Optimal Contract for

Experimentation and Production

Fahad Khalil, Jacques Lawarrée and Alexander Rodivilov

Department of Economics
University of Washington

November 16, 2016

(updated 2-27-17)

<u>Preliminary: not for circulation</u>

Abstract: Before embarking on a project, a principal must often rely on an agent to learn about its profitability. These situations are conveniently modeled as two-armed bandit problems highlighting a trade-off between learning (experimentation) and production (exploitation). We derive the optimal contract for both experimentation and production when the agent has private information about his skill in experimentation. Private information in the experimentation stage can generate asymmetric information between the principal and agent about the expected profitability of the production stage. The degree of asymmetric information is endogenously determined by the length of the experimentation stage. An optimal contract uses the timing of payments, the length of experimentation, and the output to screen the agent. To induce revelation during the experimentation, the principal utilizes the stochastic structure of asymmetric learning by agents with different skills. Both upward and downward incentive constraints can be binding. The relative probabilities of success and failure between agents of different skills imply that rent to a highly-skilled agent should be paid after early success and rent to a less-skilled agent after late success. The optimal contract may also feature payments for failure, excessive experimentation, and over- or under-production in the production stage.

We are thankful for the helpful comments of Nageeb Ali, Renato Gomes, Navin Kartik, David Martimort, Larry Samuelson, and Dilip Mookherjee.

1. Introduction

Before embarking on a project, a principal would typically like to learn about its profitability to determine how much resources to allocate to the project. When an oil company explores new areas for oil fields, it performs seismic surveys and exploration drills to figure out the amount of oil it can expect from the areas.[1] While the oil company experiments with different potential sites, it also diverts resources and delays the production of oil. This creates a trade-off between experimentation and production (exploitation) analogous to a two-armed bandit problem.[2]

An additional complexity arises if the experiments are performed by an agent who privately knows his skill in experimentation. The experimentation process itself can then create asymmetric information about the profitability of the project. Exploration drills will demonstrate the profitability of the oil field. If the agent is not very skilled at experimenting, a poor result from the exploration well only provides weak evidence of a poor project. However, if the principal is misled into believing that the agent is highly skilled, she becomes more pessimistic than the agent. A new trade-off appears for the principal. More experiments may provide more information about the profitability of the well but can also increase asymmetric information about the expected profitability. Because of this asymmetry of information, when production ultimately starts, the principal may not allocate the right amount of resources to the exploitation of the field.

In this paper, we derive the optimal contract for both experimentation and production. At the outset, the principal and agent are symmetrically informed that production cost can be high or low. Before production takes place, the principal asks the agent to gather additional information about the actual production cost. This is the experimentation stage. For most of the paper, we assume that the information gathering takes the form of looking for good news, i.e., whether cost of production is low.[3] When experimentation succeeds, it is publicly revealed that the cost is low, and production occurs under symmetric information. We say that experimentation fails when the agent does not learn the actual cost by the end of the experimentation stage. Then, production occurs under expected cost, which can be different for the principal and the agent. This will lead to a rent for the agent as we explain next.

The agent is privately informed about his experimentation skill, which can be either high or low. When cost is low, a high-skill agent has a higher likelihood of finding low cost compared to a low-skill agent. To see how experimentation can endogenously create asymmetric information between the principal and the agent, consider the case when a low-skill agent claims to have high skill, and experimentation fails. The principal is now more pessimistic that the cost is high than the low-skill agent: the agent knows that he is not very skilled at experimenting, so

[1] Other applications are the testing of new drugs, the adoption of new technologies or products, the identification of new investment opportunities, the evaluation of the state of the economy, consumer search, etc.

[2] See Bolton and Harris (1999) or Keller et al. (2005).

[3] We also show that our key results extend to the case of bad news.

his failure does not indicate strongly that the project is costly. Thus, if experimentation fails, the lying low-skill agent will have a lower expected cost of production compared to the principal, who will overcompensate the lying low-skill agent in the production stage (mistakenly believing the agent has high skill). To deter this type of lying, the principal must pay a rent to the low-skill agent.

A key contribution of our model is to study how incentives for production affects incentives to experiment and, conversely, how the asymmetric information generated in the experimentation stage impacts production. At the end of the experimentation stage, there is a non-trivial decision regarding the scale of output. This decision depends on what is learned during experimentation. Relative to the nascent literature on incentives for experimentation, the novelty of our approach is to study optimal contracts for *both* experimentation and production. As we will see, much can be learned even when experimentation fails, and this information would be lost in a model without a production stage. If the principal asks the agent to experiment longer, there is a greater chance to succeed and fine-tune the size of the project. However, experimentation is costly since it endogenously creates asymmetric information and production has to be postponed.

The asymmetric information created by experimentation impacts the optimal contract: in particular (i) the length of the experimentation period, (ii) the information rent for the agent and the timing of its payment, and (iii) the output. We discuss them briefly in turn.

(i) First, we show that the optimal length of the experimentation period could be shorter or longer than the first best length while most models of experimentation find under experimentation relative to the first best. Using time for screening the types turns out to be complex as it balances two countervailing forces: a better experimenter learns good news more quickly but also becomes pessimistic more quickly after successive failures. Under asymmetric information, the difference in expected costs, the driving force for the rent, determines the distortion in the length of the experimentation stage. When the agent lies about his type, his expected cost after failure is different from the principal's expected cost. As a result, the agent's informational rent is positively related to the difference in the expected cost of the two different types. We show that this difference is non-monotonic in time so that the principal may sometimes benefit from increasing the length of the experimentation stage.

(ii) Second, we show that it is possible that the high-skill agent also gets a rent and that the timing of payment can be used to limit the rents paid to both skill types. The rent given to the low-skill agent can be attractive to the high-skill agent, but he would face a gamble if he pretends to be a low-skill agent. This is because he will be under-compensated at the production stage if experimentation fails as he is relatively more pessimistic compared to the principal. If the net-benefit of this gamble is positive, the high-skill agent will also receive a rent. Suppose the principal rewards early success for the low-skill agent. Such a scheme would look attractive for the high-skill agent who is more likely to succeed early in the experimentation stage. But an alternative scheme, such as rewarding late failure by the low-skill agent, may also become attractive for the high-skill agent during a long experimentation stage because successive failures convince the high-skill agent that the project is high cost and experimentation is likely to fail. We show that the dynamic nature of learning by two skill types who learn at different speeds can force the principal to give a rent to both skill types.[4]

As indicated by the above arguments regarding the relative likelihoods of success and failure, we find that the timing of the payments plays a crucial role. When the principal must pay a rent to the high-skill agent, it will be as a reward for early success since the low-skill agent is less likely to succeed early. If the principal wants to reward a low-skill agent for success, it has to be late in the experimentation stage.

Remarkably, we also show that it may be optimal for the principal to reward the low-skill agent for *failure*.[5] When the optimal length of the experimentation period is short, the relative probability of success for a high-skill agent is greater than his relative probability of failure. Thus, rewarding the low-skill agent after failure becomes a useful tool to screen the skill types. One may wonder if this result depends on the assumption that the agent cannot hide success, and we show in an extension that it does not.[6]

(iii) Third, unless experimentation succeeds, the principal will use the choice of output to screen the two skill types during the production stage. Since the low-skill agent always gets a rent, we expect and indeed find that the output of the high-skill agent is distorted downward as in a standard static second best contract. However, when the high-skill agent also commands a rent, the output of the low-skill agent is distorted *upward*. A higher output for the low-skill agent makes it more costly for the high-skill agent to lie since a lying high-skill agent, being relatively more pessimistic after failure, will be under-compensated in the production stage.

Finally, we also consider an extension where the agent looks for bad news during the experimentation stage. Now it is the high-skill agent who always gets rent as he is more pessimistic after the same amount of failures. We find results analogous to the case of good news. In particular, the timing of the payment is reminiscent of that from learning good news. However, distortions in output are reversed since now it is the high-skill agent who always gets rent.

---

[4] To highlight the role of the different length of the experimentation stage for different types, we consider, in an extension, a case when the length of the experimentation stage must be identical for both types. We find that the high type can no longer command a rent.

[5] See Manso (2011) for a similar result in a model with moral hazard.

[6] If the agent could hide success, he can guarantee apparent failure in the experimentation stage. In such a case, preventing the agent from hiding success introduces additional ex post moral hazard constraints that impose additional costs on the principal, but rewarding failure can still be optimal due to the same argument based on relative probabilities of success and failure.

Our paper builds on two strands of the literature. First, it is related to the literature on principal-agent contracts with endogenous information before production.[7] It is typical to consider one shot models, where an agent exerts effort that increases the precision of a signal of the state relevant for a production decision. By modeling effort as experimentation, we introduce a dynamic learning aspect, and especially the possibility of learning with asymmetric speeds. We contribute to this literature by characterizing the structure of incentive schemes in a dynamic learning stage. Importantly, in our model, the principal can determine the degree of asymmetric information by choosing the length of the experimentation stage, and there can be over or under-experimentation.

To model information gathering, we rely on the the growing literature on contracting for experimentation following Bergmann and Hege (1998, 2005). Most of that literature has a different focus and characterizes incentive schemes for addressing moral hazard during experimentation but do not consider adverse selection.[8] Recent exceptions that introduce adverse selection are Gomes, Gottlieb and Maestri (2016) and Halac, Kartik and Liu (2016).[9] In Gomes, Gottlieb and Maestri, there is two-dimensional hidden information, where the agent is privately informed about the quality (prior probability) of the project as well as a private cost of effort for experimentation. They find conditions under which the second hidden information problem can be ignored. Halac, Kartik and Liu (2016) have both moral hazard and hidden information. They extend the moral hazard-based literature by introducing hidden information about expertise in the experimentation stage to study how asymmetric learning by the high and low-skill agents affects the bonus that needs to be paid to induce the agent to work.[10] We add to the literature by explicitly modeling a production stage following the experimentation stage, such that contracts provide incentive for production (exploitation) as well as experimentation. Furthermore, asymmetric information about expertise in the experimentation stage implies that production takes place under asymmetric information if experimentation fails. This latter aspect would be missing in a model of incentives for experimentation without a production stage. Unlike the rest of the literature, we find that over-experimentation relative to the first best can be optimal because the difference in expected production cost, the source of information rent, can decrease in time after a succession of failures.

[7] Early papers are Cremer and Khalil, 1992; Lewis and Sappington, 1997, and Cremer, Khalil, and Rochet, 1998, while Krahmer and Strausz (2011) contains recent citations.

[8] See also Horner and Samuelson (2013).

[9] See also and Gerardi and Maestri (2012) for another model where the agent is privately informed about the quality (prior probability) of the project.

[10] They show that, without the moral hazard constraint, the first best can be reached. In our model, we impose a limited liability instead of a moral hazard constraint.

2. The Model (Learning good news)

A principal hires an agent to implement a project of a variable size. Both the principal and agent are risk neutral and have a common discount factor $\delta \in (0,1]$. It is common knowledge that the marginal cost can be low or high, i.e., $c \in \{\underline{c}, \overline{c}\}$, with $0 < \underline{c} < \overline{c}$. The probability that $c = \underline{c}$ is denoted by $\beta_0 \in (0,1)$. Before the actual *production stage* (exploitation), the agent can gather information regarding the production cost, which is called the *experimentation stage*.

The experimentation stage

During the experimentation stage, the agent gathers information about the cost of the project. The experimentation stage takes place over time, $t \in \{1,2,3, \dots .T\}$, where T is the maximum length of the experimentation stage and is determined by the principal. In each period t, experimentation costs $\gamma > 0$, and we assume that this cost γ is paid by the principal at the end of each period. Thus, there is no moral hazard aspect in this model. We assume that it is always optimal to experiment at least once.

In the base model, we also assume that information gathering takes the form of looking for good news (see section 5 for the case of bad news). If the cost is actually low, the agent learns whether the cost is low with probability λ in each period $t \leq T$. If the agent learns that the cost is low (*good news*) in a period t, we will say that the experimentation was successful.[11] The experimentation stage then stops. If the agent fails to learn that the cost is low in a period $t < T$, the agent continues to experiment, but both the agent and the principal become more *pessimistic* about the likelihood of the cost being low. We say that experimentation has failed if the agent fails to learn that cost is low in all T periods.

We assume that the agent is privately informed about his experimentation skill represented by λ. Therefore, the principal faces an adverse selection problem. As we will see next, this implies that the principal and agent may update their beliefs differently during the experimentation stage. The agent's private information about his skill λ determines his type, and we will refer to an agent with high or low skill as a high or low-type agent. With probability ν, the agent has high skill in experimenting, i.e., the agent is a high type, $\theta = H$. With probability $(1 - \nu)$, he is a low type, $\theta = L$. Thus, we define the learning parameter with the type superscript: $\lambda^\theta = Pr(type\ \theta\ learns\ c = \underline{c} | c = \underline{c})$, where $0 < \lambda^L < \lambda^H < 1$.[12] If experimentation fails to reveal low cost in a period, agents with different types form different beliefs about the expected cost of the project. Given $t - 1$ failures, we denote by β_t^θ the updated

---

[11] We assume that the agent cannot hide the evidence of the cost being low. We will revisit this assumption below.
[12] If $\lambda^\theta = 1$, the first failure would be a perfect signal regarding the project quality.

belief of a θ-type agent that the cost is actually low at the beginning of period t. For period $t > 1$, we have $\beta_t^\theta = \frac{\beta_{t-1}^\theta(1-\lambda^\theta)}{\beta_{t-1}^\theta(1-\lambda^\theta)+(1-\beta_{t-1}^\theta)}$, which can be re-written in terms of β_0 as:

$$\beta_t^\theta = \frac{\beta_0(1-\lambda^\theta)^{t-1}}{\beta_0(1-\lambda^\theta)^{t-1}+(1-\beta_0)}.$$

The θ-type agent's expected cost at the beginning of period t is then given by:

$$c_t^\theta = \beta_t^\theta \underline{c} + (1-\beta_t^\theta)\, \overline{c}.$$

Two aspects of learning are worth noting. First, after each period of failure during experimentation, β_t^θ falls, there is more pessimism that the true cost is low, and the expected cost c_t^θ increases and converges to $\overline{c}$. Second, for the same number of failures during experimentation, there is *relatively more pessimism* when the agent is a high type than a low type. An example of how the expected cost c_t^θ converges to $\overline{c}$ for each type is presented in Figure 1 below.

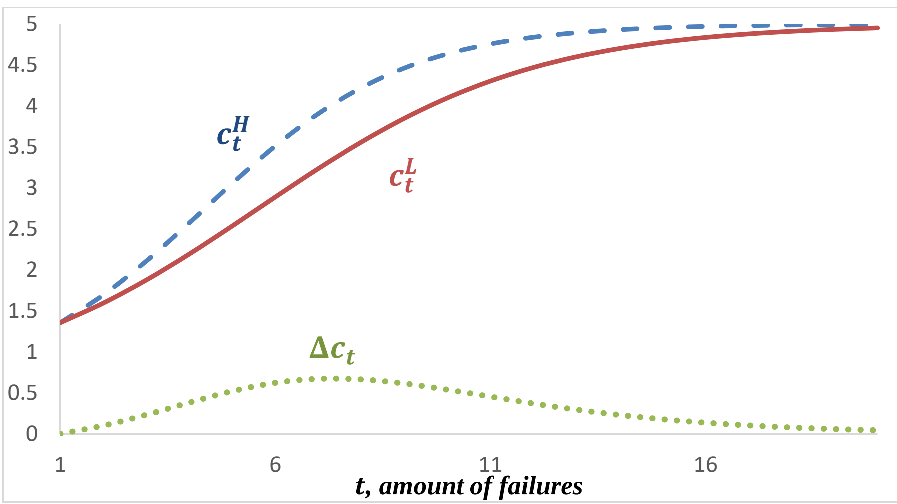

Figure 1. Expected cost with $\lambda^H = 0.35$, $\lambda^L = 0.2$, $\beta_0 = 0.7$, $\underline{c} = 0.5$, $\overline{c} = 5$.

For future use, we also note that the difference in the expected cost, $\Delta c_t = c_t^H - c_t^L > 0$, is a *non-monotonic* function of time, and both $c_t^H$and c_t^L approach $\overline{c}$ as failure persists over time. Furthermore, there exists a unique time period t_Δ such that Δc_t achieves the highest value at this time period, where

$$t_\Delta = arg \max_{1\le t\le T} \frac{(1-\lambda^L)^t - (1-\lambda^H)^t}{(1-\beta_0+\beta_0(1-\lambda^H)^t)(1-\beta_0+\beta_0(1-\lambda^L)^t)}.$$

The production (exploitation) stage

After the experimentation stage ends, production takes place in the production stage. The principal's value of the project is $V(q)$, where $q > 0$ is the size of the project. The function $V(\cdot)$ is strictly concave, twice differentiable on $(0, +\infty)$, and satisfies the Inada conditions. The size of the project and the payment to the agent are determined by the contract offered by the principal before the experimentation stage takes place. If experimentation reveals that cost is low in a period t, experimentation stops and production takes place based on $c = \underline{c}$. If experimentation fails, production occurs based on the expected cost in period T.[13]

The contract

Before the experimentation stage takes place, the principal offers the agent a menu of dynamic contracts. Relying on the revelation principle, we use a direct truthful mechanism, where the agent is asked to announce his type, denoted by $\hat{\theta}$. A contract specifies, for each type of agent, the length of the experimentation stage, the size of the project, and a transfer as a function of the publicly observable history, which, in this case, is whether or not the agent succeeded while experimenting. In terms of notation, we represent the observable history in the case of success by including $\underline{c}$ as an argument in the wage and output. In the case of failure, we include the expected cost $c^{\hat{\theta}}_{T^{\hat{\theta}}}$.

A contract is defined formally by $\varpi^{\hat{\theta}} = \left(T^{\hat{\theta}}, \left\{w_t^{\hat{\theta}}(\underline{c}), q_t^{\hat{\theta}}(\underline{c})\right\}_{t=1}^{T^{\hat{\theta}}}, \left\{w^{\hat{\theta}}\left(c^{\hat{\theta}}_{T^{\hat{\theta}}}\right), q^{\hat{\theta}}\left(c^{\hat{\theta}}_{T^{\hat{\theta}}}\right)\right\}\right)$, where $T^{\hat{\theta}}$ is the maximum duration of the experimentation stage for the announced type $\hat{\theta}$, $w_t^{\hat{\theta}}(\underline{c})$ and $q_t^{\hat{\theta}}(\underline{c})$ are the agent's wage and the output produced if he observed $c = \underline{c}$ in period $t \leq T^{\hat{\theta}}$ and $w^{\hat{\theta}}\left(c^{\hat{\theta}}_{T^{\hat{\theta}}}\right)$ and $q^{\hat{\theta}}\left(c^{\hat{\theta}}_{T^{\hat{\theta}}}\right)$ are the agent's wage and the output produced if the agent fails $T^{\hat{\theta}}$ consecutive times.

An agent of type θ, announcing his type as $\hat{\theta}$, receives expected utility $U^{\theta}\left(\varpi^{\hat{\theta}}\right)$ at time zero from a contract $\varpi^{\hat{\theta}}$:

$$U^{\theta}\left(\varpi^{\hat{\theta}}\right) = \beta_0 \sum_{t=1}^{T^{\hat{\theta}}} \delta^t \left(1-\lambda^{\theta}\right)^{t-1} \lambda^{\theta} \left(w_t^{\hat{\theta}}(\underline{c}) - \underline{c}\, q_t^{\hat{\theta}}(\underline{c})\right)$$

$$+\delta^{T^{\hat{\theta}}} \left(1-\beta_0+\beta_0\left(1-\lambda^{\theta}\right)^{T^{\hat{\theta}}}\right)\left(w^{\hat{\theta}}\left(c^{\hat{\theta}}_{T^{\hat{\theta}}}\right) - c^{\theta}_{T^{\hat{\theta}}}\, q^{\hat{\theta}}\left(c^{\hat{\theta}}_{T^{\hat{\theta}}}\right)\right).$$

[13] We assume that the agent will learn the exact cost later but it is not contractible.

Conditional on the actual cost being low, which happens with probability β_0, the probability of succeeding for the first time in period $t \le T^{\hat{\theta}}$ is given by $\left(1-\lambda^\theta\right)^{t-1}\lambda^\theta$. If the agent succeeds, he will produce $q_t^{\hat{\theta}}(\underline{c})$ and will be paid $w_t^{\hat{\theta}}(\underline{c})$ by the principal. In addition, it is possible that experimentation fails. This is the case either if the cost is actually high ($c=\bar{c}$), which happens with probability $1-\beta_0$, or, if the agent fails $T^{\hat{\theta}}$ times despite $c=\underline{c}$, which happens with probability $\beta_0(1-\lambda^\theta)^{T^{\hat{\theta}}}$. In this case, the agent produces $q^{\hat{\theta}}\left(c^{\hat{\theta}}_{T^{\hat{\theta}}}\right)$ based on expected cost and is paid $w^{\hat{\theta}}\left(c^{\hat{\theta}}_{T^{\hat{\theta}}}\right)$.

The optimal contract will have to satisfy the following incentive compatibility constraint for all θ and $\hat{\theta}$:

$$(IC) \qquad U^\theta\left(\varpi^\theta\right) \ge U^\theta\left(\varpi^{\hat{\theta}}\right).$$

The principal's expected payoff at time zero from a contract ϖ^θ offered to the agent of type θ is

$$\pi^\theta\left(\varpi^\theta\right) = \beta_0 \sum_{t=1}^{T^\theta} \delta^t \left(1-\lambda^\theta\right)^{t-1}\lambda^\theta \left(V\left(q_t^\theta(\underline{c})\right) - w_t^\theta(\underline{c}) - \frac{\sum_{s=1}^{t}\delta^s}{\delta^t}\gamma\right)$$

$$+\delta^{T^\theta}\left(1-\beta_0+\beta_0\left(1-\lambda^\theta\right)^{T^\theta}\right)\left(V\left(q^\theta\left(c^\theta_{T^\theta}\right)\right) - w^\theta\left(c^\theta_{T^\theta}\right) - \frac{\sum_{s=1}^{T^\theta}\delta^s}{\delta^{T^\theta}}\gamma\right).$$

Thus, the principal's objective function is:

$$\nu\pi^H(\varpi^H) + (1-\nu)\pi^L(\varpi^L).$$

To summarize, the timing is as follows:

(1) The agent learns his type θ.
(2) The principal offers a contract to the agent. In case the agent rejects the contract, the game is over and both parties get payoffs normalized to zero; if the agent accepts the contract, the game proceeds to the experimentation stage with duration as specified in the contract.
(3) The experimentation stage begins.
(4) If the agent learns that $c=\underline{c}$, the experimentation stage stops and the production stage starts with output and transfers as specified in the contract.
In case no success is observed during the experimentation stage, the production stage starts with output and transfers as specified in the contract.

2.1 The First Best Benchmark

Suppose the agent's type θ is common knowledge *before* the principal offers the contract. The first-best solution is found by maximizing the principal's profit such that $U^{\theta}(\varpi^{\theta}) \geq 0$. Since the expected cost is rising as long as success is not obtained, the first-best solution is characterized by a termination date T_{FB}^{θ}, where:

$$T_{FB}^{\theta} \in \underset{T^{\theta}}{arg\,max} \Big\{ \pi^{\theta}(\varpi^{\theta}) = \beta_0 \textstyle\sum_{t=1}^{T^{\theta}} \delta^t \left(1-\lambda^{\theta}\right)^{t-1} \lambda^{\theta} \left(V\left(q_t^{\theta}(\underline{c})\right) - \underline{c} q_{t^{\theta}}^{\theta}(\underline{c}) - \frac{\sum_{s=1}^{t} \delta^s}{\delta^t} \gamma \right) + \delta^{T^{\theta}} \left(1 - \beta_0 + \beta_0 \left(1-\lambda^{\theta}\right)^{T^{\theta}}\right) \left(V\left(q^{\theta}(c_{T^{\theta}}^{\theta})\right) - c_{T^{\theta}}^{\theta} q^{\theta}(c_{T^{\theta}}^{\theta}) - \frac{\sum_{s=1}^{T^{\theta}} \delta^s}{\delta^{T^{\theta}}} \gamma \right) \Big\}.$$

Note that T_{FB}^{θ} is bounded and it is the highest t^{θ} such that

$$\delta \beta_{t^{\theta}}^{\theta} \lambda^{\theta} \left[V\left(q_{t^{\theta}}^{\theta}(\underline{c})\right) - \underline{c} q_{t^{\theta}}^{\theta}(\underline{c}) \right] + \delta\left(1 - \beta_{t^{\theta}}^{\theta} \lambda^{\theta}\right) \left[V\left(q^{\theta}(c_{t^{\theta}}^{\theta})\right) - c_{t^{\theta}}^{\theta} q^{\theta}(c_{t^{\theta}}^{\theta}) \right] \geq \gamma + \left[V\left(q^{\theta}(c_{t^{\theta}-1}^{\theta})\right) - c_{t^{\theta}-1}^{\theta} q^{\theta}(c_{t^{\theta}-1}^{\theta}) \right]$$

The intuition is that, by extending the experimentation stage by one additional period, the agent of type θ can learn that $c = \underline{c}$ with probability $\beta_{t^{\theta}}^{\theta} \lambda^{\theta}$. If the agent succeeds, the efficient output will be produced such that $V'\left(q_{t^{\theta}}^{\theta}(\underline{c})\right) = \underline{c}$ for any t^{θ}. The transfer covers the actual cost and no rent is given to the agent. In case the agent fails, the efficient output based on the current *expected* cost, such that $V'\left(q^{\theta}(c_{t^{\theta}}^{\theta})\right) = c_{t^{\theta}}^{\theta}$ for any t^{θ}.[14] The transfer covers the expected cost and no expected rent is given to the agent. For example, if $\lambda^L = 0.2$, $\lambda^H = 0.22$, $\underline{c} = 0.5$, $\overline{c} = 20$, $\beta_0 = 0.4$, $\delta = 0.9$, $\gamma = 2$, and $V = 10\sqrt{q}$, then the first-best termination date for the high type agent is $T_{FB}^{H} = 4$, whereas it is optimal to allow the low type agent to experiment for at most three periods, $T_{FB}^{L} = 3$.

Note that the first-best termination date of the experimentation stage T_{FB}^{θ} is a *non-monotonic* function of the agent's type. This non-monotonicity is a result of two countervailing forces. In any given period of the experimentation stage, the high type is more likely to learn $c = \underline{c}$ (conditional on the actual cost being low) since $\lambda^H > \lambda^L$. This suggests that the principal should allow the high type to experiment longer. However, at the same time, the high type agent becomes relatively more pessimistic with repeated failures. This can be seen by looking at the probability of success conditional on reaching period t, given by $\beta_0\left(1-\lambda^{\theta}\right)^{t-1}\lambda^{\theta}$, over time. In Figure 2, we see that this conditional probability of success for the high type becomes smaller than that for the low type at some point. Given these two countervailing forces, the first-best

---

14 Note that our definition of the first-best stopping time slightly differs from that of Halac et al. (2016). In their paper the discount factor δ does not affect the first-best stopping time. In contrast, we are following the traditional tradeoff between "exploration" and "exploitation" from the literature on experimentation.

stopping time for the high type agent can be shorter or longer than that of the type L agent depending on the parameters of the problem. For example, if we now change λ^H to 0.4 and β_0 to 0.5 then the low type agent is allowed to experiment longer, that is, $T_{FB}^H = 4 < T_{FB}^L = 7$.

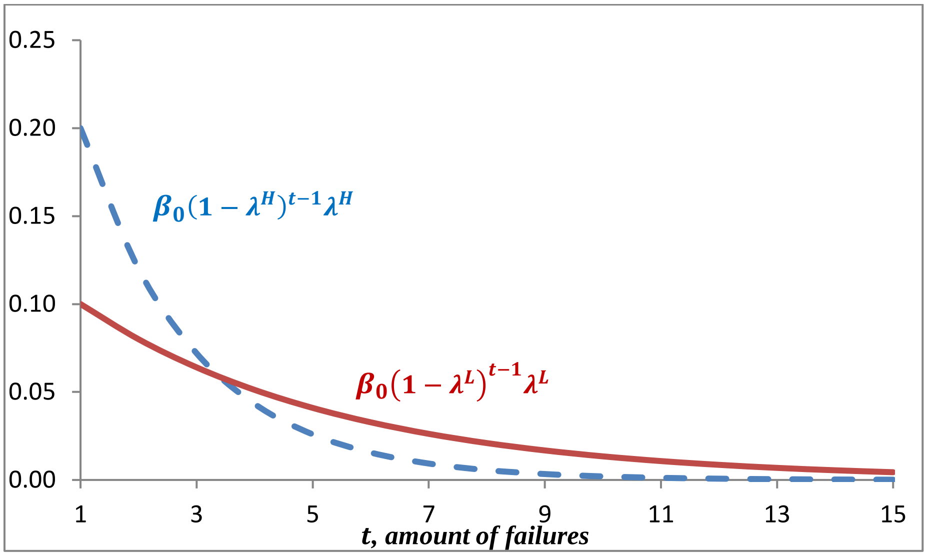

Figure 2. Probability of success with $\lambda^H = 0.4$, $\lambda^L = 0.2$, $\beta_0 = 0.5$.

In Figure 2, we depict the conditional probability of success, $\beta_0\left(1-\lambda^\theta\right)^{t-1}\lambda^\theta$, for both types. For any parameters these curves cross exactly once as can be seen in the figure. Then, for any $t \leq 3$ the high type has a higher conditional probability of success than the low type (conditional on the project being good) since $\lambda^H > \lambda^L$. However, for $t > 3$ the low type is more likely to succeed since the type H agent is now much more pessimistic.

2.2 Asymmetric information

Assume now that the agent privately knows this type. To understand the role of beliefs in generating rent in the production stage, we start with a benchmark without an experimentation stage but with asymmetric information about expected cost of production. The principal can only screen the agents with the output and payments. We obtain a standard second best contract, where the hidden parameter is the expected marginal cost (e.g., Baron-Myerson (1982), Laffont-Tirole (1986)). Suppose in this case, a type θ agent's belief is denoted by β^θ, which implies that the expected cost at the production stage is $c^\theta = \beta^\theta \underline{c} + \left(1 - \beta^\theta\right)\overline{c}$. Suppose that the high-type is more pessimistic than the low type about the cost being low: $\beta^H < \beta^L$. This implies that $c^H > c^L$. As a result, the principal's optimization problem now is:

$$\max_{q(c^H),w(c^H),q(c^L),w(c^L)} \upsilon\left[V\left(q(c^H)\right) - w(c^H)\right] + (1-\upsilon)\left[V\left(q(c^L)\right) - w(c^L)\right]$$

(IR^H) $w(c^H) - c^H q(c^H) \geq 0$,

(IR^L) $w(c^L) - c^L q(c^L) \geq 0$,

$(IC^{H,L})$ $w(c^H) - c^H q(c^H) \geq w(c^L) - c^H q(c^L)$,

$(IC^{L,H})$ $w(c^L) - c^L q(c^L) \geq w(c^H) - c^L q(c^H)$.

It can be easily shown that the optimal contract resembles a standard second-best contract with adverse selection. In particular, (IR^H) and $(IC^{L,H})$ constraints are binding, the low type gets a positive informational rent and produces the first-best output: $V'\left(q(c^L)\right) = c^L$. The high type gets zero rent and his output is distorted as follows: $V'\left(q(c^H)\right) = c^H + \frac{(1-\upsilon)}{\upsilon}(c^H - c^L)$. As in a standard adverse selection model $q^{SB}(c^H) < q^{FB}(c^H) < q^{FB}(c^L) = q^{SB}(c^L)$.

We now return to our main case where an experimentation stage precedes production. Recall that asymmetric information arises in our setting because the two types learn asymmetrically in the experimentation stage, and not because there is any inherent difference in their ability to implement the project. Furthermore, private information can exist only if experimentation fails. If an agent experiences success before the terminal date, T^θ, the true cost $c = \underline{c}$ is revealed. Thus, it is only if there is failure during the entire experimentation stage that an agent has private information about expected cost.

We assume the agent must be paid his expected production costs whether experimentation succeeds or fails. We will call them the *ex post* (i.e., after experimentation) (IR) constraints. Before presenting these new (IR) constraints again, we introduce some notation by defining y_t^θ and x^θ as follows:

$$y_t^\theta \equiv w_t^\theta(\underline{c}) - \underline{c} q_t^\theta(\underline{c}) \text{ for } 1 \leq t \leq T^\theta,$$

$$x^\theta \equiv w^\theta\left(c_{T^\theta}^\theta\right) - c_{T^\theta}^\theta q^\theta\left(c_{T^\theta}^\theta\right).$$

Therefore, the new ex post (IR) constraints can be written as:

$\left(IRS_t^\theta\right)$ $y_t^\theta \geq 0$ for $t \leq T^\theta$,

$\left(IRF_{T^\theta}^\theta\right)$ $x^\theta \geq 0$,

where the S and F are to denote success and failure. Note that they imply that the ex ante (IR) that we presented earlier are redundant. In contrast, if the principal only had to satisfy an ex ante participation constraint $U^\theta\left(\varpi^\theta\right) \geq 0$, she could use the fact that the high type is relatively more likely to succeed (conditional on $c = \underline{c}$) to screen the agent without distorting the duration of the experimentation stage. In other words, since success during the experimentation stage is a random event that is correlated with the agent's type, we can apply well-known ideas from mechanisms à la Crémer-McLean (1985) that says the principal can still receive the first best profit. We explain this next in the context of our model.

To implement the first best, the principal has to counter the incentive of the low type to pretend to be a high type. Relative to the first best payments, the principal can change the payments to the high type only. She can increase the payment in case of success and lower it otherwise, while keeping the high type at zero expected rent, his level of utility in the first best contract. This payment scheme will lower the rent of the *low* type who is less likely to succeed in the experimentation stage. By choosing the appropriate transfers, the principal can obtain the first best level of profit. While this scheme ensures a zero expected rent for each type, it also implies a large positive ex post rent in case of success and a large negative ex post rent in case of failure. When such schemes are allowed, the first best can be reached.[15]

Before presenting the (IC) constraints it is useful to simplify the notation by denoting with P_T^{θ} the probability that an agent of type θ does not succeed during the T periods of the experimentation stage:

$$P_T^{\theta} = 1 - \beta_0 + \beta_0 \left(1 - \lambda^{\theta}\right)^T.$$

Also recall that y_t^{θ} is the rent payment to the θ type who succeeds in period t and x^{θ} is the rent payment to the θ type who failed during the entire experimentation stage. The two incentive constraints are given next.

$$(IC^{L,H})\ \beta_0 \sum_{t=1}^{T^L} \delta^t (1 - \lambda^L)^{t-1} \lambda^L y_t^L + \delta^{T^L} P_{T^L}^L x^L$$

$$\geq \beta_0 \sum_{t=1}^{T^H} \delta^t (1 - \lambda^L)^{t-1} \lambda^L y_t^H + \delta^{T^H} P_{T^H}^L \left[x^H + \Delta c_{T^H} q^H\left(c_{T^H}^H\right)\right],$$

$$(IC^{H,L})\ \beta_0 \sum_{t=1}^{T^H} \delta^t (1 - \lambda^H)^{t-1} \lambda^H y_t^H + \delta^{T^H} P_{T^H}^H x^H$$

$$\geq \beta_0 \sum_{t=1}^{T^L} \delta^t (1 - \lambda^H)^{t-1} \lambda^H y_t^L + \delta^{T^L} P_{T^L}^H \left[x^L - \Delta c_{T^L} q^L\left(c_{T^L}^L\right)\right],$$

We denote the right hand side of $(IC^{L,H})$ constraint by ρ_L, which is the rent to the low type when the constraint is binding. We will see later that it is also possible that the $(IC^{H,L})$ becomes binding in this problem. We denote the right hand side of $(IC^{H,L})$ by ρ_H, which is the rent to the high type when the constraints is binding.

[15] See Theorem 1 in Halac, Kartik, and Liu (2016) for a formal proof in a case without production and a fixed up-front payment.

In this problem, it is the low type who has an incentive to claim to be a high type: since a high type must be given his expected cost following failure, a low type will have to be given a rent to truthfully report his type as his expected cost is lower, that is, $c^L_{T^H} < c^H_{T^H}$.

We now return to the problem with ex post (IR) constraints. The optimal contract is presented in Proposition 1 and derived in Appendix A. The principal has three tools to screen the agent: the length of the experimentation period, the timing of the payments and the output for each type. We examine them below.

Proposition 1

(i) *In the optimal contract, each type may under-experiment or over-experiment relative to the first best.*
(ii) *The low type always receives a rent and* $(IC^{L,H})$ *always binds.*
(iii) *In case 1, when* $(IC^{H,L})$ *is slack, the principal has no restriction when paying the rent to the low type (whether rewarding success or failure) beside what* $(IC^{L,H})$ *requires.*
(iv) *In case 2, when* $(IC^{H,L})$ *binds, the principal must reward early failure and late success when paying the rent to the low type. If, in addition, the high type receives a rent, the principal must pay this rent by rewarding success in the very first period.*
(v) *The high type under-produces relative to the first best output. The low type over-produces if the high type receives a rent and produces at the first best level otherwise.*

Proof: See Appendix A.

2.2.1 The length of the experimentation period: over- or under-experimentation

We already know from the first best contract that the length of the experimentation period is non-monotonic in types. So, in general, we cannot say whether the low type or high type experiments longer. This result extends to the second best and we cannot say, in general, whether T^L is larger or smaller than T^H.

Moreover, we find that each type may under- or over-experiment compared to the first best.[16] The reason is that the driving factor for the rent, and therefore the length of the experimentation stage, is the difference in expected costs between the types after experimentation fails. When the agent lies about his type, his expected cost after failure is different from principal's expected cost. As a result, the agent's informational rent is positively related to the difference in the expected cost different types have. As we saw earlier in Figure 1,

---

[16] One exception: when $(IC^{L,H})$ is not binding, the length of the experimentation period is identical to the first best for the low type.

the difference in the expected cost, $\Delta c_t = c_t^H - c_t^L > 0$, is a *non-monotonic* function of time. Because of that, the principal will sometimes benefit from increasing the length of the experimentation stage to reduce rent.

One important aspect of the length of the experimentation period is whether T^L is larger or smaller than a critical value $\hat{T}^L$. This critical value determines which type is more likely to succeed or fail during the experimentation stage. In any period $t < \hat{T}^L$, the high type who chooses the contract designed for the low type is relatively more likely to succeed than fail compared to the low type. For $t > \hat{T}^L$, the opposite is true. This feature plays an important role in structuring the optimal contract. The critical value $\hat{T}^L$ determines whether the principal will choose to reward success or failure in the optimal contract. We now turn to this issue and provide the precise derivation of $\hat{T}^L$.

2.2.2 The timing of the payments: rewarding success or failure

In part (ii) of proposition 1, we show that $(IC^{L,H})$ is binding, which implies a positive rent ρ_L to the low type. In the dynamic optimization problem, the principal has to determine the optimal payment plan for each type of agent.

If $(IC^{H,L})$ is not binding, we show in case 1 of Proposition 1 that the principal can use any combination of y_t^L and x^L: there is no restriction on how the principal pays ρ_L to the low type as long as $\beta_0 \sum_{t=1}^{T^L} \delta^t (1-\lambda^L)^{t-1} \lambda^L y_t^L + \delta^{T^L} P_{T^L}^L x^L = \delta^{T^H} P_{T^H}^L \Delta c_{T^H} q^H\left(c_{T^H}^H\right)$. Therefore, the principal can reward either success, or failure, or both.

If $(IC^{H,L})$ is binding, the high type has an incentive to claim to be a low type and we are in case 2 of proposition 1. This is the more interesting case and it allows us to characterize the stochastic structure of the dynamic problem. In particular, we will explain the derivation of the critical cut-off period $\hat{T}^L$ mentioned above.

We show in Appendix A that the principal will pay a rent either after success or after failure.[17] We also show in Lemmas 1-3 in Appendix A that the principal will pay a rent after success in at most one period. This allows us to simplify the problem as $y_j^\theta > 0$ for at most one

[17] In the knife-edge case where $T^L = \hat{T}^L$, the principal is free to use a combination of x^L and $y_{T^L}^L$ to pay the low-type's rent.

$j = t$. Given this, the principal must decide whether to pay the rent after success in some period t during the experimentation stage ($y_t^L > 0$) or at the end following failure ($x^L > 0$).

If she pays the rent after success in some period j, such that $1 \leq j \leq T^L$, then

$$y_j^L = \frac{\rho_L}{\beta_0 \delta^j (1-\lambda^L)^{j-1} \lambda^L} > 0, \quad (1)$$

and $y_t^L = 0$ for $t \neq j$, and $x^L = 0$. If she pays after failure, then

$$x^L = \frac{\rho_L}{\delta^{T^L} P_{T^L}^H} > 0, \quad (2)$$

and $y_t^L = 0$ for all $t \leq T^L$, where $P_t^\theta = 1 - \beta_0 + \beta_0 \left(1 - \lambda^\theta\right)^t$.

As the relative likelihood of reaching different periods is different for each type, paying the rent to the low type early or late has different incentive effects. The principal will have to pay close attention to when she pays the rent to the low type. A concern is not to encourage the high type from pretending to be low in order to claim the rent ρ_L. We argue below that misreporting his type is a gamble for the high type: he has a chance to obtain the low-type's rent ρ_L, but he will incur an expected loss in the production stage if he fails during the experimentation stage. If this gamble results in a negative expected return, the high type has no incentive to pretend to be the low type. The principal would then pay zero rent to the high type ($y_t^H = x^H = 0$), while paying ρ_L to the low type using any combination of y_t^L and x^L as long as she does not violate the $(IC^{H,L})$.

In sum, we have two cases. First, if the high type does not want to mimic the low type, , it does not matter if the rent to the low type is paid after success or after failure. This is case 1 in Proposition 1. Second, it is possible that the timing of the payment induces the high type to mimic the low type. To illustrate how the timing of payments matters, suppose that the principal chooses to reward the agent for an early success, say success in period 1 only: $\rho^L = y_1^L > 0$. This payment scheme should look attractive to the high type who is more likely to succeed in period 1 than the low type. This explains why, under certain parameters, the high type incentive constraint $(IC^{H,L})$ may become binding. We call this situation case 2 in Proposition 1. The principal then must modify the payment scheme.

The principal will discourage the high type from mimicking the low type by rewarding early failure or late success. To understand the nature of restrictions imposed by dynamic learning by different types at different speeds, we compare the relative incentive effects on the *high* type of rewarding the *low* type after success or after failure. If the principal rewards the low type only after failure, the high type's expected utility from misreporting (i.e., the RHS of the $(IC^{H,L})$ constraint) is:

$$U_F^H(\varpi^L) = \delta^{T^L} P_{T^L}^H \left[x^L - \Delta c_{T^L} q^L\left(c_{T^L}^L\right)\right],$$

which can be re-written using (2) as,

$$= \frac{\delta^{T^L} P_{T^L}^H}{\delta^{T^L} P_{T^L}^L} \rho_L - \delta^{T^L} P_{T^L}^H \Delta c_{T^L} q^L\left(c_{T^L}^L\right).$$

If the principal rewards the low type only after success in some period j, with $1 \leq j \leq T^L$, the high type's expected utility from misreporting (i.e., the RHS of the $IC^{H,L}$ constraint) is:

$$U_S^H(\varpi^L) = \beta_0 \delta^j (1-\lambda^H)^{j-1} \lambda^H y_j^L - \delta^{T^L} P_{T^L}^H \Delta c_{T^L} q^L\left(c_{T^L}^L\right),$$

which can be rewritten using (1) as,

$$= \frac{\beta_0 \delta^j (1-\lambda^H)^{j-1} \lambda^H}{\beta_0 \delta^j (1-\lambda^L)^{j-1} \lambda^L} \rho_L - \delta^{T^L} P_{T^L}^H \Delta c_{T^L} q^L\left(c_{T^L}^L\right).$$

Comparing these two expressions, we can study the nature of the gamble for the high type when he misreports and derive the optimal payment schemes. The choice to reward the low type after success or failure depends on which scheme yields a lower RHS of $(IC^{H,L})$. The first term in each expression is the expected gain from obtaining ρ_L by misreporting, and the second is the possibility of incurring a loss if experimentation fails. The second term is identical in the two expressions so we focus on the first term.

If the low type is rewarded for success in period j, the relative probability of success of a high type in period j is:

$$\frac{\beta_0 \delta^j (1-\lambda^H)^{j-1} \lambda^H}{\beta_0 \delta^j (1-\lambda^L)^{j-1} \lambda^L},$$

which is the coefficient of ρ_L in $U_S^H(\varpi^L)$. If the low type is rewarded only after failure, the relative probability of failure conditional on reaching T^L is:

$$\frac{\delta^{T^L} P^H_{T^L}}{\delta^{T^L} P^L_{T^L}},$$

which is the coefficient of ρ_L in $U_F^H(\varpi^L)$. The relative probability of success for a high type decreases with j, while the relative probability of failure for a high type is constant given T^L. We show in the appendix that there is a j such that the RHS of $(IC^{H,L})$ under success or failure equal each other. This is achieved when the two coefficients of ρ_L are equal to each other. We can now formally define $\hat{T}^L(=j)$ by setting the two coefficients equal:

$$\frac{(1-\lambda^H)^{\hat{T}^L-1}\lambda^H}{(1-\lambda^L)^{\hat{T}^L-1}\lambda^L} \equiv \frac{P^H_{T^L}}{P^L_{T^L}}.$$

Thus, if the principal wants to reward the low type after success, it will only be optimal if the experimentation stage lasts long enough. If $T_L < \hat{T}^L$, then, $\frac{(1-\lambda^H)^{j-1}\lambda^H}{(1-\lambda^L)^{j-1}\lambda^L} > \frac{P^H_{T^L}}{P^L_{T^L}}$ for all j and the high type will have an advantage over the low type in obtaining any reward given after success. To provide the rent ρ_L to the low type, the principal will have to reward failure. If $T_L > \hat{T}^L$, the principal does not have to reward the low type for failure but can reward success instead. Indeed, it is optimal to pay the reward after success in the last period as $\frac{\beta_0\delta^j(1-\lambda^H)^{j-1}\lambda^H}{\beta_0\delta^j(1-\lambda^L)^{j-1}\lambda^L}$ is declining over time and the rent is the smallest when $j = T^L$.

As mentioned earlier, even when the principal uses this optimal payment scheme to pay the rent to the low type, it is possible that the high type may want to mimic the low type. When this is the case, we show in the appendix that the principal will pay the high type's rent by rewarding success in the first period only. Intuitively, this is the period when success is most likely to come from a high type than a low type.

It is unusual to have both types earning a rent and we will provide some intuition using an example. Consider $V(q) = 3.5\sqrt{q}$, $\beta_0 = 0.7$, $\underline{c} = 0.1$, $\overline{c} = 10$, $\delta = 0.9$, $\gamma = 1$, $\lambda^L = 0.14$, $\lambda^H = 0.35$. Then $\hat{T}^L = 5$ and the principal optimally chooses $T^H = 8$, $T^L = 12$ and grant rent only to the low type. However, if we increase only the parameter $\lambda^H = 0.82$, then $\hat{T}^L = 3$ and the principal optimally chooses $T^H = 3$, $T^L = 11$ and grants rent to both types.

Intuitively, when the high type is very efficient in learning the true cost ($\lambda^H = 0.82$), by rewarding the low type for success the principal makes it more likely that the high type will receive this reward. Thus, ideally the principal would like to reward the low type for a sequence of failures for as many periods as possible. However, as T^L increases, the high type becomes more confident that the project is bad and failure is more likely. Rewards after failure now become attractive for the high type. This happens once $T^L > \hat{T}^L$. At that point, the principal should no longer reward failure but must switch to reward success from the low type.

2.2.3 The output: under- or over- production

Finally, the principal can use the choice of output to screen the types and limit the rent to both types. We derive the formal output scheme in the appendix but present the intuition here. If experimentation was successful, there is no asymmetric information and no reason to distort the output. Both types produce the first best output. If experimentation failed to reveal the cost, asymmetric information will induce the principal to distort the output to limit the rent. This is a familiar result in contract theory. In a standard second best contract à la Baron-Myerson, the type who receives rent produces the first best level of output while the type with no rent under-produces relative to the first best.

When only the low type's incentive constraint binds, the low type produces the first best output while the high type under-produces relative to the first best. To limit the rent of the low type, the high type is asked to produce a lower output.

The dynamic feature of our model, however, creates a possibility that the high type's incentive constraint also binds. To limit the rent of the high type, the principal will then *increase* the output of the low type and require over-production relative to the first best. To understand the intuition behind this result, recall that the rent of the high type mimicking the low type has two components. The first component is the rent promised to the low type after failure in the experimentation stage. The second component is negative and comes from the higher expected cost of producing the output required from the low type $q^L(c^L_{T^L})$. By making this output higher, the principal can strengthen the negative component and lower the rent of the high type.

3. Extensions

3.1 Identical length of experimentation stage for *both* types

As a special case of our model we consider an environment where it is not feasible to screen the agent with the duration of the experimentation stage.[18] That is, the principal must choose an identical length of the experimentation stage for both types. We prove that $(IC^{H,L})$ is never binding. Therefore, the main message of this section is that it is the principal's desire to have different lengths of the experimentation stage that resulted in $(IC^{H,L})$ being binding in the main model.

A contract is now defined formally by $\varpi^{\hat{\theta}} = \left(\tilde{T}, \left\{w_t^{\hat{\theta}}(\underline{c}), q_t^{\hat{\theta}}(\underline{c})\right\}_{t=1}^{\tilde{T}}, \left\{w^{\hat{\theta}}\left(c_{\tilde{T}}^{\hat{\theta}}\right), q^{\hat{\theta}}\left(c_{\tilde{T}}^{\hat{\theta}}\right)\right\}\right)$, where $\tilde{T}$ is the maximum duration of the experimentation stage *regardless* of the announced type.

First, we re-examine the first best case by assuming that the agent's type θ is common knowledge *before* the principal offers the contract.

Since the expected cost is rising as long as success is not obtained for both types, the first-best solution is characterized by a termination date $\tilde{T}_{FB}$, where:

$$\tilde{T}_{FB} \in \underset{\tilde{T}}{arg\,max} \left\{ \begin{array}{c} \upsilon\pi^H(\varpi^H) + (1-\upsilon)\pi^L(\varpi^L) = \\ \upsilon \left(\begin{array}{c} \beta_0 \sum_{t=1}^{\tilde{T}} \delta^t (1-\lambda^H)^{t-1}\lambda^\theta \left(V\left(q_t^H(\underline{c})\right) - \underline{c}q_t^H(\underline{c}) - \frac{\sum_{s=1}^{t}\delta^s}{\delta^t}\gamma\right) \\ +\delta^{\tilde{T}}\left(1-\beta_0+\beta_0(1-\lambda^H)^{\tilde{T}}\right)\left(V\left(q^H(c_{\tilde{T}}^H)\right) - c_{\tilde{T}}^H q^H(c_{\tilde{T}}^H) - \frac{\sum_{s=1}^{\tilde{T}}\delta^s}{\delta^{\tilde{T}}}\gamma\right) \end{array} \right) \\ +(1-\upsilon) \left(\begin{array}{c} \beta_0 \sum_{t=1}^{\tilde{T}} \delta^t (1-\lambda^L)^{t-1}\lambda^\theta \left(V\left(q_t^L(\underline{c})\right) - \underline{c}q_t^L(\underline{c}) - \frac{\sum_{s=1}^{t}\delta^s}{\delta^t}\gamma\right) \\ +\delta^{\tilde{T}}\left(1-\beta_0+\beta_0(1-\lambda^L)^{\tilde{T}}\right)\left(V\left(q^L(c_{\tilde{T}}^L)\right) - c_{\tilde{T}}^L q^L(c_{\tilde{T}}^L) - \frac{\sum_{s=1}^{\tilde{T}}\delta^s}{\delta^{\tilde{T}}}\gamma\right) \end{array} \right) \end{array} \right\}$$

The optimal $\tilde{T}_{FB}$ is bounded and it is the highest t such that

$$\begin{array}{c} \upsilon\left(\delta\beta_t^H\lambda^H\left[V\left(q_t^H(\underline{c})\right) - \underline{c}q_t^H(\underline{c})\right] + \delta(1-\beta_t^H\lambda^H)\left[V\left(q^H(c_t^H)\right) - c_t^H q^H(c_t^H)\right]\right) \\ +(1-\upsilon)\left(\delta\beta_t^L\lambda^L\left[V\left(q_t^L(\underline{c})\right) - \underline{c}q_t^L(\underline{c})\right] + \delta(1-\beta_t^L\lambda^L)\left[V\left(q^L(c_t^L)\right) - c_t^L q^L(c_t^L)\right]\right) \\ \geq \gamma + \upsilon\left[V\left(q^H(c_{t-1}^H)\right) - c_{t-1}^H q^H(c_{t-1}^H)\right] + (1-\upsilon)\left[V\left(q^L(c_{t-1}^L)\right) - c_{t-1}^L q^L(c_{t-1}^L)\right] \end{array}$$

[18] For example, the FDA requires all the firms to go through the same amount of trials before they are allowed to release new drugs on the market.

Recall that the first-best termination date of the experimentation stage T^{θ}_{FB} is a *non-monotonic* function of the agent's type. Since the expected cost is rising as long as success is not obtained for both types, we immediately conclude that

$$\min_{\theta} T^{\theta}_{FB} \leq \tilde{T}_{FB} \leq \max_{\theta} T^{\theta}_{FB}.$$

This implies that, when the principal is restricted to set the same duration of the experimentation for both agents even if the type of the agent was known, one type will over experiment whereas the other will under experiment.

The principal's optimization problem when the type of the agent is *not known* is to choose contracts ϖ^H and ϖ^L to

$$\max_{\varpi^H,\varpi^L \in \varpi} \begin{matrix} \upsilon \left\{ \begin{matrix} \beta_0 \sum_{t=1}^{\tilde{T}} \delta^t (1-\lambda^H)^{t-1}\lambda^H \left[V\left(q^H_t(\underline{c})\right) - y^H_t - \underline{c} q^H_t(\underline{c}) - \frac{\sum_{s=1}^{t}\delta^s \gamma}{\delta^t} \right] \\ + \delta^{\tilde{T}} P^H_{\tilde{T}} \left[V\left(q^H(c^H_{\tilde{T}})\right) - x^H - c^H_{\tilde{T}} q^H(c^H_{\tilde{T}}) - \frac{\sum_{s=1}^{\tilde{T}}\delta^s \gamma}{\delta^{\tilde{T}}} \right] \end{matrix} \right\} \\ +(1-\upsilon) \left\{ \begin{matrix} \beta_0 \sum_{t=1}^{\tilde{T}} \delta^t (1-\lambda^L)^{t-1}\lambda^L \left[V\left(q^L_t(\underline{c})\right) - y^L_t - \underline{c} q^L_t(\underline{c}) - \frac{\sum_{s=1}^{t}\delta^s \gamma}{\delta^t} \right] \\ + \delta^{\tilde{T}} P^L_{\tilde{T}} \left[V\left(q^L(c^L_{\tilde{T}})\right) - x^L - c^L_{\tilde{T}} q^L(c^L_{\tilde{T}}) - \frac{\sum_{s=1}^{\tilde{T}}\delta^s \gamma}{\delta^{\tilde{T}}} \right] \end{matrix} \right\} \end{matrix} \text{ subject to}$$

$(IC^{H,L})$ $\beta_0 \sum_{t=1}^{\tilde{T}} \delta^t (1-\lambda^H)^{t-1}\lambda^H y^H_t + \delta^{\tilde{T}} P^H_{\tilde{T}} x^H$

$\geq \beta_0 \sum_{t=1}^{\tilde{T}} \delta^t (1-\lambda^H)^{t-1}\lambda^H y^L_t + \delta^{\tilde{T}} P^H_{\tilde{T}} [x^L - \Delta c_{\tilde{T}} q^L(c^L_{\tilde{T}})]$,

$(IC^{L,H})$ $\beta_0 \sum_{t=1}^{\tilde{T}} \delta^t (1-\lambda^L)^{t-1}\lambda^L y^L_t + \delta^{\tilde{T}} P^L_{\tilde{T}} x^L$

$\geq \beta_0 \sum_{t=1}^{\tilde{T}} \delta^t (1-\lambda^L)^{t-1}\lambda^L y^H_t + \delta^{\tilde{T}} P^L_{\tilde{T}} [x^H + \Delta c_{\tilde{T}} q^H(c^H_{\tilde{T}})]$,

$\left(IRS^{\theta}_t\right)$ $y^{\theta}_t \geq 0$ for $t \leq T^{\theta}$, and

$\left(IRF^{\theta}_{T^{\theta}}\right)$ $x^{\theta} \geq 0$.

Proposition 2. *If the duration of the experimentation stage must be chosen independently of the announced type, the high type gets no rent. The principal can choose any combinations of payments to the low type such that* $\rho^L = \delta^{\tilde{T}} P^L_{\tilde{T}} \Delta c_{\tilde{T}} q^H(c^H_{\tilde{T}})$. *Each type may under-experiment or over-experiment relative to the first best.*

Proof: See Appendix B.

Based on propositions 1 and 2, we learn that it is the principal's desire to have different lengths of the experimentation stage that results in $(IC^{H,L})$ being binding. If $\hat{T}$ is the same for both types, $(IC^{H,L})$ is not binding as in a static problem – the gamble is negative. Indeed, using $(IC^{H,L})$, the high type pretending to be the low type receives

$$\delta^{\tilde{T}} P_{\tilde{T}}^{H} \Delta c_{\tilde{T}} \left(q^{H}\left(c_{\tilde{T}}^{H}\right) - q^{L}\left(c_{\tilde{T}}^{L}\right) \right).$$

We show in appendix B that we have $q^{H}\left(c_{\tilde{T}}^{H}\right) < q_{FB}^{H}\left(c_{\tilde{T}}^{H}\right) < q^{L}\left(c_{\tilde{T}}^{L}\right)$. As a result, the high type receives a strictly negative utility if he mimics a low type. [19]

As in the main model, there is no distortion in the output relative to the first-best level after success. After failure, only the high type output is distorted. There is underproduction by the high type when experimentation fails: $q_{SB}^{H}\left(c_{\tilde{T}}^{H}\right) < q_{FB}^{H}\left(c_{\tilde{T}}^{H}\right)$.

3.2. Success might be hidden: ex post moral hazard

A notable result from the main section was that the principal may want to reward failure, or wait until later periods to reward success. The implementation of schemes with such properties relies on our assumption that the outcome of experiments in each period is publicly observable. If the agent were able to suppress a finding of success[20], he would gain by hiding success, or postponing the revelation of success.

In this subsection, we allow the agent to engage in *ex post moral hazard* by hiding success when it occurs. Specifically, we assume that success is privately observed by the agent, and that an agent who finds success in some period j can choose to announce or reveal it at any period $t \geq j$ that he finds optimal. Thus, we assume that success generates hard information that can be presented to the principal when desired, but it cannot be fabricated. The agent's decision to reveal success is affected not only by the payment and the output tied to success/failure in the particular period j, but also by the payment and output in all subsequent periods of the experimentation stage. The dynamic optimization problem for the principal when success is privately observed by the agent becomes more complex as the principal has to deal with both

[19] In this case, the intuition from the static second best contract applies as in our benchmark case without experimentation but with asymmetric information about expected cost. We found that the $(IC^{H,L})$ is not binding since the high type also produces less than the low type.

[20] In some settings, the principal can observe success easily. For example, a clinical research organization hired by a pharmaceutical company may have difficulty hiding a revolutionary drug's success. However, success might be much more difficult to ascertain when information gathering does not involve extreme outcomes.

adverse selection and ex post moral hazard problems simultaneously. We show that our finding of rewarding the agent after failure does not depend on success being observed publicly.

Note first that if the agent succeeds but hides it, the principal and the agent's beliefs are different at the production stage: the principal's expected cost is given by $c^{\theta}_{T^{\theta}}$ while the agent knows the true cost is $\underline{c}$. In addition to the existing (IR) and (IC) constraitns, the optimal scheme must now satisfy two new ex post moral hazard constraints:

(EMH^{θ}) $y^{\theta}_{T^{\theta}} \geq x^{\theta} + (c^{\theta}_{T^{\theta}} - \underline{c})q^{\theta}(c^{\theta}_{T^{\theta}})$ for $\theta = H, L$, and

(EMP^{θ}_{t}) $y^{\theta}_{t} \geq \delta y^{\theta}_{t+1}$ for $t \leq T^{\theta} - 1$.

The first constraint makes it unprofitable for the agent to *hide* success in the last period. The second one makes it unprofitable to *postpone* revealing success in prior periods. The two together imply that the agent cannot gain by postponing or hiding success. The principal's problem is exacerbated by having to address the ex post moral hazard constraints in addition to all the constraints presented before. First, as formally shown in the appendix C, both $(IC^{H,L})$ and $(IC^{L,H})$ may be slack, and either or both may be binding.[21] Since the two (EM) constraints imply that both types will receive rent due to ex post moral hazard, these rents may be sufficient to satisfy the (IC) constraints. Second, private observation of success increases the cost of paying a reward after failure. When the principal rewards failure with $x^{\theta} > 0$, the (EMH) constraint forces her to also reward success in the last period ($y^{\theta}_{T^{\theta}} > 0$ because of (EMH)) and in all previous periods ($y^{\theta}_{t} > 0$ because of (EMP)). However, we show below that it can still be optimal to reward failure.

Proposition 3. *When success can be hidden, the principal must reward success in every period for each type. When both the $(IC^{L,H})$ and $(IC^{H,L})$ constraints bind and the optimal $T^{L} \leq \hat{T}^{L}$, it is optimal to reward failure for the low type.*

Proof: See Appendix C.

While details and the formal proof are in the appendix, we now provide some intuition why rewarding failure remains optimal even when the agent privately observes success. We also provide an example below where this occurs in equilibrium. Recall from the previous section, where success is publicly observed, that the principal relied on rewarding a low type after failure when the high type's $(IC^{H,L})$ was binding. Since a low type was relatively more likely to fail compared to a high type when the experimentation period is relatively short, rewarding the low type when experimentation failed was the best way to deter misreporting by the high type. This intuition remains intact even when success is privately observed since the relative probability of

21 Unlike the case when success is public, the $(IC^{L,H})$ may not always be binding.

success between types is not affected by the two ex post moral hazard constraints above. An increase of \$1 in x^{θ} causes an increase of \$1 in $y^{\theta}_{T^{\theta}}$, which in turn causes an increase in all the previous y^{θ}_{t} according to the discount factor. Therefore, the increases in $y^{\theta}_{T^{\theta}}$ and y^{θ}_{t} are not driven by the relative probability of success between types. And, just as in Proposition 1, we again find that it is optimal to reward failure when the low type experiments for a relatively brief length of time and both $IC^{H,L}$ and $IC^{L,H}$ are binding. For example, when $\beta_0 = 0.7$ $\gamma = 2$, $\lambda^L = 0.28$, $\lambda^H = 0.7$ the principal optimally chooses $T^H = 1$, $T^L = 2$ and grants rent only to the low type by rewarding failure since $\hat{T}^L = 3$.

While we have focused on how the ex post moral hazard affects the benefit of rewarding failure, it is clear that those constraints also affect the other optimal variables of the contract. For instance, the constraint (EMP) can be relaxed by decreasing either T^{θ} (which will decrease $c^{\theta}_{T^{\theta}}$) or $q^{\theta}\left(c^{\theta}_{T^{\theta}}\right)$. So we expect a shorter experimentation stage and a lower output when success can he hidden.

3.3. Learning bad news

In our main model, we assumed that the object of experimentation was to look for good news, with success meaning the discovery of low cost. In this section, we show that our main results survive if the object of experimentation is to seek bad news, where success in an experiment means discovery of high cost $c = \overline{c}$. For instance, stage 1 of a drug trial looks for bad news by testing the safety of the drug. Following the literature on experimentation we call an event of observing $c = \overline{c}$ by the agent "success" although this is a bad news for the principal. In contrast to the case with good news, if the agent's type were common knowledge, the principal and agent both become more *optimistic* if success is not achieved in a particular period and *relatively more optimistic* when the agent is a high type than a low type. Also, as time goes by without learning that the cost is high, the expected cost becomes lower due to Bayesian updating and converges to $\underline{c}$.

Denoting again by β^{θ}_{t}, the updated belief of agent θ that the cost is actually *high* at the beginning of period t after $t-1$ failures, for $t > 1$, we have:

$$\beta^{\theta}_{t} = \frac{\beta_0\left(1-\lambda^{\theta}\right)^{t-1}}{\beta_0\left(1-\lambda^{\theta}\right)^{t-1}+\left(1-\beta_0\right)}.$$

The agent θ's expected cost is then $c^{\theta}_{t} = \beta^{\theta}_{t}\overline{c} + \left(1-\beta^{\theta}_{t}\right)\underline{c}$. An agent of type θ, announcing his type as $\hat{\theta}$, receives expected utility $U^{\theta}\left(\varpi^{\hat{\theta}}\right)$ at time zero from a contract $\varpi^{\hat{\theta}}$:

$$U^{\theta}\left(\varpi^{\hat{\theta}}\right) = \beta_0 \sum_{t=1}^{T^{\hat{\theta}}} \delta^t \left(1-\lambda^{\theta}\right)^{t-1}\lambda^{\theta}\left(w^{\hat{\theta}}_{t}(\overline{c}) - \overline{c}q^{\hat{\theta}}_{t}(\overline{c})\right)$$

$$+\delta^{T^{\hat{\theta}}}\left(1-\beta_0+\beta_0\left(1-\lambda^{\theta}\right)^{T^{\hat{\theta}}}\right)\left(w^{\hat{\theta}}\left(c^{\hat{\theta}}_{T^{\hat{\theta}}}\right)-c^{\theta}_{T^{\hat{\theta}}}q^{\hat{\theta}}\left(c^{\hat{\theta}}_{T^{\hat{\theta}}}\right)\right).$$

Consider first the case when the type of the agent is known by the principal before the contract is offered. Since the expected cost is *decreasing* over time as long as success is not observed, the first-best solution is characterized by a termination date T^{θ}_{FB}.

As with learning good news, the first-best termination date of the experimentation stage is a *non-monotonic* function of the agent's type. In addition, the difference in the expected cost is now negative, $\Delta c_t = c^H_t - c^L_t < 0$ since the H type is relatively more optimistic after the same amount of failures.

Under asymmetric information about the agent's type, the intuition behind the key incentive problem is again similar to that under learning good news. However, it is now the high type who has an incentive to claim to be a low type. Given the same length of experimentation, following failure, the expected cost is higher for the low type. Thus, a high type now has an incentive to claim to be a low type: since a low type must be given his expected cost following failure, a high type will have to be given a rent to truthfully report his type as his expected cost is lower, that is, $c^H_{T^L} < c^L_{T^L}$. We denote the right hand side of $(IC^{H,L})$ constraint by ρ_H, which is the rent to the high type when the constraint is binding.

$$(IC^{H,L})\ \beta_0\sum_{t=1}^{T^H}\delta^t\,(1-\lambda^H)^{t-1}\lambda^H[w^H_t(\overline{c})-\overline{c}q^H_t(\overline{c})]+\delta^{T^H}P^H_{T^H}\left[w^H\left(c^H_{T^H}\right)-c^H_{T^H}q^H\left(c^H_{T^H}\right)\right]$$

$$\geq \beta_0\sum_{t=1}^{T^L}\delta^t\,(1-\lambda^H)^{t-1}\lambda^H[w^L_t(\overline{c})-\overline{c}q^L_t(\overline{c})]+\delta^{T^L}P^H_{T^L}\left[w^L\left(c^L_{T^L}\right)-c^H_{T^L}q^L\left(c^L_{T^L}\right)\right],$$

It is also possible that the $(IC^{L,H})$ becomes binding in this problem:

$$(IC^{L,H})\ \beta_0\sum_{t=1}^{T^L}\delta^t\,(1-\lambda^L)^{t-1}\lambda^L[w^L_t(\overline{c})-\overline{c}q^L_t(\overline{c})]+\delta^{T^L}P^L_{T^L}\left[w^L\left(c^L_{T^L}\right)-c^L_{T^L}q^L\left(c^L_{T^L}\right)\right]$$

$$\geq \beta_0\sum_{t=1}^{T^H}\delta^t\,(1-\lambda^L)^{t-1}\lambda^L[w^H_t(\overline{c})-\overline{c}q^H_t(\overline{c})]+\delta^{T^H}P^L_{T^H}\left[w^H\left(c^H_{T^H}\right)-c^L_{T^H}q^H\left(c^H_{T^H}\right)\right].$$

The details of the optimization problem mirror the case of Proposition 1. Regarding the length of the experimentation periods, we find again that each type may over- or under- experiment relative to the first best.

Regarding the timing of payments, we find that the relative likelihood ratios of success and failure continue to characterize the optimal payment scheme as it was the case with learning good news. We find similar restrictions when both (IC) constraints bind as in proposition 1. Specifically, the high-type's rent is paid in case of success in the first period and the low-type's rent is paid after early failure (when $T^L < \hat{T}^L$) or after success in the last period (when $T^L > \hat{T}^L$). The parallel between good news and bad news is remarkable but not difficult to explain. In both cases, the agent is looking for news. The types determine how good the agent is at obtaining this news. The contract gives incentives for each type of agent to reveal his type, not the actual news.

The type of news, however, determines the optimal production and length of experimentation decisions. Regarding the distortion in the output, we find that the low type under-produces relative to the first best output. The high type over-produces if the low type receives a rent and produces at the first best level otherwise. This is also similar to proposition 1 except that the distortions in output are switched for the two types.

Proposition 4:

(i) In the optimal contract, each type may under-experiment or over-experiment relative to the first best.

(ii) The high type always receives a rent and $(IC^{H,L})$ always binds.

(iii) In case 1, when $(IC^{L,H})$ is slack, the principal has no restriction when paying the rent to the high type (whether rewarding success or failure) beside what $(IC^{H,L})$ requires.

(iv) In case 2, when $(IC^{L,H})$ binds, the principal must pay the high type rent by rewarding success in the very first period. If the low type receives a rent, this rent is paid by rewarding early failure and late success.

(v) The low type under-produces relative to the first best output. The high type over-produces if the low type receives a rent and produces at the first best level otherwise.

Proof: See Appendix D.

4. Conclusions

In this paper, we have studied the interaction between experimentation and production where the length of the experimentation stage determines the degree of asymmetric information at the production stage. This interaction affects the optimal project scale after success as well as after failure. While success in experimentation typically resolves uncertainty in a two-armed bandit model, learning still occurs after successive failures, and it determines the scale of the project. While there has been much recent attention on studying incentives for experimentation in two-armed bandit settings, details of the production stage are typically suppressed to focus on incentives for exploration. In reality, each stage impacts the other in interesting ways and our paper is a step towards studying this interaction.

There is also a significant literature on endogenous information gathering in contract theory but typically relying on static models of learning. By modeling experimentation in a dynamic setting, we have endogenized the degree of asymmetry of information in a principal agent model and also related it to the length of the learning stage.

By analyzing the stochastic structure of the dynamic problem, we clarify how the principal can rely on the relative probabilities of success and failure of the two types in order to screen them. The rent to a high type should come after early success and to the low type for late success. If the experimentation stage is not long enough, the principal has no recourse but to pay the low type's rent after failure, which is another novel result. While our main section relies on publicly observed success and experimenting for 'good news', we show that our main insights survive if the agent can hide success or if we changed to model to learn 'bad news'. Without a production stage with a scalable project size after failure, there would be under experimentation relative to the first best. With a scalable project size, we find a new result that over-experimentation can be also optimal. Over production can occur in the production stage. Analogous results are obtained whether we consider a good news or a bad news model of experimentation.

References

Bergemann D. and U. Hege (1998). Venture capital financing, moral hazard, and learning. *Journal of Banking & Finance* 22 (6-8), 703-735.

Bergemann D. and U. Hege (2005). The Financing of Innovation: Learning and Stopping. *Rand Journal of Economics* 36 (4), 719-752.

Bolton P. and Harris C. (1999). Strategic Experimentation. *Econometrica* 67(2), 349–374.

Cremer J. and Khalil F. (1992). Gathering Information before Signing a Contract. *American Economic Review* 82(3), 566-578.

Cremer J., Khalil F. and Rochet J.-C. (1992). Contracts and productive information gathering. *Games and Economic Behavior* 25(2), 174-193.

Gerardi D. and Maestri L. (2012). A Principal-Agent Model of Sequential Testing. *Theoretical Economics* 7, 425–463.

Gomes R., Gottlieb D. and Maestri L. (2016). Experimentation and Project Selection: Screening and Learning. *Games and Economic Behavior* 96, 145-169.

Halac M., N. Kartik and Q. Liu (2016). Optimal Contracts for Experimentation. *Review of Economic Studies, forthcoming*.

Horner J. and L. Samuelson (2013). Incentives for Experimenting Agents. *RAND Journal of Economics* 44(4), 632-663.

Keller G., Rady S. and Cripps M. (2005). Strategic Experimentation with Exponential Bandits. *Econometrica* 73, 39–68.

Khramer D. and Strausz R. (2011) Optimal Procurement Contracts with Pre-Project Planning. *Review of Economic Studies* 78, 1015-1041.

Lewis T. and Sappington D. (1997) Information Management in Incentive Problems. *Journal of Political Economy* 105(4), 796-821.

Manso G. (2011). Motivating Innovation. *Journal of Finance* 66 (2011), 1823–1860.