Subjective Performance of Patent Examiners and Implicit Contracts

Corinne Langinier* and Philippe Marcoul†

October 2012
– Preliminary Draft –

Abstract

Patent offices should be concerned with both patent quality and quantity. On the one hand, more patent applications generate more revenue through fees paid by patent applicants, which is the main source of revenue for self-funded patent offices. On the other hand, patents should be granted to only deserving innovations and, thus, patent offices should be concerned with quality. In a context where both patent quantity and quality are accounted for, we investigate the impact of different bonus systems on the examiners’ incentive to prosecute patent applications. We consider contracts in which a patent office can offer bonuses on quantity quotas (explicit contract) and on quality outcome (either an implicit contract or an explicit contract based on a quality proxy). An examiner makes costly unobservable efforts to achieve these two goals. We find that when only explicit contracts are offered, the patent office is less eager to rely on a quality proxy when this measure becomes noisier. This creates a distortion in the provision of efforts, which is partly corrected when implicit contracts can be offered. Overall, the benefit of the patent office is a non-monotonic function of the noise.

Keywords: Patents; Examiners

JEL classification: D23, D86, O34 (Intellectual Property Rights)

*University of Alberta, Edmonton. Corinne.Langinier@ualberta.ca
†University of Alberta, Edmonton. marcoul@ualberta.ca

We would like to thank Talia Bar and participants in the seminars at University of Calgary, Wilfrid Laurier University, and participants at CEA and IIOC for many helpful comments. All remaining errors are ours.
1 Introduction

The patent system is in crisis.\(^1\) In order to improve the quality of the patent prosecution process, it is fundamental to have a better understanding of its internal functioning. Patent examiners are key players as they grant temporary monopoly rights to applicants. It is therefore important to analyze their patenting behaviors and the structure of their incentives.

In an effort to improve the quality of the patenting process, many reforms have been proposed. Among the most recent ones, as of February 2010, the U.S. Patent and Trademark Office (PTO) has started to implement a new examiner count system to evaluate examiner’s work.\(^2\) Until then, examiners’ production quotas were based on outdated criteria established in the 1970s, and the PTO has been urged to modernize evaluations of patent examiners (GAO 2007, 2008). As of October 2010, the USPTO has announced that it is adopting new, more comprehensive procedures for measuring the quality of patent examination.\(^3\)

When setting up examiners’ rewards, the PTO should be concerned with both quantity and quality of patents. On the one hand, more patent applications generate more gross revenue as more fees are paid by applicants. The PTO is a self-funded agency that relies essentially on fees.\(^4\) On the other hand, the PTO is supposed to grant patents to only new, non-obvious and useful innovations and, therefore, should be concerned with quality of patents.

During the patent prosecution process, an examiner responsible of a particular patent application must assess the novel content of the innovation. To determine whether it satisfies the patentability requirements, he has to search for prior art information, which corresponds to the set of existing innovations prior to the filing of the patent application. At any point in time, an examiner has to treat new cases and to dispose of ongoing cases. Therefore, there is an implicit trade-off between the effort that the examiner is putting into searching for potentially invalidating information (which is related to quality) and the effort to put into processing more applications (which is related to quantity).

---

\(^1\) See for instance Bessen and Meurer (2008) or the U.S. Chamber of Commerce (2009).
\(^2\) In February 2010, the USPTO has published a press release (http://www.uspto.gov/news/pr/2010/10_08.jsp) on its website.
\(^4\) In 2009 in the U.S., patents and trademarks generated total revenue of $ 1,927 million. At the same time, the functioning costs of the USPTO were $1,981 millions, 70% of which were labor costs.
In this paper we provide a theoretical framework to analyze contracts that the PTO might offer to patent examiners. We consider multitask examiners who are offered (explicit and/or implicit) contracts based on different (objective or subjective) performance measures linked to quantity and quality. In this setting, we investigate how contracts based on subjective or objective performance measures affect the provision of efforts of patent examiners.

We consider a model with two players: the PTO and an examiner. The PTO offers a salary scheme to the examiner based on two tasks: task 1, related to quantity, is to process more applications and task 2, related to quality, is to search for invalidating information. Nevertheless, as task 2 is non-contractible, the PTO can either use a proxy (e.g., number of legally disputed patents, number of citations made by the examiner in the patent), or rely on implicit contracting (where the quality screening is done by a more senior patent examiner). The examiner incurs costly efforts: an effort on task 1 (processing more information) and one on task 2 (searching for invalidating information). Efforts are not observable by the PTO, only outcomes in terms of both quantity and quality are observed.

In this setting, we find that when a bonus is offered for each task and if a proxy is used for task 2, the PTO is less eager to rely on the performance measure on task 2 when this measure becomes noisier. In fact, the use of a proxy on task 2 introduces a distortion in the provision of efforts as the examiner undersupplies effort on task 2. The distortion is worsened when we include limited liability on the part of the examiner.

Instead of relying on a proxy for task 2, the PTO and the examiner could rely on an implicit contract in which the PTO promises a bonus to the examiner if the quality of the screening is satisfactory (as defined by peer review). In this setting, we find that implicit contracts correct the distortion induced by the performance measurement noise on task 2. Depending on the noise of the measure of performance on task 2, all contracts are explicit (little noise), or the examiner is rewarded on task 2 through an implicit contract (for a noisy measure). The benefit of the PTO is a non-monotonic function of the noise measure.

Even though there is an abundant patent literature, only recently attention has been brought to the patent prosecution system. From a theoretical viewpoint, a few contributions analyze the strategic behavior of examiners. The granting of questionable patents may be due to the poor
knowledge of relevant prior art – the existing set of related inventions. Langinier and Marcoul (2011) propose a model of a bilateral search of information in which both innovator and examiner provide prior art information to prove the novel content of an innovation. The main focus is on the strategic behavior of patent applicants while searching and revealing information. Caillaud and Duchêne (2011) analyze also the determinant of patent quality but they are essentially concerned with the “overload” problem faced by the patent office. Atal and Bar (2010) study the incentives of innovators to search for prior art before and after undertaking any R&D investments. As in most of the patent literature, these contributions assume that both PTO and examiners have identical objective functions, and there is no strategic behavior on the part of patent examiners. However, it seems realistic to consider that they have different incentives, as the PTO is a federal self-funded institution for which many examiners are working with, most likely, different objective functions. To the best of our knowledge, there exist no theoretical contribution that analyzes the PTO’s incentive problems.

A growing number of empirical studies have started to open the “black box” of the process by which patents are granted. A few surveys have been conducted in patent offices. Friebel et al. (2006) provide an analysis of the objectives of the European Office, and the nature of its internal organization. They study the way internal organization shapes the ability of the patent office to pursue these objectives. Cockburn, Kortum and Stern (2003) propose a detailed exposition of the U.S. patent examination process, and they provide empirical evidence of the existence of heterogeneity among patent examiners and in the examination process. Sampat (2005), and Alcacer and Gittelman (2006) provide empirical evidence of the role played by patent applicants and examiners in revealing information regarding prior art. Sampat (2005) finds strong evidence that in some fields examiners are less informed than applicants and face particular challenges in searching for information. On the other hand, Alcacer and Gittelman (2006) show that many citations (prior art information) are listed by examiners, and that firm-level effects (e.g., experience of applicants, nationality) seem to explain most of the variance of examiner citation shares (Alcacer, Gittelman and Sampat, 2009). These analyses suggest that patent applicants and examiners do not necessarily have the same information, or the same incentives to search and reveal pertinent information. In this paper we take the information provided by the applicant as given and we focus on the examiners’ incentive to prosecute patents.
Subjective performance measures have been widely analyzed in contract theory literature as objective performance measures are not always available or are imperfect (Baker, Gibbons and Murphy, 1994; MacLeod and Malcomson, 1998). When rewards depend only on imperfect measures, workers’ incentives are not perfectly aligned with the principal’s objective. This problem may be mitigated by using subjective performance measures (non-verifiable and non-contractible). However, agreements based on subjective performance evaluation have to be self-enforcing as they cannot be part of an enforceable in court contract. In the context of a repeated principal agent model with one task that can be objectively or subjectively assessed, Baker and al. (1994) show that both explicit and implicit contracts can be complements or substitutes. In an efficient contract, the weight given to one performance measure should be decreasing in the precision and sensitivity of the other measures. We built on Baker and al. (1994) model in which we consider that two tasks must be performed, each of which has a different weight for the principal.

In a recent contribution, Schottner (2008) considers a repeated principal agent model in which one or two agents can perform three tasks whose contribution to the firm is non verifiable. In her model, the three tasks contribute to the unique outcome that defines the firm’s value. In our setting, each task contributes differently to the principal outcome. Multitasking has been analyzed by MacDonald and Marx (2001) for instance, but they do not consider any implicit contract.

The paper is organized as follows. In section 2 we present the patent prosecution process and the incentives awards of patent examiners. The model is introduced in section 3. Section 4 is devoted to the analysis of different incentive schemes offered by the PTO. In section 5 we analyze under what circumstances implicit contract will be favored to explicit contracts. Section 6 concludes.

2 Patent prosecution process and incentive awards

The patent prosecution process and the incentives awards of patent examiners presented in this section correspond to the U.S. system which is very well documented (e.g., see NAPA, 2005).

Once a patent application is received at the U.S. Patent and Trademark Office (USPTO),
after completion of administrative checks it is sent to an Art Unit (group of examiners with similar expertise) where it is processed by a Supervisory Patent Examiner (SPE) who attributes the patent application to a particular examiner in his group. The examiner, who is responsible for the patent application, must then determine whether the innovation is new, novel and non-obvious. To do so, he has to search for prior art information which corresponds to existing innovation prior to the filing of the patent application.

2.1 Patent prosecution process

The patent prosecution process can be divided into four tasks: search, examination, amendment review and post examination (NAPA, 2005). During the search task, the examiner reviews the application, searches for prior art information (in general the examiner starts with a search of patented innovations closely related to the claimed invention, and also foreign patents and non patented information), and analyzes the claims. The examination task requires that the examiner compares the invention described in the patent application to the prior art search results, and to the prior art provided by the applicant. He must determine the proper scope of the invention claimed by the inventor. Then, he prepares the first action letter (called FAOM, first action on the merits) in which he either allows the patent to be granted or proposes a non-final rejection (such that the applicant can respond to it), and he submits the FAOM to his superior for review if needed. If there is an amendment to a non-final rejection, the examiner searches for more information and writes the second and, generally the last action letter in which he rejects or grants the patent. Lastly, in the post examination stage, the examiner finalizes the disposal of the patent application.

Examiners work closely with applicants (or attorneys) to either narrow the scope of the patent application, or to split up the invention and file separate applications. It can take years before the final action (approval or rejection).

2.2 Incentive awards

The examiners award system has been in place since 1976 and is mainly based on the number of patent applications processed. New hires usually enter at around level \( GS - 7 \) (general schedule) pay scale, whereas experienced examiners are at level \( GS - 9 \) or above.
A production expectancy goal based on the technological field and the experience of the
examiner is set for each examiner. Factors have been established at the PTO to account for
experience. For instance, for the same application (same technological field) the production
expectancy goal of a GS – 7 examiner will be 39.3 hours, whereas it will be 27.5 hours for an
examiner at the level GS – 12, and 20.4 hours for a GS – 14 examiner.

An examiner is given credit (or count) at two different times during the examination process:
when a patent application is first examined (new case with a first action letter), and when it
is disposed of (allowed, rejected, abandoned). Once the quality of either of these actions is
approved by a SPE, the examiner receives credit for it. One production unit corresponds to two
counts. Expected annual productivity is calculated by assuming that 80% of the 2,080 hours in a
52 work week per year of 40 hours per week will be spent examining applications (NAPA, 2005).
For instance, an examiner with a goal of 31.6 hours per application would need to complete 53
applications or 106 counts, whereas an examiner with a goal of 14.3 hours per application needs
to complete 116 applications or 232 counts. On average, 87 patent applications per year must
be processed per examiner.

As his career progresses, an examiner is expected to examine more cases. The production
quotas system is measured by two different equations depending on experience. The balance
disposal (BD) of experience examiners is $BD = (N + D)/2$ and a new hire $BD = (2N + D)/3$
where $N$ is a new case and the first office action taken has with the case, and $D$ is
disposal (allowance, rejection or abandonment of the case). Each work period is two weeks (that
corresponds to the case deadline) but the measure of each examiner production comes quarterly.

Examiners get three different types of incentive awards: an annual gain-sharing award (which
accounts for 1 to 6% of base salary), a special achievement award (which accounts for 3% of
base salary) and a pendency reduction award (up to 1%). The first award corresponds to the
promotion to the next GS level. Examiners have non competitive promotion possibility at the
potential rate of one GS level per year up to GS-13. In order to get promoted they need to
exceed their production quotas by more than 10% on average over the fiscal year, with few
errors (the evaluation process is described later). In the period 1999-2003 between 60% to 73%
of patent examiners got promoted (NAPA, 2005). The second award corresponds to a bonus
that examiners get whenever they exceed their production goals by at least 10% on average over
four consecutive quarters. If their production goals is above 110%, they get a bonus of 5% of current salary, over 120 it is a 7% and over 130% it is a 9% bonus. Within the period 1999 and 2003, between 63% and 77% of patent examiners got a bonus (NAPA, 2005). According to NAPA (2005), the last reward (related to workflow management averaged over two consecutive quarters) seems to be based on obsolete measures and only concerns a small group of examiners as only 28% to 44% of examiners got this award between 1999 and 2003.

Examiners who consistently fail to meet their quotas get first an oral warning, then a written warning and eventually, they can be dismissed. In 2004, 329 oral warnings have been issued, 48 written warnings and 17 removals (NAPA, 2005).

Furthermore, the USPTO does annual quality reviews on about 2 to 3 percent of patents that are granted. This is part of a patent quality measurement program that focuses on the claims that have been allowed. This is done once the USPTO has notified the applicant that a patent will be allowed, but before it is published. A reviewer determines whether an examiner made an error in at least one claim that was allowed in the patent. According to the USPTO, the error rate is relatively constant over time (in 2000 it was 6.6%, 4.2% in 2002 and 5.3% in 2004) and it varies among technological centers (from 2.5% to 9% in different technological centers in 2004). If a reviewer determines that an examiner made an error, the case is reopened. Within the period 2000 and 2004, 302 to 401 applications have been reopened.

Overall, examiners evaluations are based on quotas (number of patent applications that are processed), errors (if the examiner is relatively junior, there is systematically a more senior examiner who signs patent applications for the assistant examiner), random check of patents that have been issued.

Because of the nature of the patent prosecution process, new hires need to be trained for the job. Since January 2006, new employees have to participate in a eight month training at the Patent Training Academy (PTA). During the first two months of examinations, new examiners do not have production quotas. Then, they start to have to meet increasing quotas. At the end of the training each examiner is required to take a proficient test. In fact, in the first year, examiners are evaluated semi-annually and can potentially be promoted twice. New examiners have a two-year probation period during which they must demonstrate their ability to work as an
examiner for the USPTO. Before January 2006, new examiners had only two weeks of training (during which they were introduced to software, and they did acquire a better knowledge about patents). After two weeks, SPEs were responsible to train the new examiners.

The necessary skills that are recognized as being fundamental to be an examiner are decision-maker skills, ability to compartmentalize, writing and reading. Examiners need to be multitask, as they simultaneously have to be able to determine the patentability of an innovation for new cases, and also have to make decisions on older cases. They need to be knowledgeable of the patent law: if correctly granted patents should not be invalidated in court in case of prosecution.

3 The Model

We consider a repeated game with two players, the PTO and a patent examiner. The PTO offers a salary scheme (a fixed salary and bonuses) to the examiner based on his performance. The examiner chooses unobservable efforts to process patent applications and assess the patentability of the innovation.

In order to determine a salary scheme, the PTO values both quantity and quality of granted patents. In terms of quantity, we assume that the PTO has a targeted quantity of patents that must be processed (based for instance on the backlog, or on the number of applications needed to fund the institution). We denote $y_1$ this first task and, to keep things simple, we assume that $y_1$ only equals zero or one. Therefore, if the examiner reaches the targeted quantity, $y_1 = 1$. It is also the PTO responsibility to insure that only deserving innovations be granted patents. Let $y_2$ be this second task, where $y_2$ also equals zero or one, with $y_2 = 1$ corresponds to the case where no mistakes have been made. As it will become clearer in what follows, the PTO does not necessarily put the same weight on quantity and quality.

Once the examiner receives an application, to complete these two tasks (task 1 related to quantity and task 2 related to quality) he exerts costly efforts: an effort $e_1 \in [0, 1]$ to perform satisfactory on the first task (to process as many patents as possible), and an effort $e_2 \in [0, 1]$ to perform on the second task (to grant patents to only deserving innovations). The second

---

5 It is costly for society to have non-deserving innovation being patented as it reduces competition, increases deadweight loss and also may reduce incentive to innovate.
task implies that the examiner exerts some effort to search for information that will prove the innovation is not novel (i.e., the examiner will find invalidating information).

Notice that even if task 1 is perfectly observable and contractible, task 2 is not. The measure of quality is hard to assess. We assume that a senior examiner checks some of the patent applications that have been processed by a patent examiner, but this is a subjective assessment.

The timing in each period is the following. First the PTO offers the examiner a salary scheme. Second, the examiner accepts or rejects it. If he rejects it his outside option value is \( w \) (reservation salary). If he accepts, he receives a patent application, and makes costly efforts \((e_1, e_2) \in [0, 1]^2\). Third, both players observe the realization of the examiner contribution, and the realization of the objective performance measure. Fourth, if the examiner has successfully accomplished task 1, the PTO pays the bonus specified in the contract; if task 2 has been accomplished successfully depending on whether it is an implicit or an explicit contract, the PTO chooses whether to pay the bonus or not.

In the baseline model we consider that the PTO offers a bonus for each task.\(^6\) The PTO’s payoff is affected by both tasks. The outcome \( y_1 \) on task 1 generates a value \( \lambda \) whereas the outcome \( y_2 \) on task 2 generates a value \((1 - \lambda)\), where \( \lambda \) is the weight that the PTO puts on task 1 (related to quantity) whereas \((1 - \lambda)\) is the weight it puts on task 2 (quality). Therefore, the PTO’s expected payoff when the examiner contributes to both tasks for a compensation \( w \) is

\[
p_1 \lambda + p_2 (1 - \lambda) - w,
\]

where \( p_1 \) (respectively, \( p_2 \)) is the probability that the examiner performs satisfactorily on the first task (respectively, the second task), \( \lambda \) is the value of the outcome related to task 1 (when the outcome in quantitative terms has been reached), and \((1 - \lambda)\) the value of the outcome related to task 2.

The rationale for this objective function is as follows. Even though the PTO values both quantity and quality of granted patents, it puts some potentially different weight on them. Typically, the PTO could set targets in term of the number of applications that have to be

\(^6\)The case of a bonus based on both tasks will be presented as an extension of the model.
processed by an examiner and in terms of the average quality of the examiner’s work (e.g., how long it would take to an experienced examiner to “invalidate” an accepted application).

We assume that the essential inputs in the patent examination process are the examiner’s efforts $e_1$ and $e_2$. When the examiner reaches the targeted quantity, the output on task 1 is $y_1 = 1$ and when he reaches some measure of quality then $y_2 = 1$. Therefore, we define the probability that he performs satisfactorily on task 1 as

$$p_1 = \Pr (y_1 = 1 \mid e_1) = \frac{1}{2} (1 + e_1), \quad (1)$$

and on task 2

$$p_2 = \Pr (y_2 = 1 \mid e_2) = \frac{1}{2} (1 + \varphi e_2), \quad (2)$$

where $\varphi \in [0, 1]$ is a measure of examiners heterogeneity, with $\varphi = 1$ stands for a skilled examiner, whereas $\varphi < 1$ represents an averaged skilled examiner. Therefore, the more skilled an examiner, the higher the probability of finding that the innovation is patentable (or of finding that there exists invalidating prior art).

The cost incurred by the examiner when he exerts $(e_1, e_2)$ is

$$C (e_1, e_2) = \frac{e_1^2}{2} + \frac{e_2^2}{2} + \theta e_1 e_2, \quad (3)$$

where $\theta \in [0, 1]$ denotes the conflict between these two tasks as perceived by the examiner. For the examiner, increasing the effort to comply with patent quality standards raises the marginal cost of working on patent processing rate. If $\theta = 0$, both tasks are independent, whereas if $\theta > 0$ both tasks are substitutes. Therefore, an increase in $\theta$ has two effects: it increases the overall cost incurred by the examiner but it also increases the substitutability between the tasks.

As far as patent processing rate is concerned, we assume that the PTO rewards examiners based on a formulaic bonus approach. In other words, the PTO commits to a reward scheme that precisely takes into account the number of processed applications (whether rejected or accepted). Although, the processing rate is easily measured and contracted upon, we assume that the outcome of task 2 (quality level achieved) is observable by a supervisor but is not contractible. However, proxies (e.g., applicant’s satisfaction surveys, number of legally disputed patents, or number of citations included in the patent) for the performance on task 2 are available for
contracting purposes. Hence, when a proxy is included in the contract, the examiner essentially works towards achieving a higher proxy score which may involve a distortion with respect to the true performance (for instance, the examiner may “over please” applicants and this may be different than providing quality work).

We thus consider that there is a second performance measure $P_2$ of task 2 that is also affected by the effort $e_2$. It is an imperfect proxy for the examiner’s contribution to quality. For simplicity we also assume that $P_2$ can only take the value zero or one. When a proxy is used, the probability that the proxy yields a “successful” signal is

$$p_\beta = \Pr (P_2 = 1 \mid e_2) = \frac{\mu}{2} (1 + \varphi e_2),$$

(4)

where $\mu$ is a positive random variable distributed according to

$$\mu \sim \mathcal{L} (\mu = 1, \sigma^2).$$

The examiner receives a private information $\mu$ before choosing his levels of effort. The interpretation is similar to Baker, Gibbons and Murphey (1994). The examiner knows that if $\mu$ is small, high effort on task 2 will increase the probability to get $y_2 = 1$ but not $P_2 = 1$. For high levels of $\mu$ but smaller than 1, a high effort on task 2 will increase both probability to get $y_2 = 1$ and $P_2 = 1$. For $\mu > 1$, a small effort will increase the probability to get $P_2 = 1$, but not $y_2 = 1$.

For instance, consider that the proxy used is related to the number of citations included by the examiner. Thus, when he receives an application, the examiner also learns whether his task to find relevant information will be easier to perform if he has recently granted a similar patent.

We assume that the distribution and its support are such that

$$\frac{\mu}{2} (1 + \varphi e_2) < 1.$$

The examiner’s compensation $w$ can be composed of different bonuses. It can be $s$ which is only the base salary if none of the tasks have been completed. If only task 1 has been successfully completed it is $s + b_1$. If only task 2 has been completed, it depends on whether the PTO offers only an implicit contract on the subjective performance measure or an explicit contract on the imperfect proxy. The compensation is $s + b_2$ if only an implicit contract is offered on task 2, $s + \beta_2$ when only an explicit contract is offered on an imperfect proxy for task 2, $s + b_2 + \beta_2$ when both
implicit and explicit contracts have been offered. If both tasks have been completed, depending on whether both implicit and explicit contracts are offered on task 2, the compensation can be $s + b_1 + b_2$, $s + b_1 + b_2$, or $s + b_1 + b_2 + b_2$. When the examiner chooses his effort levels, his expected compensation is thus

$$w = s + p_1 b_1 + p_2 b_2 + p_2 b_2.$$ (5)

To summarize, the PTO can offer a compensation package $(s, b_1, b_2)$ where $s$ is a fixed salary, $b_1$ an explicit contract bonus associated to task 1 based on an objective measure of performance, $b_2$ an implicit contract bonus associated to task 2 based on a subjective performance measure and $b_2$ an explicit contract bonus based on an imperfect proxy of task 2 which is an objective performance measure. If only explicit contract bonuses are offered, the package will be $(s, b_1, b_2)$, whereas it will be $(s, b_1, b_2)$ if only implicit contract on task 2 is offered.

4 Optimal Incentive Schemes

4.1 First best effort levels

As a benchmark, we first consider the case in which efforts are observable and everything is contractible. The expected benefit of the PTO is

$$B(e_1, e_2) = p_1 \lambda + p_2 (1 - \lambda) - w,$$ (6)

where $p_1$ and $p_2$ are defined by equations (1) and (2).

The PTO solves the following program

$$\max_{e_1, e_2} B(e_1, e_2),$$

subject to

$$w - C(e_1, e_2) \geq \bar{w}.$$ (7)

The first best effort levels are found by equating the expected marginal benefit of effort $(\partial (p_1 \lambda + p_2 (1 - \lambda))/\partial e_i)$ with its marginal cost $(\partial C(e_1, e_2)/\partial e_i)$ for each task $i$, with $i = 1, 2$, or, equivalently,

$$\frac{1}{2} \lambda = e_1 + \theta e_2.$$ (8)
and
\[
\frac{1}{2} \varphi (1 - \lambda) = e_2 + \theta e_1.
\] (9)

To express the first best effort levels we need to consider two cases: 
(i) when the PTO puts relatively more weight on quantity \((\lambda - \varphi (1 - \lambda) > 0)\) and (ii) when it puts more weight on quality \((\lambda - \varphi (1 - \lambda) \leq 0)\).

In case (i), when the PTO values more quantity, the first best efforts levels are
\[
e_1^o = \frac{\lambda - (1-\lambda)\theta \varphi}{2(1-\theta^2)} \quad \text{and} \quad e_2^o = \frac{\varphi (1-\lambda)-\theta \lambda}{2(1-\theta^2)},
\]
if \(\theta < \varphi (1 - \lambda) / \lambda\) and
\[
e_1^o = \frac{1}{2} \lambda \quad \text{and} \quad e_2^o = 0,
\]
if \(\theta \geq \varphi (1 - \lambda) / \lambda\).

When the PTO puts relatively more weight on quantity, the examiner will put more effort into processing more applications, \(e_1^o(\theta) > e_2^o(\theta)\) for any \(\theta \in [0, 1]\). For low values of \(\theta\), both efforts are decreasing as \(\theta\) increases. If we start from a situation where the tasks are independent \((\theta = 0)\), an increase in \(\theta\) raises the overall cost of the examiner as well as the substitutability between the tasks. The first effect is bigger than the second one and, therefore, the examiner puts less effort in searching for invalidating information and in processing applications. As \(\theta\) raises more, the conflict between the two tasks increases, and after a certain cut-off value for \(\theta\), the examiner starts increasing the effort to perform task 1 (quantity) and keep lowering his effort to perform task 2 (quality) until it reaches a point where the effort in task 1 is constant and the examiner makes no more effort in searching for invalidating information.

In case (ii), when the PTO values more the quality, the first best effort levels are
\[
e_1^o = \frac{\lambda - (1-\lambda)\theta \varphi}{2(1-\theta^2)} \quad \text{and} \quad e_2^o = \frac{\varphi (1-\lambda)-\theta \lambda}{2(1-\theta^2)},
\]
if \(\theta < \lambda / \varphi (1 - \lambda)\), and
\[
e_1^o = 0 \quad \text{and} \quad e_2^o = \frac{1}{2} \varphi (1 - \lambda),
\]
if \(\theta \geq \lambda / \varphi (1 - \lambda)\).
Not surprisingly, when the PTO puts more weight on quality (low values of $\lambda$), the examiner searches more to invalidate the patent, $e_2^o(\theta) > e_1^o(\theta)$ for any $\theta \in [0,1]$. In this case $e_1^o(\theta)$ is non-increasing whereas $e_2^o(\theta)$ first decreases before increasing. When the susceptibility between tasks is very high, the examiner makes no more effort in performing task 1 ($e_1^o(.) = 0$) whereas the effort on task 2 becomes constant.

The optimal wage will be $w^o(\theta) = C(e_1^o, e_2^o) + \bar{w}$.

The effort on task 1 (respectively, on task 2) is non-decreasing (respectively, non-increasing) with $\lambda$. For very low values of $\lambda$, the examiner makes no effort to search for quantity, whereas for high values of $\lambda$ he makes no effort on task 2.

From these two different cases we state the first result.

**Lemma 1** Too much conflict between tasks is potentially conducive to no effort at one task.

When the PTO puts more emphasize on quantity (case (i)), the (first best) benefit function of the PTO is

\[
B(e_1^o, e_2^o) = \begin{cases} 
\frac{1}{2} + \frac{\lambda^2 + (1 - \lambda)^2 \varphi^2}{8(1 - \theta^2)} - \theta \frac{\lambda(1 - \lambda) \varphi}{4(1 - \theta^2)} - \bar{w} & \text{for } \theta < \frac{\varphi(1 - \lambda)}{\lambda} \\
\frac{1}{2} + \frac{\lambda^2 \varphi^2}{8} - \bar{w} & \text{for } \theta \geq \frac{\varphi(1 - \lambda)}{\lambda}
\end{cases}
\]

Similarly, when the PTO favors quality (case (ii)), the (first best) benefit function of the PTO is

\[
B(e_1^o, e_2^o) = \begin{cases} 
\frac{1}{2} + \frac{\lambda^2 + (1 - \lambda)^2 \varphi^2}{8(1 - \theta^2)} - \theta \frac{\lambda(1 - \lambda) \varphi}{4(1 - \theta^2)} - \bar{w} & \text{for } \theta < \frac{\lambda}{\varphi(1 - \lambda)} \\
\frac{1}{2} + \frac{(1 - \lambda)^2 \varphi^2}{8} - \bar{w} & \text{for } \theta \geq \frac{\lambda}{\varphi(1 - \lambda)}
\end{cases}
\]

In both cases, the benefit function of the PTO is first strictly decreasing with $\theta$: as the conflict between the two tasks increases, the examiner’s efforts decrease and the PTO obtains a lower benefit. After a certain threshold, the examiner does not make any effort to perform one task and the other effort becomes independent of $\theta$ and thus the benefit function is independent of the $\theta$ as well.

**4.2 Explicit contracts**

We now consider that only explicit contracts on both tasks can be proposed by the PTO. Task 1, related to quantity, is contractible and thus can be part of an explicit contract, whereas only
a proxy on task 2 can be used. Therefore the PTO offers a compensation package \((s, b_1, \beta_2)\). In this context, we analyze different scenario: first, we assume that there is no limited liability for the examiner (he can get negative bonuses). Second, there exists a constraint of limited liability.

### 4.2.1 No limited Liability

In absence of any limited liability, the PTO must insure that the examiner will accept the contract and that the incentive compatibility constraint will be satisfied. The program of the PTO becomes

\[
\begin{align*}
\text{Max}_{e_1, e_2, s, b_1, \beta_2} & \quad B(e_1, e_2), \\
\text{subject to} & \quad w - C(e_1, e_2) \geq \overline{w},
\end{align*}
\]  

(10)

and

\[
(e_1, e_2) \in \text{arg max}\{ w - C(e_1, e_2) \},
\]  

(11)

where \(w = s + p_1 b_1 + p_\beta \beta_2\) where \(p_1\) and \(p_\beta\) are defined by equations (1) and (4).\(^7\) For given \(b_1\) and \(\beta_2\), the optimal effort levels are

\[
e_1^* = \frac{b_1 - \theta \mu \varphi \beta_2}{2(1 - \theta^2)},
\]  

(13)

and

\[
e_2^* = \frac{\mu \varphi \beta_2 - \theta b_1}{2(1 - \theta^2)}.
\]  

(14)

Not surprisingly, the higher the compensation \(b_1\), the higher the effort on task 1 and the lower the effort on task 2. On the other hand, the higher the compensation \(\beta_2\), the lower the effort on task 1 and the higher the effort on task 2.

The determination of total compensation \(w\) is done by using (13) and (14) and replacing these effort levels in (11), we obtain

\[
s + E_\mu \left\{ \frac{1}{2} (1 + e_1^*) b_1 + \frac{\mu}{2} (1 + \varphi e_2^*) \beta_2 - \frac{1}{2} (e_1^*)^2 - \frac{1}{2} (e_2^*)^2 - \theta e_1^* e_2^* \right\} \geq \overline{w},
\]  

(15)

\(^7\)The problem with the shape of this bonus is that it is not optimal. It would be optimal to give a bonus based on both tasks, instead of a bonus for each task (see extension for the case of a bonus on both tasks).
which is equivalent to

\[ s \geq s^* = \bar{w} - E_{\mu} \left\{ \frac{b_1 + \mu \beta_2}{2} + \frac{b_1^2 + \nu^2 \beta_2^2 - 2\mu \beta_1 \nu \beta_2}{8(1-\theta)^2} \right\}. \]

Using the fact that \( E_{\mu} (\mu^2) = \text{var} \mu + (E_{\mu})^2 \), we obtain

\[ s^* = \frac{\bar{w}}{1} - \left( \frac{b_1 + \beta_2}{2} + \frac{b_1^2 + \nu^2 \beta_2^2 + \nu^2 \beta_2^2 \sigma^2 - 2\theta \varphi b_1 \beta_2}{8(1-\theta)^2} \right). \]

The total expected compensation paid by the PTO is thus computed as

\[ w = \bar{w} - \left( \frac{b_1 + \beta_2}{2} + \frac{b_1^2 + \nu^2 \beta_2^2 + \nu^2 \beta_2^2 \sigma^2 - 2\theta \varphi b_1 \beta_2}{8(1-\theta)^2} \right) + E_{\mu} \left\{ \frac{1}{2} (1 + \frac{b_1 - \mu \beta_2}{2}) b_1 + \frac{1}{2} (1 + \varphi \frac{\mu \beta_2 - \theta b_1}{2(1-\theta^2)}) \mu \beta_2 \right\}, \]

or, after simplifications

\[ w^* = \bar{w} + \frac{b_1^2 + \beta_2^2 (1 + \sigma^2) - 2\theta \varphi b_1 \beta_2}{8(1-\theta^2)}. \]

Replacing \( w^* \) in the objective function, we obtain the following unconstrained program

\[ \text{Max } B (b_1, \beta_2) = \left\{ \frac{1}{2} (1 + e_1^* \lambda) + \frac{1}{2} (1 + \varphi e_2^* \lambda) \right\} - \left\{ \frac{b_1^2 + \beta_2^2 (1 + \sigma^2) - 2\theta \varphi b_1 \beta_2}{8(1-\theta^2)} \right\} - \bar{w}. \] (P)

When the PTO puts relatively more weight on quantity \((\lambda - \varphi (1 - \lambda) > 0)\) the resolution of this maximization program yields the optimal levels of compensation

\[ b_1^* = \lambda - \frac{\varphi \theta (1 - \lambda)}{1 + \sigma^2 - \theta^2} \sigma^2, \] (16)

and

\[ \beta_2^* = \frac{(1 - \lambda)(1 - \theta^2)}{1 + \sigma^2 - \theta^2}, \] (17)

and the optimal fixed salary is

\[ s^* = \bar{w} - \left( \frac{b_1^* + \beta_2^*}{2} + \frac{b_1^2 + \beta_2^2 (1 + \sigma^2) - 2\theta \varphi b_1^* \beta_2^*}{8(1-\theta)^2} \right). \] (18)

In this case, for any value of \( \theta \) and \( \sigma^2 \), the bonus on task 1 is higher than the bonus on task 2. As the PTO puts more weight on the quantity, it will reward more the examiner on task 1. When the PTO puts more weight on quality \((\lambda - \varphi (1 - \lambda) < 0)\), the optimal levels of compensation are identical to (16), (17) for relatively low values of \( \theta \) \((\theta < \theta' \) where the threshold \( \theta' \) is calculated in the appendix\), and are

\[ b_1^* = 0, \] (19)
and

\[ \beta^*_2 = \frac{(1-\lambda-\lambda\theta^2)}{1+\sigma^2}, \]  

(20)

for \( \theta > \theta' \). When the proxy is not too noisy (\( \sigma^2 < (1-2\lambda)/\lambda \)), the PTO gives a higher bonus on the proxy on task 2. However, for a noisier measure (\( \sigma^2 > (1-2\lambda)/\lambda \)), and if the conflict between the tasks is not too strong (\( \theta < \theta'' \)), the PTO will give a higher bonus on task 1 than on the proxy. When the tasks are relatively independent, even though the PTO puts more weight on the quality, it still provides a higher bonus for quantity. If the two tasks become more dependent, the PTO reward more task 1 than task 2.

We summarize these findings in the following proposition.

**Proposition 1 (Explicit contracts)** When efforts are non-observable, the PTO offers to the examiner the compensation scheme \((s^*, b_1^*, \beta^*_2)\) defined by (16), (17) and (18). This compensation scheme yields the following equilibrium PTO’s benefit

\[
B(b_1^*, \beta^*_2) = \frac{1}{2} + \frac{\lambda^2-(1-\lambda)^2(2-2\varphi^2)(1-\lambda)\lambda}{8(1-\varphi)} + \frac{(1-\lambda)^2(1+\theta^2-2\varphi^2)}{8(1-\varphi)(1+\sigma^2-\theta^2)} \sigma^2 - \frac{(1-\lambda)^2\theta^2(1-\varphi^2)}{8(1-\varphi)(1+\sigma^2-\theta^2)} (\sigma^2)^2 - w. 
\]

(21)

When the performance is hard to objectively measure (i.e., through contractible measure), the PTO’s mission is harder to fulfill as

\[
\frac{\partial B(b_1^*, \beta^*_2)}{\partial \sigma^2} < 0.
\]

Thus, for any \( \sigma^2 > 0 \), the PTO’s benefit is suboptimal compared to the first best,

\[
B(e_1^0, e_2^0) > B(b_1^*, \beta^*_2).
\]

Comparative statics with respect to \( \sigma^2 \) gives expected results as both bonuses decrease with \( \sigma^2 \). When the performance measure on task 2 becomes noisier, the PTO is less eager to rely on this performance measure and therefore the PTO decreases \( \beta^*_2 \). However, to avoid too much emphasis on task 1, the PTO decreases the bonus on task 1 as well. Overall, the power of incentives tends to decrease as a noisier performance measure increases the fixed part of the agent’s optimal compensation (\( \partial s^*/\partial \sigma^2 > 0 \)). As the PTO puts more weight on quantity, (i.e., \( \lambda \) increases), the bonus on task 1 increases (\( \partial b_1^*/\partial \lambda > 0 \)) whereas the bonus on task 2 decreases.
(\partial \beta_2^* / \partial \lambda < 0). As there is more conflict between the two tasks for the examiner, both bonuses decrease (\partial b_1^* / \partial \theta < 0 and \partial \beta_2^* / \partial \theta < 0).

By using the previous findings, we can rewrite the expected equilibrium effort levels as 
\[ \bar{e}_i^* = E_{\mu} \{ e_i^* \} \] for \( i = 1, 2 \). Simple calculations allow us to derive the expected equilibrium efforts and to compare them with the first best effort levels. When the PTO puts relatively more weight on quantity (case (i), i.e., for \( \lambda - \varphi (1 - \lambda) > 0 \), the expected levels of effort are

\[ \bar{e}_1^* = \begin{cases} 
  e_1^0 & \text{for } \theta < \tilde{\theta} \\
  \bar{e}_1^*(\theta) & \text{for } \theta \geq \tilde{\theta}
\end{cases} \] (22)

and

\[ \bar{e}_2^* = \begin{cases} 
  \frac{\varphi(1-\lambda)-\theta \lambda}{2(1-\theta^2)} - \frac{(1-\lambda)\varphi}{2(1-\theta^2+\sigma^2)} \sigma^2 & \text{for } \theta < \tilde{\theta} \\
  0 & \text{for } \theta \geq \tilde{\theta}
\end{cases} \] (23)

where \( \tilde{\theta} \) is such that \( \bar{e}_2^* = 0 \) (see appendix for the existence of \( \tilde{\theta} \)). For low values of \( \theta \) (i.e., \( \theta < \tilde{\theta} \)), there is no distortion in the effort level on task 1, \( \bar{e}_1^* = e_1^0 \) whereas the examiner will put less effort into task 2, \( \bar{e}_2^* < e_2^0 \). For higher values of \( \theta \) (i.e., \( \theta \geq \tilde{\theta} \), the asymmetric information creates a distortion in the provision of both efforts, \( \bar{e}_1^* < e_1^0 \) and \( \bar{e}_2^* \leq e_2^0 \). Therefore, for any value of \( \theta > e_2^0 \), \( \bar{e}_1^* \leq e_1^0 \) and \( \bar{e}_2^* \leq e_2^0 \). The examiner exerts more effort on task 1 than on task 2, \( \bar{e}_1^* > \bar{e}_2^* \).

Both levels of effort are non increasing in \( \theta \). As \( \theta \) increases from a very small value, both efforts decrease. Recall that an increase in \( \theta \) increase the conflict between the tasks but also raises the overall cost of examination. After \( \theta \) reaches the cut-off value \( \tilde{\theta} \), the examiner no longer exert effort on task 2 and exert a constant effort on task 1.

When the PTO puts relatively more weight on quality (case (ii), i.e., for \( \lambda - \varphi (1 - \lambda) \leq 0 \), the expected levels of effort are

\[ \bar{e}_1^* = e_1^0, \] (24)

and

\[ \bar{e}_2^* = \begin{cases} 
  \frac{\varphi(1-\lambda)-\theta \lambda}{2(1-\theta^2)} - \frac{(1-\lambda)\varphi}{2(1-\theta^2+\sigma^2)} \sigma^2 & \text{for } \theta < \frac{\lambda}{(1-\lambda)\varphi} \\
  \bar{e}_2^*(\frac{\lambda}{(1-\lambda)\varphi}) & \text{for } \theta \geq \frac{\lambda}{(1-\lambda)\varphi}
\end{cases} \] (25)

In this case, \( \bar{e}_2^* < e_2^0 \). However, it might be the case that the examiner exerts more effort on task 1 than on task 2. Indeed, if \( \varphi \in \left[ \frac{\lambda}{(1-\lambda)}, \frac{\lambda}{(1-\lambda)(1+\sigma^2)} \right] \) there exists \( \tilde{\theta} \) such that \( \bar{e}_1^* = \bar{e}_2^* \).
Then, for values of \( \theta < \hat{\theta} \), \( \bar{e}_2^* < e_1^0 = \bar{e}_1^* \) and for \( \theta \geq \hat{\theta} \), \( \bar{e}_2^* \geq e_1^0 = \bar{e}_1^* \). Therefore, depending on the value of \( \theta \), even if the PTO puts more weight on quality, the examiner can exert less effort on quality.

The use of a proxy on task 2 introduces a distortion in the provision of both efforts. The examiner has a tendency to undersupply efforts on both quality and quantity. This distortion increases when the proxy becomes noisier (i.e., when \( \sigma^2 \) increases).

When there is no conflict between the two tasks (\( \theta \to 0 \)), the expected efforts are

\[
\bar{e}_1^* = e_1^0 = \frac{\lambda}{2},
\]
\[
\bar{e}_2^* = \frac{(1-\lambda)\varphi}{2(1+\sigma^2)} < e_2^0 = \frac{\varphi(1-\lambda)}{2}.
\]

Therefore, the examiner has a tendency to undersupply effort on task 2 which is not contractible.

Comparative statics for effort levels show that the higher the weight the PTO puts on the quantity, the lower the effort on quality (\( \partial e_2^*/\partial \lambda < 0 \)) and the higher the effort on quantity (\( \partial e_1^*/\partial \lambda > 0 \)).

### 4.2.2 Limited liability

When limited liability is imposed on the examiner’s incentive scheme, an extra constraint must be added to the previous maximization program (10) subject to (11), (12), which is

\[
s \geq 0, \ b_1 \geq 0, \text{ and } \beta_2 \geq 0.
\] (26)

The resolution of this program is identical to the previous one except that now

\[
s^* = \text{Max}\{\bar{w} - \left(\frac{b_1 + \beta_2^2}{2} + \frac{b_1^2 + \beta_2^2 (1+\sigma^2) - 2\theta \varphi b_1^2 \beta_2^2}{2(1-\sigma^2)}\right), 0\}.
\]

If the limited liability constraint is not binding, the solution is identical to the one derived in the no liability case. On the other hand, if it is binding then

\[
s^* = 0
\]

and

\[
w = \frac{1}{2}(1 + \frac{b_1 - \theta \varphi \beta_2}{2(1-\sigma^2)})b_1 + \frac{1}{2}(1 + \frac{\varphi (1+\sigma^2) - \theta b_1}{2(1-\sigma^2)})\beta_2.
\]
After simplifications, we obtain the following program

\[
\max_{b_1, b_2} \frac{1}{2}(1 + \frac{b_1 - \theta \varphi \beta_2}{2(1 - \theta^2)}) \lambda + \frac{1}{2}(1 + \varphi \beta_2 - \theta b_2) (1 - \lambda) - \frac{1}{2}(1 + \frac{b_1 - \theta \varphi \beta_2}{2(1 - \theta^2)}) b_1 - \frac{1}{2}(1 + \varphi \frac{\beta_2(1 + \sigma^2) - \theta b_2}{2(1 - \theta^2)}) \beta_2.
\]

The first order conditions yield the following bonus levels

\[
b^*_{1l} = \frac{1}{2} \lambda - \frac{\varphi(1-\lambda)\theta}{2(1-\theta^2+\sigma^2)} \sigma^2 - (1 - \theta^2) \frac{1+\sigma^2+\theta}{1-\theta^2+\sigma^2}
\]

and

\[
\beta^*_{2l} = \frac{1-\theta^2}{1-\theta^2+\sigma^2} \frac{(1-\lambda)}{2} - \frac{(1+\varphi\theta)}{\varphi^2}
\]

A straightforward comparison yields

\[
\beta^*_2 > \beta^*_{2l},
\]

and

\[
b^*_1 > b^*_{1l}.
\]

Thus limited liability worsens the distortion described in the previous subsection.

Note also that when there is no noise on the signal \((\sigma^2 \to 0)\), first best cannot be achieved as

\[
b^*_{1l} (\sigma^2 = 0) = \frac{1}{2} \lambda - (1 + \frac{\theta}{\varphi}) < b^*_1 (\sigma^2 = 0) = \lambda,
\]

and

\[
\beta^*_{2l} (\sigma^2 = 0) = \frac{(1-\lambda)}{2} - \frac{(1+\varphi\theta)}{\varphi^2} < \beta^*_2 (\sigma^2 = 0) = (1 - \lambda).
\]

### 4.3 Explicit and Implicit contracts

We now introduce implicit contract on task 2 and allow the PTO to offer an explicit contract on contractible outcome (quantity) and an implicit contract on the non contractible outcome linked to the quality. The work of an examiner is somewhat specialized. When the examiner receives an application, he has to go through a series of check regarding the validity of the application. One of the most important step is the search for prior art regarding the innovation. In this particular process, each examiner develops his own search methods. Arguably, this activity is hard to codify. However, the quality of the prior art background search is observable for experienced examiners as they do recognize whether their peer have done a satisfactory work.

The relationship between the PTO and its examiners is a repeated one. Thus, although task 2 cannot be explicitly contracted upon, the PTO and the examiner can agree on implicit
contracts where the PTO can promise bonus payments if the quality of the examiner’s work is satisfactory. We now consider this possibility. These contracts are self enforced in the sense that if one of the parties reneges on his promise the other party will cease any informal relationship in the future. In our setting, this simply means that if the PTO decides not to reward an examiner for a satisfactory performance on task 2, other examiners will likely learn it and lose trust in the PTO’s word to honor its promises. We assume that players use trigger strategies when one of them fail from abiding by the informal agreement.

The principal solves the following program

$$\max_{(b_1, b_2)} B(b_1, b_2) = \left\{ \frac{1}{2} (1 + e_1^*) \lambda + \frac{1}{2} (1 + \varphi e_2^*) (1 - \lambda) - \frac{b_1^2 + b_2^2 - 2 \theta \varphi b_1 b_2}{8 (1 - \theta^2)} - \pi \right\}$$

subject to

$$\frac{B(b_1, b_2) - B(b_1^*, b_2^*)}{r} \geq b_2,$$  \hspace{1cm} (27)

where \( r \) is the discount rate. At any period \( t \), the PTO has no incentive to renege as long as the benefit it gets from not reneging from that period on \( (B(b_1, b_2) r/(1 + r)) \) is higher than the benefit it will get from reneging at \( t \) and going back to explicit contracts during the next period \( (B(b_1, b_2) + b_2 + B(b_1, b_2^*) / r) \). The constraint (27) is equivalent to

$$b_2 \geq \frac{8 (1 - \theta^2) r}{\sigma^2},$$

which simply states that if the PTO reneges on its bonus promise, it can still use formal contract using the proxy for task 2 performance. The PTO will not renege if the difference between the benefit from implicitly contracting and the benefit derived from explicit contracting exceeds the short term gain from reneging on its commitment (i.e., \( b_2 \)). The solution to this program is summarize in the following proposition.

**Proposition 2** The optimal implicit bonuses offered by the PTO are as follows:

1. For \( \sigma^2 < \bar{\sigma}^2 = \frac{8 (1 - \theta^2) r}{\varphi (1 - \lambda)} \), the non-reneging constraint is binding and

$$b_{1r}^* = \lambda - (1 - \lambda) \theta \varphi + \theta \frac{8 (1 - \theta^2)}{\sigma^2} r$$ and

$$b_{2r}^* = \frac{8 (1 - \theta^2)}{\sigma^2} r$$

2. For \( \sigma^2 \geq \bar{\sigma}^2 \), the non-reneging constraint is not binding, and implicit contracting implements \( b_1^* \) and \( b_2^* \).
It is interesting to observe that since implicit contracting corrects the distortion induced by performance measurement noise on task 2, it also corrects any distortion on task 1. Plugging in the expressions of the optimal implicit bonuses, we obtain the PTO’s benefit with optimal (implicit) incentives $B(b_1^*, b_2^*)$. The equilibrium benefit of the PTO differs markedly from the benefit obtained with a performance proxy. In particular, we have the following result

**Corollary 2** The equilibrium profit with implicit contracting always (weakly) increases with $\sigma^2$.

**Proof.** We have

$$
\frac{\partial B(b_1^*, b_2^*)}{\partial \sigma^2} = 0 \text{ for any } \sigma^2 \geq \bar{\sigma}^2 \equiv \frac{8 (1 - \theta^2) r}{\varphi (1 - \lambda)},
$$

$$
\frac{\partial B(b_1^{\ast_r}, b_2^{\ast_r})}{\partial \sigma^2} > 0 \text{ for any } \sigma^2 < \bar{\sigma}^2 \equiv \frac{8 (1 - \theta^2) r}{\varphi (1 - \lambda)}.
$$

While the PTO’s benefit with explicit contracting decreases with $\sigma^2$, the benefit with implicit contracts increases when the proxy becomes more noisy. To understand this result one must recall that when $\sigma^2$ increases $B(b_1^*, \beta_2^*)$ decreases. In other words, the fallback position of the PTO becomes less attractive and constraint (27) is relaxed.

### 4.4 Comparison of implicit and explicit contracts

Let us define $\sigma^{2*}$ as

$$
\sigma^{2*} = \bar{\sigma}^2 \frac{(1 - \lambda) \varphi}{(1 - \lambda) \varphi - 4r + \sqrt{(4r)^2 + 8r(1 - \lambda) \varphi}} < \bar{\sigma}^2
$$

A comparison between implicit and explicit contracts shows that

**Proposition 3** The PTO’s benefit is

1. decreasing with $\sigma^2$ for any $\sigma^2 \leq \sigma^{2*}$, and all contracts are explicit;
2. increasing for any $\sigma^2$ such that $\sigma^{2*} < \sigma^2 \leq \bar{\sigma}^2$, and performance on task 2 is rewarded through an implicit contract;
3. constant and first best for any $\sigma^2 > \bar{\sigma}^2$ and performance on task 2 is rewarded through an implicit contract.
When the proxy used on quality is not very noisy, only explicit contracts are offered by the PTO. Indeed, the PTO’s benefit is higher when it offers a bonus on verifiable outcomes, and the proxy is not too noisy, so that the PTO can rely on it. However, when the proxy becomes noisier, the PTO is better off if it offers only implicit contracts. First, the PTO’s benefit is increasing as the non-reneging constraint is binding. Then, the constraint is no longer binding and therefore the profit is independent on the noise.

5 Self-funding Constraint

We now take into account an important aspect of the problem: the PTO is a self-funded agency. In several countries (e.g., U.S., Canada and U.K.) patent offices are self-funded, and, as such, they must get revenues from applicant fees to cover part of their operating costs. The role of patenting fees have been studied in the literature (for a recent survey see de Rassenfosse and van Potthelsberghe, 2011). Patent fees vary across time and across countries (Lerner, 2000). Patent offices in different countries might want to change their fees to balance their budget but also to adjust to other patent offices. Patent fees are divided between application fees and renewal fees where the later fees contain some uncertainty as not all patents are renewed. An optimal fee structure may deviate from the socially optimal scheme when the PTO is bound by a budget constraint (Gans et al., 2004). Indeed, a self funded PTO has an incentive to encourage too many rewards and to reduce renewal fees. These results are also derived in a different context by Baudry and Dumont (2009). Therefore, here we consider unique fees and we do not take into account the randomness of the non renewal. In a simple way we assume that the fee paid by the applicant \( F \) must cover the operating costs that are represented by the wages. Each patent application will generate a fee \( F \) and will cost a wage \( w \) to be processed.

Application fees will depend on the disposal of the patent (whether it is rejected or accepted). To simplify we assume that an unique fee is paid when the applicant applies for a patent.

We thus add self-funding constraint to the maximization program of the PTO program which is

\[
F \geq \frac{w}{N},
\]

where \( F \) represents the fee paid by the patent applicant, \( w \) is the wage paid to the examiner.
and $N$ is the number of patent applications processed by the examiner. This number $N$ should be linked to the task $1$ (quantity). For instance, the examiner is supposed to process $\overline{N}$ patent applications (in which case the task is accomplished and $y_1 = 1$). Therefore $N \leq \overline{N}$. To simplify we assume that $N = 1$. It is obviously a fairly simple representation of a much more complex constraint as we only consider one examiner. Therefore the self-funding constraint becomes

$$F \geq w.$$  

We first analyze the impact of this constraint on the first best levels of effort before studying the impact on the explicit contracts and implicit contract.

### 5.1 First best effort levels with self-funded PTO

The program of the PTO becomes

$$\max_{e_1, e_2} B(e_1, e_2),$$

subject to

$$w - C(e_1, e_2) \geq \overline{w},$$

and

$$F \geq w.$$  

For relatively high values of $F$ ($F \geq \frac{1}{8}(\lambda^2 + (1 - \lambda)^2 + \bar{w})$), the self-finding constraint is not binding and the first best effort levels are unchanged ($e_1^0, e_2^0$).

For lower values of $F$ (as long as $F < w_\theta(\theta)$) the self-finding constraint will be binding and the level of efforts will be affected. The resolution of the program (provided in appendix) gives the following levels of effort

$$e_1' = \Phi_s \frac{\lambda - (1 - \lambda)\theta}{2(1 - \theta^2)},$$

$$e_2' = \Phi_s \frac{(1 - \lambda) - \lambda\theta}{2(1 - \theta^2)},$$

where

$$\Phi_s = \left[ \frac{8(F - \bar{w})(1 - \theta^2)}{\lambda^2 + (1 - \lambda)^2(1 - 2\lambda\theta)} \right]^{\frac{1}{2}}.$$  

For $\lambda > 1/2$ and $\theta < \lambda/\varphi (1 - \lambda)$ (or $\lambda \leq 1/2$ and $\theta < \lambda/\varphi (1 - \lambda)$) we have

$$e_1' = \Phi_s e_1^0,$$

$$e_2' = \Phi_s e_2^0.$$
and $\Phi_s < 1$ (see appendix). Therefore the level of efforts are suboptimal ($e'_i < e^*_i$ for $i = 1, 2$) as the PTO must reduce the wage that it will offer to the examiner.

**Lemma 3** For a given patenting fee $F < w^\alpha(\theta)$, the levels of efforts are suboptimal.

### 5.2 Explicit contracts with self-funded PTO

This will add a self-funding constraint to the previous program as defined by (10), (11) and (12), which is

$$g(p_1)F \geq w,$$

(29)

where $F$ represents the fees paid by the patent applicant, $w$ is the wage paid to the examiner and $g(p_1)$ is the number of disposed patents with $g'(.) > 0$. To simplify we assume that $g(p_1) = p_1$. In this relatively simple setting, the self-funding constraint is such that the revenue raised from applicant fees must be superior or equal to the wage paid to the examiner. It is obviously a fairly simple representation of a much more complex constraint as we only consider one examiner.

If the fees paid by applicants are very high, the self funded constraint will not be binding and the optimal levels of bonuses are those defined above (22) and (23). However, if these fees are not too high, the constraint will be binding and it will affect the bonuses as well as the provision of efforts. Intuitively, the PTO will increase $b^*_1$ while it decreases $\beta^*_2$. Indeed, the PTO will increase the incentives to process more applications and reduces the incentive to provide more quality.

### 5.3 Explicit and implicit contracts with self-funded PTO

[To be completed]

### 5.4 Endogenous fees

We have assumed that the fees paid by the applicants are fixed and, therefore, by adding a self-funding constraint, the PTO will change the incentives provided to the examiner. Nevertheless, it can be the case that the PTO decides to set the bonuses at a certain level and to adjust the fees in order to satisfy the self-funding constraint. In this case it becomes important to
consider the supply of patent applications. Indeed, as patenting fees increase, it is likely that
less innovators will apply for patents. This is especially true for small applicants.

However, by reducing the number of applications, the PTO makes it harder for the examiner
to fulfill their quotas requirement. As less applicants file for patents, less applications will be
examined and it will become impossible to reach the objective on task 1.
References


Appendix

First Best effort levels

From Equation (8) and (9) we can easily derive the first best effort levels

\[ e_1^o = \frac{\lambda - (1 - \lambda)\theta \varphi}{2(1 - \theta^2)} \quad \text{and} \quad e_2^o = \frac{\varphi(1 - \lambda) - \theta \lambda}{2(1 - \theta^2)}. \]

These effort levels will be positive or negative depending on the sign of \( \lambda - (1 - \lambda)\theta \varphi \). The Hessian matrix is semi-definite negative which insures that we have a maximum.

We can then plug these effort levels within the benefit function of the PTO and we therefore derive the first best benefit of the PTO

\[ B(e_1^o, e_2^o) = \max \left\{ \frac{1}{2} + \frac{\lambda^2 + (1 - \lambda)^2 \varphi^2}{8(1 - \theta^2)} - \theta \frac{\lambda(1 - \lambda)\varphi}{4(1 - \theta^2)}, \max \left\{ \frac{1}{2} + \frac{\lambda^2}{8}, \frac{1}{2} + \frac{(1 - \lambda)^2 \varphi^2}{8} \right\} - \overline{w} \right\}. \]

First Best effort levels with self-funding constraint

The program can be rewritten as a Lagrangian

\[ L = B(e_1, e_2) + \gamma_1(w - C(e_1, e_2) - \overline{w}) + \gamma_2(F - w) \]

where \( \gamma_1 \) and \( \gamma_2 \) are the Lagrange multipliers.

*** SOC

Explicit contracts

We calculate the expected equilibrium effort levels \( \overline{e}_i^* = E_\mu \{ e_i^* \} \) for \( i = 1, 2 \) where

\[ e_1^* = \frac{b_1^*-\theta \mu \varphi \beta_2^*}{2(1 - \theta^2)}, \]

and

\[ e_2^* = \frac{\mu \varphi \beta_2^*-\theta \beta_2^*}{2(1 - \theta^2)}, \]

and

\[ b_1^* = \lambda - \frac{\varphi(1 - \lambda)}{1 + \sigma^2 - \theta^2} \sigma^2, \]

and

\[ \beta_2^* = \frac{(1 - \lambda)(1 - \theta^2)}{1 + \sigma^2 - \theta^2}. \]
Therefore

\[
\tau_1^* = E_{\mu} \left[ \frac{\beta^*_1 - \theta \mu \varphi \beta^*_2}{2 (1 - \theta^2)} \right] = \frac{1}{2 (1 - \theta^2)} E_{\mu} \left[ \frac{\varphi (1 - \lambda)}{1 + \sigma^2 - \theta^2} \right] \lambda - \theta \varphi (1 - \lambda) \sigma^2 + \mu (1 - \theta^2)
\]

\[
= \frac{1}{2 (1 - \theta^2)} E_{\mu} \left[ \frac{\lambda - \theta \varphi (1 - \lambda) \sigma^2 + \mu (1 - \theta^2)}{2 (1 - \theta^2)} \right] = e_1^0(\theta),
\]

and

\[
\tau_2^* = E_{\mu} \left[ \frac{\mu \varphi \beta^*_2 - \theta \beta^*_1}{2 (1 - \theta^2)} \right] = \frac{1}{2 (1 - \theta^2)} E_{\mu} \left[ \frac{\mu \varphi (1 - \lambda) (1 - \theta^2)}{1 + \sigma^2 - \theta^2} \right] - \theta (\lambda - \frac{\varphi (1 - \lambda) \sigma^2}{1 + \sigma^2 - \theta^2})
\]

\[
= \frac{1}{2 (1 - \theta^2)} E_{\mu} \left[ (1 - \lambda) \frac{\varphi (1 - \theta^2) \mu + \theta^2 \sigma^2}{1 + \sigma^2 - \theta^2} - \theta \lambda \right]
\]

\[
= \frac{e_2^0(\theta) - \theta \lambda}{2 (1 - \theta^2)} < e_2^0(\theta),
\]

where \( e_2^0 = \frac{\varphi (1 - \lambda) - \theta \lambda}{2 (1 - \theta^2)} \).

Existence of \( \bar{\theta} \)

There exists \( \bar{\theta} \) such that \( \tau_2^*(\bar{\theta}) = 0 \). It is equivalent to solving a cubic equation.

By construction \( \bar{\theta} < \frac{\varphi (1 - \lambda)}{\lambda} \), then \( \tau_1^*(\bar{\theta}) < \frac{\lambda}{2} \).