilon \left(\lambda_1 q \left(G + \Delta\right) + \ (1-\lambda_1)(1-q) \left(B + \lambda_2 \Delta\right)\right) - \epsilon \lambda_1 q \theta^{\emptyset}(gg) \Delta \\ & -\epsilon \left(\lambda_1 q + (1-\lambda_1)(1-q)\right) (1 - \theta^{\emptyset}(gg)) \lambda_2 \Delta \\ & -\epsilon (1-\lambda_1)(1-q) \theta^{\emptyset}(gg) R + \epsilon \lambda_1 q (1 - \theta^{\emptyset}(gg)) R \\ = \ & \epsilon \left(\lambda_1 q G + \ (1-\lambda_1)(1-q) B\right) \\ & +(1-\lambda_1)(1-q) \theta^{\emptyset}(gg) \left(\lambda_2 \Delta - R\right) + \epsilon \lambda_1 q (1 - \theta^{\emptyset}(gg)) R \\ & +\epsilon \lambda_1 q \left(1 - \theta^{\emptyset}(gg)\right) \Delta (1 - \lambda_2) \end{aligned}$$

which is strictly positive by Assumption 1. Hence, it is optimal to raise $h(gg)$.

Case 2: If $\theta^{\emptyset}(gg) < p(g)$ and $w_1^G(gg) = 0$, the left hand side of $IC_1(g)$ will increase when $h(gg)$ is increased by ϵ. As $w_1^G(gg) = 0$, this perturbation changes the maximand by

$$\begin{aligned}
&\epsilon\left(\lambda_1 q\left(G+\Delta\right)+\ (1-\lambda_1)(1-q)\left(B+\lambda_2\Delta\right)\right)-\epsilon\lambda_1 q\theta^{\emptyset}(gg)\Delta \quad (39)\\
&-\epsilon\left(\lambda_1 q+(1-\lambda_1)(1-q)\right)(1-\theta^{\emptyset}(gg))\lambda_2\Delta\\
=\ &\epsilon\left(\lambda_1 qG+\ (1-\lambda_1)(1-q)B\right)+\epsilon\lambda_1 q\left(1-\theta^{\emptyset}(gg)\right)\Delta\left(1-\lambda_2\right)\\
&+(1-\lambda_1)(1-q)\theta^{\emptyset}(gg)\lambda_2\Delta \quad (40)
\end{aligned}$$

which is positive by Assumption 1. Hence, it is also optimal to increase $h(gg)$ in Case 2 too, and the claim is proved.

Claim 2. $\mathrm{IC}_1(g)$ and $\mathrm{IC}_2(b)$ bind, and either $\theta^{\emptyset}(bb)=1$ or $w_1^{\emptyset}(bb)=0$.

Proof. Suppose $\mathrm{IC}_2(b)$ does not bind. Then $w_2^{\emptyset}(bb)=0$ or else $w_2^{\emptyset}(bb)$ could be lowered, and hence $(1-\theta^{\emptyset}(bb))R>0$. Now raise $\theta^{\emptyset}(bb)$ by $\varepsilon>0$. This respects all constraints and increases the principal's payoff by

$$\begin{aligned}
&\varepsilon\Delta\left(\lambda_1(1-q)-\ \left(\lambda_1(1-q)+(1-\lambda_1)q\right)\lambda_2\right) \quad (41)\\
=\ &\left(\lambda_1(1-q)+(1-\lambda_1)q\right)\ \varepsilon\Delta\left(p(b)-\ \lambda_2\right)>0
\end{aligned}$$

since the inventor is ahead. Therefore, $\mathrm{IC}_2(b)$ must bind.

Suppose $\mathrm{IC}_1(g)$ does not bind and recall from Claim 1 that $h(gg)=1$. Then, $w_1^G(gg)=0$. Except for the slack constraint $\mathrm{IC}_1(g)$, $h(bg)$ and $\theta^{\emptyset}(bg)$ only enter IC_2(b) which binds. Reducing the right hand side of $\mathrm{IC}_2(b)$ is advantageous because the principal can either lower $w_2^{\emptyset}(bb)$ or raise $\theta^{\emptyset}(bb)$ (the latter is strictly advantageous from (41)). Therefore, if $\mathrm{IC}_1(g)$ is slack the right hand side of $\mathrm{IC}_2(b)$ must be zero, which implies $\theta^{\emptyset}(bg)=1$ and $h(bg)=0$. However, together with $w_1^G(gg)=0$ this violates $\mathrm{IC}_1(g)$. Therefore $\mathrm{IC}_1(g)$ must bind.

Finally, suppose $\theta^{\emptyset}(bb)<1$ and $w_1^{\emptyset}(bb)>0$. Then, by raising $\theta(bb)$ by ε, reducing $w_1^{\emptyset}(bb)$ by ϵR and increasing $w_2^{\emptyset}(bb)$ by ϵR, the sum of the wages is constant, all constraints are respected, and the principal payoff goes up by (41). This proves the claim.

Claim 3. At the optimum, $\theta^{\emptyset}(bb)>0$ and $w_2^{\emptyset}(bb)=0$.

Proof: If $\theta^{\emptyset}(bb)=0$, then as $\mathrm{IC}_2(b)$ binds we must have $h(bg)=\theta^{\emptyset}(bg)=0$. Then the right hand side of $\mathrm{IC}_1(g)$ is zero, but this contradicts the fact that $\mathrm{IC}_1(g)$ is binding and $h(gg)=1$. Thus, $\theta^{\emptyset}(bb)>0$.

Now suppose $w_2^{\emptyset}(bb)>0$. As $\theta^{\emptyset}(bb)>0$ we can lower $\theta^{\emptyset}(bb)$ by ε and $w_2^{\emptyset}(bb)$ by ϵR without violating any constraints (recall we are omitting $\mathrm{IC}_1(b)$ from the program). The principal's expected payoff goes up by

$$\varepsilon\Delta\left(\lambda_1(1-q)+(1-\lambda_1)q\right)\ \left(-p(b)+\lambda_2+\frac{R}{\Delta}\right)>0$$

since the inventor is not well ahead. This proves the claim.

¿From $\mathrm{IC}_2(b)$ we now have

$$1-\theta^{\emptyset}(bb)=h(bg)(1-p(b))+(1-h(bg))(1-\theta^{\emptyset}(bg)) \quad (42)$$

and from $\mathrm{IC}_1(g)$ as $h(gg) = 1$

$$w_1^G(gg) = (1 - h(bg)) \left[\frac{\theta^{\emptyset}(bg)}{p(g)} - 1 \right] R \geq 0 \tag{43}$$

Omitting all constants from the principal's expected payoff, setting $w_1^{\emptyset}(bb) = w_2^G(gg) = w_2^{\emptyset}(bb) = 0$, and adding the constant term $-\lambda_1(1-q)\Delta$ we obtain

$$\begin{aligned} & -\lambda_1 q w_1^G(gg) + \lambda_1(1-q)(1-\theta^{\emptyset}(bb))(\lambda_2 - 1)\Delta + (1-\lambda_1)q(1-\theta^{\emptyset}(bb))\lambda_2\Delta \\ = \quad & -\lambda_1 q w_1^G(gg) + (1-\theta^{\emptyset}(bb))\Delta\left(\lambda_1(1-q)(\lambda_2-1) + (1-\lambda_1)q\lambda_2\right) \end{aligned}$$

Next, using (42), the principal's maximand becomes

$$\begin{gathered} -\lambda_1 q w_1^G(gg) + \left(h(bg)(1-p(b))\right. \\ \left. +(1-h(bg))(1-\theta^{\emptyset}(bg))\right)\left(\lambda_1(1-q)(\lambda_2-1) + (1-\lambda_1)q\lambda_2\right)\Delta \end{gathered}$$

and (up to constants) and using (43) this the same as

$$\begin{aligned} & -\lambda_1 q(1-h(bg)) \left[\frac{\theta^{\emptyset}(bg)}{p(g)} - 1 \right] R \\ & \qquad +(1-h(bg))(p(b) - \theta^{\emptyset}(bg))\left(\lambda_1(1-q)(\lambda_2-1) + (1-\lambda_1)q\lambda_2\right)\Delta \\ = \quad & -(1-h(bg))\lambda_1 \left[q\left[\frac{\theta^{\emptyset}(bg)}{p(g)} - 1 \right] R + (1-q)\left(\lambda_2 - p(b)\right)\left[\frac{\theta^{\emptyset}(bg)}{p(b)} - 1 \right]\Delta \right] \end{aligned}$$

The derivative with respect to $\theta^{\emptyset}(bg)$ is positive if (15) holds, so in this case $\theta^{\emptyset}(bg) = 1$. Then, the derivative with respect to $h(bg)$ is

$$\begin{aligned} & \lambda_1 q \left[\frac{1}{p(g)} - 1 \right] R + \lambda_1(1-q)\left(\lambda_2 - p(b)\right)\left[\frac{1}{p(b)} - 1 \right]\Delta \\ < \quad & \lambda_1 q \left[\frac{1}{p(g)} - \frac{p(b)}{p(g)} \right] R + \lambda_1(1-q)\left(\lambda_2 - p(b)\right)\left[\frac{1}{p(b)} - 1 \right]\Delta \\ = \quad & \lambda_1(1-p(b)) \left[q\frac{1}{p(g)} R + (1-q)\left(\lambda_2 - p(b)\right)\frac{1}{p(b)}\Delta \right] \end{aligned}$$

which is negative by (15). Thus $h(bg) = 0$.

If (15) does not hold then instead it is optimal to set $\theta^{\emptyset}(bg) = p(g)$ (the minimum $\theta^{\emptyset}(bg)$, from (43)). Again it can be checked that $h(bg) = 0$ is optimal.

Since we have shown that $h(bg) = 0$, (42) implies $\theta^{\emptyset}(bg) = \theta^{\emptyset}(bb)$.

Finally, $\mathrm{IC}_1(b)$ and $\mathrm{IC}_2(g)$ are satisfied at zero cost by setting $h(gb) = 1$. ■

Proof of Lemma 6

Suppose the inventor is well ahead. We need to show that the problem of maximizing the principal's expected payoff subject to (32)-(36), and with the extra constraint

$\theta^{\emptyset}(m) = 1$ for all m, is solved by $h(gb) = 1$ and $h(bg) = 0$. All wages are zero, *except* that

$$w_1^G(gg) = \frac{1-p(g)}{p(g)}R$$

We shall omit $\text{IC}_1(b)$ and $\text{IC}_2(g)$ from the program, and later show they are satisfied. As in the proof of Lemma 5 one can show that $\text{IC}_1(g)$ and $\text{IC}_2(b)$ must bind. Then $w_1^{\emptyset}(bb) = w_2^G(gg) = 0$ and

$$w_1^G(gg) = (1-h(bg))\frac{1-p(g)}{p(g)}R$$

$$w_2^{\emptyset}(bb) = h(bg)\,(1-p(b))\,R \tag{44}$$

As $\theta^{\emptyset}(m) = 1$ for all m, the principal's expected payoff is, omitting constants:

$$\begin{aligned} & -\lambda_1 q w_1^G(gg) - (\,\lambda_1(1-q) + (1-\lambda_1)q)\, w_2^{\emptyset}(bb) \\ = \quad & -(1-\lambda_1)\,((1-h(bg))(1-q) + h(bg)q)\,R \end{aligned} \tag{45}$$

so that $h(bg) = 0$ is optimal. Moreover, by setting $h(gb) = 1$ we guarantee that $\text{IC}_1(b)$ and $\text{IC}_2(g)$ hold. ■

References

[1] Sandeep Baliga and Tomas Sjöström (1997), "Not Invented Here", Harvard Discussion Paper No. 1797

[2] George Baker, Robert Gibbons and Kevin J. Murphy (1994), "Subjective Performance Measures in Optimal Incentive Schemes," *Journal of Political Economy,* 109, 1125-56

[3] Guillermo A. Calvo and Stanislaw Wellisz (1979), "Hierarchy, Ability and Income Distribution," *Journal of Political Economy*, 87, 991-1010.

[4] H. Lorne Carmichael (1988), "Incentives in Academics: Why is there Tenure?," *Journal of Political Economy*, 96, 453-472.

[5] James S. Coleman (1990), *Foundations of Social Theory.* Harvard University Press, Cambridge

[6] Mathias Dewatripont and Eric Maskin (1995), "Contractual Contingencies and Incomplete Contracts," *RAND Journal of Economics,* 26, 704-719

[7] R. Donald Gamache and Robert Lawrence Kuhn (1989), "The Creativity Infusion." New York: Harpers and Row

[8] Bengt Holmström (1982), "Moral Hazard in Teams," *Bell Journal of Economics,* 13: 324-340.

[9] Bengt Holmström (1999), "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies,* 66: 169-182.

[10] Hideshi Itoh (1991), "Incentives to Help in Multi- Agent Situations", *Econometrica* 59:611-636

[11] Edward P. Lazear (1989) "Pay, Inequality and Industrial Politics", *Journal of Political Economy,* 97: 561-80

[12] Steven D. Levitt (1995), "Optimal Incentive Schemes when only the Agents' Best Output Matters," *RAND Journal of Economics*, 24: 744-760.

[13] Steven D. Levitt and Christopher M. Snyder (1997), "Is No News Bad News? Information Transmission and the Role of 'Early Warning' in the Principal-Agent Model," *RAND Journal of Economics,* 28: 641-61.

[14] Eric Maskin and Jean Tirole (1992),"The Principal- Agent Relationship with an Informed Principal, II: Common Values," *Econometrica,* 60, 1-42

[15] Paul Milgrom (1988), "Employment Contracts, Influence Activities and Efficient Organization Design," *Journal of Political Economy*, 96: 42-60

[16] Max Planck (1949), *Scientific Autobiography and Other Papers.* New York: Philosophical Library.

[17] Canice Prendergast and Lars Stole (1996), "Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning," *Journal of Political Economy*, 104:1105-1134

[18] Canice Prendergast and Robert H. Topel (1996), "Favoritism in Organizations," *Journal of Political Economy*, 104, 958- 978

[19] Julio Rotemberg and Garth Saloner (1995), "Overt Interfunctional Conflict (and its Reduction through Business Strategy)", *RAND Journal of Economics*, Winter 1995, 26: 630-653

[20] Jean Tirole (1992), "Collusion and the Theory of Organizations," in *Advances in Economic Theory: Sixth World Congress*, vol. 2, edited by Jean-Jacques Laffont. Cambridge: Cambridge University Press