Motivating Information Acquisition in Strategic Settings

Andres Zambrano, UCLA *

November 27, 2011

JOB MARKET PAPER

Abstract

Successful innovators have become billionaires by generating breakthrough technologies that are later widely adopted in the society. This paper proposes a model of delegated expertise to explain why this is (constrained) efficient. To develop intuition, I first study the optimal design of contracts when a principal delegates a decision to a single agent of whether to pursue a risky project or a safe one. Before taking the decision, the agent can acquire unobservable information about the risky project by exerting an unobservable effort that determines the quality of the information. The optimal contract suggests that the principal should reward the agent for outcomes that are significantly better than the safe return to encourage more information acquisition and the selection of the desired project by the principal. I then apply this structure to study the problem faced by a population of agents who must decide between two technologies and the acquired information becomes public, thus creating incentives for free-riding. The optimal contract divides the team between non-experimenters and a few experimenters, and splits the total returns among experimenters when the unknown project yields significantly greater returns than the safe project.

Keywords: Information Acquisition, Strategic Experimentation, Private Information, Contract

JEL codes: D82, D83, D86, C72

---

*E-mail: azambrano@ucla.edu. Please download the most recent version at https://sites.google.com/site/andreszam/. The author is indebted to Andrew Atkeson and Hugo Hopenhayn for their helpful advice during all these years. I also gratefully acknowledge the comments from Simon Board, Jernej Copic, Fernando Leibovici, Moritz Meyer-ter-Vehn, Ichiro Obara, John Riley, Venky Venkateswaran, Pierre-Olivier Weill, and Bill Zame, as well as participants of the Macro Lunch seminar, Macro Proseminar, IO Proseminar and Theory Proseminar at UCLA, and Universidad del Rosario. Funding from the Colombian Central Bank and NBER is also acknowledged. Remaining errors are mine.

1 Introduction

Recognized innovators such as Bill Gates and Steve Jobs have become billionaires by creating breakthrough technologies that are later widely used in the society. Could we achieve similar levels of innovation without having to pay them so much for such new technologies? Why is it that other innovators who create more marginal improvements are not paid as much? (Ebersberger et al., 2008; Marsili and Salter, 2005) This paper proposes a model of delegated expertise with multiple agents to explain these features.

Models of delegated expertise were first proposed by Lambert (1986) and Demski and Sappington (1987). In such models a principal must hire an agent to decide between a risky and a safe project. Before taking the decision, the agent can acquire unobservable information about the risky project by exerting unobservable effort. This framework shares features of moral hazard and hidden information, and incentives must be used to motivate both information acquisition and the (partial) revelation of the obtained information. Therefore, incentives in this scenario are potentially different than the ones in standard moral hazard problems. I use this one principal-one agent model to develop intuition on the results that are later obtained for the team problem case. The first contribution of this paper is to characterize the optimal contract in this simplified environment.

The studied setup is very general and the only restriction imposed is that signals can be ordered as in Milgrom (1981). In contrast to the standard moral hazard literature, the unobservable action taken by the agent determines the precision of a privately observed signal and thus generates mean-preserving spreads on the ex-ante distribution of the unknown return. Under limited liability for both individuals, the optimal wages suggest that the agent should be rewarded with the return of the risky project only if it is significantly better than the safe return.

The intuition why contracts reward experimenters only for extreme good outcomes has two components. The moral hazard component suggests that we should reward the agent whenever the observed outcome of the risky project is greater than the safe return, since a greater effort reduces the probability of choosing the risky project when in fact it is worse than the safe one. However, a higher effort does not affect importantly the probability of adopting the risky project when the unknown return is either close to the safe return or close to the maximum possible return. Thus, the moral hazard alone suggests incentives should be low-powered for outcomes close to the safe return or the highest possible return, and high powered otherwise.

There is a second component associated with the adverse selection problem. The principal must also provide incentives for the agent to choose the risky project whenever she observes a sufficiently high signal. Therefore, the higher the signal the agent receives, the more the principal wants him to undertake the risky project. Since the signal and the unknown return are positively correlated, the principal wants to increase the reward for the agent when higher returns are realized.

Since the principal wants to pay the agent the least possible, she will pay nothing to the agent for outcomes that are worse than the safe project. It will also pay nothing for outcomes that are better but close enough to the safe return since the moral hazard is not as strong in such region. As argued before, moral hazard is also not as strong for extreme positive outcomes; however, that is the same region where the adverse selection component suggests the strongest incentives must be provided to encourage the choice of the risky project. Therefore, the optimal contract rewards the agent with the whole output when the realized return is significantly better than the safe one. Although this instrument is highly nonlinear, it is not enough to achieve a first best since the effects of a greater precision are shown to be bounded.

I then use the structure to study optimal contracts in team environments where a population of agents can produce public information about the unknown technology. The acquired information then is used by every agent to chose between the risky and the safe project. In this case, the public quality of the information creates incentives for free-riding, thus decreasing the overall amount of available information recollected. Agents prefer to wait until others incur in the cost of experimentation since the next stage they will observe the relevant signals and take a more informed decision. Hence, the probability of choosing a new and better technology decreases importantly because of the underlying incentives.

The second contribution of this paper is to characterize optimal contracts that increase the information in this economy. I solve for the optimal contracts that maximize total welfare and satisfy budget balance, when the precision of the signals and its content is not observed by the social planner. Contracts in this case will specify a wage schedule for agents that are willing to experiment and a different wage schedule for non-experimenters. The separation of the agents occurs endogenously since a fixed cost of experimenting prevents everybody to become an experimenter.

Although the second setup seems quite different than the one principal-one agent model, it is shown that the main tradeoffs are still the same ones. There is a moral hazard and an adverse selection problem as before since the planner must increase the amount of information gathered and encourage the agents to undertake the best project. However, there is also a new component associated with the free riding going in the opposite direction of the moral hazard effect. The planner has now some pressure to reward experimenters for low returns since otherwise experimenters will prefer to become non-experimenters, decreasing the amount of information in the economy and increasing the probability of being rewarded for low outcomes.

As long as a regularity condition is satisfied or the size of the population is large enough, it is shown that the last component is dominated by the moral hazard one. In this case, the optimal contract should split among experimenters the total output if the risky technology is chosen and its return is significantly better than the return of the previous technology. On the other hand, output should be split among nonexperimenters if the risky technology is not significantly better than the safe one.

The rationale is the same as in the case of one agent. Since agents can learn more accurately about the risky technology by having more people experimenting or by increasing the precision per experimenter, the planner also substitutes the extensive margin by a greater intensive margin to provide more powerful incentives. This constrained efficient outcome resembles the features exposed at the beginning where a few successful innovators will get paid a large amount of money, whereas not so successful entrepreneurs, and the rest of the society will not be paid as much. Several comparative statics are derived throughout the paper related to the optimal contract, the equilibrium, and the first best allocation.

The model applies to a variety of situations. The one principal one agent problem is tightly connected to the optimal compensation of CEOs who must be motivated by the shareholders to undertake risky projects that could potentially lead to higher returns. In this scenario, the effort exerted by the CEO in learning about the portfolio of projects and the learned information is usually never observed by the shareholders, only the project chosen and the realized returns are observed. In a managerial environment, the optimal contract can be implemented using restricted stocks conditional on a performance threshold. When the optimal contract is constrained to be monotone as in Innes (1990), the optimal monotone contract is an option with strike price greater than the return of the safe project.

The model also reflects the environment when an expert must be hired to give advice about a new technology that is available. The latter is the case of consultants or specialized researchers who have been previously prepared for these tasks. The principal who hires such agents usually does not have the expertise to understand the information and that is why he must hire them.

The team problem resembles the technology adoption decision faced by farmers in developing countries. Foster and Rosenzweig (1995) studied the adoption of HYVs in India and found that imperfect knowledge about the management of the new seeds was a significant barrier to adoption. Moreover, they found that farmers learn through their neighbors, but that they do not fully incorporate the social returns of their experimentation. Hence, the rate of adoption was much slower than the desired one. In this context the amount of trials performed by farmers as well as the information gathered by them are usually unobservable to the social planner. The optimal contract suggests that experimenting farmers should be subsidized when a new seed is adopted and ends up very successful, otherwise they should be heavily taxed.

The framework also resembles the problem faced by managers who must encourage innovation among her employees to increase the profits of the firm. Workers have to divide their time between undertaking known tasks or exploring new ideas. If a worker comes with a new innovation that improves the existing technology the firm will adopt it and every worker will benefit from the adoption. However, innovations are risky ventures with a high probability of failure, thus agents prefer to put more effort on known tasks which returns are well known, while expecting that other workers take the risky decision. Again issuing options and restricted stock to experimenters are shown to be useful to encourage more risk-taking.

As a last example, the proposed framework also fits the situation of industries with high levels of innovation such as pharmaceuticals. Pharmaceuticals must invest in potential drugs which effectiveness is unknown. Once its effectiveness is proven, other firms can replicate the drug as a generic without incurring in a significant cost. In the absence of property rights, free-riding reduces innovation and the potential discovery of new drugs. The investment chosen by firms in the earlier stage of adoption of new technologies is usually not disclosed. Also R&D expenditures and the processes of invention are typically hidden and can only be revealed after the product is in the market and competitors can imitate it. Given these constraints, the optimal contract suggests the use of patents only for breakthroughs and not marginal innovations.

1.1 Literature Review

The structure used in the one principal and one agent model was first studied by Lambert (1986) and Demski and Sappington (1987), who used a simplified environment with two or three possible outcomes. Similar models where later developed as in Feess and Walzl (2004), Chade and Kovrijnykh (2011) and Gromb and Martimort (2007). In contrast with these papers, I allow for a continuum of outcomes, which permits a more complete characterization of optimal contracts. The closest structure to my model is the one studied by Malcomson (2009) who focuses in optimal distortions of the final decision as a mechanism to encourage more information acquisition.

This paper, on the other hand, characterizes the optimal contract as the incentive mechanism.

The information structure used in this paper is similar to the one used in Szalay (2009) and Persico (2000). However, the first model is used in a procurement environment where the acquired information is induced to be completely revealed, this is not the case in this paper. In the second one the agents acquire information to learn about their value for an object, not the value for a principal as in our model. A similar information structure is also used in Moscarini and Smith (2001). Nevertheless, their model focus in the optimal actions of a single decision maker when information can be acquired over time, and not in the strategic interaction.

The acquisition of information is also related to bandit problems where an agent can learn about the return of a project by undertaking it as in Manso (2011) and Ederer (2008). This paper departs from this framework by enriching the information acquisition process and allowing agents to invest in the precision of their signals. Bonatti and Horner (2011) study moral hazard in teams over time where the return of a project is unknown and effort determines the rate of arrival of the return. Our setup is different in that individuals invest one time on a signal before deciding to undertake the risky project.

The next section introduces a principal agent setup with a single agent acquiring private information to give intuition about the optimal contracts on a simplified framework. The third section presents the model with multiple agents and characterizes the equilibrium, the first best allocation, and the optimal contract when the precision chosen by each agent is unobservable. In the next section I discuss how to implement the contract in different real-world applications. In the last section I conclude.

2 Principal-Agent Problem

Consider the case of a risk neutral principal who has to decide between a safe project with known returns and a risky project with unknown returns. The principal can hire a risk neutral agent to acquire information and recommend him which project to pursue.[1] The amount of information gathered by the agent is unobservable to the principal, as well as the realization of the signals acquired. Hence, this is a problem that involves hidden actions and hidden information, and the contracts designed by the principal can only be a function of the realized final outcome and the chosen project. It is assumed that both individuals have limited liability.

2.1 Model

There are two available projects that cannot be pursued simultaneously. There is a safe project with net return $y_s > 0$. There is also a risky one whose return $y_r \in [0, \overline{y}]$ is unknown, with $\overline{y} > y_s$. Let both individuals have the same nondegenerate prior belief $g(y_r)$ over the unknown return with finite mean μ_0.

The agent can generate information about the risky project by acquiring a continuum of e independent signals at a cost $C(e)$; this cost can be associated with R&D expenditures or the cost of running trials. Assume the cost function satisfies $C(0) = 0$, is increasing and is strictly convex in e. Moreover, assume there is a fixed cost c whenever the agent acquires information, that is only if $e > 0$.[2]

In other words, there is no cost if the there is no information acquisition; however, if the principal decides to experiment, agent must first incur in a fixed cost, for example setting up the lab. Alternatively, the fixed cost can be interpreted as if there is a minimum number of signals that must be purchased when information is acquired. The fixed cost is also observationally equivalent to the outside option of an agent, and thus can be interpreted as the expected minimum wage to be paid to the agent when the principal decides to hire him. In a context of bandit problems, where signals are the same returns of the risky project, the fixed cost can also be

[1]The individual can be in fact risk averse or risk lover, just let the returns perceived by the agent be measured in utils and let the agent maximize a Von Neuman-Morgenstern utility function

[2]This last assumption is used in this section to derive comparative statics related to when an agent must be hired, but is not crucial for the results obtained here. On the other hand, the assumption is very important for the subsequent section with multiple agents to obtain a well defined solution when the size of the population becomes very large.

thought as the ex ante expected return of the risky project, μ_0, and the individuals are initially pessimistic about it.

Each independent signal x_k is drawn from the distribution $f(x_k|y_r)$, for each $k \in [0, e]$. We will refer to e as the precision of the information and denote by x a sufficient statistic of the signals. Let the conditional pdf and cdf of x be denoted by $f(x|y_r, e)$ and $F(x|y_r, e)$, respectively, with support $[\underline{x}, \overline{x}]$.[3] Assume both functions are twice differentiable in e and x. Note, however, that the prior distribution of y_r is independent of e. Let the sufficient statistics be ordered, following Milgrom (1981), a signal x is more favorable than signal x' if the posterior distribution $g(y_r|x, e)$ first order stochastically dominates the posterior distribution $g(y_r|x', e)$.

It will be assumed that both the precision e and the sufficient statistic for the signals, x, are privately observed by the agent. The only observable variables for the principal are the chosen project and the final return of the project y_s or y_r. Thus the principal designs the optimal wage to be paid to the agent as a function of these variables, that is he chooses $w(y_s) = w_s$ and $w(y_r)$. It will also be assumed that individuals have a limited liability constraint. The wage for the agent cannot be lower than 0 and the principal cannot pay more than the return he receives. Formally, optimal wages must satisfy $0 \leq w(y) \leq y$. Given that a project j is chosen, the payoff for the principal is given by $y_j - w(y_j)$ and the payoff for the agent is $w(y_j) - C(e)$.

The game consists of two stages. In the first stage the principal designs a payment schedule and makes a take it or leave it offer to the agent. The agent accepts or rejects the contract. If she accepts the contract, she chooses to acquire e independent signals, which are privately observed by her. In the second stage, the agent updates her beliefs and chooses which project to pursue.[4] Finally a return y is realized and the principal

[3] Alternatively, the acquisition of information can be modelled as the purchase of a signal x with precision e defined in the Blackwell (1951) sense. That is experiment X is more precise than experiment X' if you can mimic signal X' by adding noise to signal X. Formally, experiment X is sufficient for (more precise than) experiment X' if for every $x' \in X'$ there is a probability distribution over X, $g(x; x')$ where $x \in X$, such that $\int g(x; x') f(x|y)\, dx = f(x'|y)$ for any y.

[4] Of course when $e = 0$ no information is acquired and the prior is not updated. In the Blackwell

pays to the agent the contracted wage $w(y)$.

2.2 First Best

Suppose the precision chosen by the agent and the sufficient statistic are observed by the principal, also assume there are no limited liability constraints. The principal faces the following problem:[5]

$$\max_{e,w(y_r),w(y_s)} \mathbb{E}_x \left[\max_{j_x \in \{s,r\}} \mathbb{E}_{y_{j_x}} \left[y_{j_x} - w\left(y_{j_x}\right) | x, e \right] \right]$$

$$\text{s.t } \mathbb{E}_x \left[\mathbb{E}_{y_{j_x}} \left[w\left(y_{j_x}\right) | x, e \right] \right] - C(e) \geq 0$$

where j_x is the project chosen by the principal when x is observed. The first best can be obtained by either a constant payment from the principal to the agent equal to the cost or by having the agent pay the principal for the returns of the project. Either of the alternatives lead us to solve the following problem:

$$\max_{e} \mathbb{E}_x \left[\max_{j_x \in \{s,r\}} \mathbb{E}_{y_{j_x}} \left[y_{j_x} | x, e \right] \right] - C(e)$$

Since there are two stages, we proceed to solve the individual's problem using backward induction. That is, I will first determine which project is going to be chosen given the information acquired. Second, I will determine the optimal wages that implement a certain precision. Third, the optimal precision is found given the principal decides to hire the agent. Finally, I will characterize when the principal decides to hire an agent as a function of y_s and c.

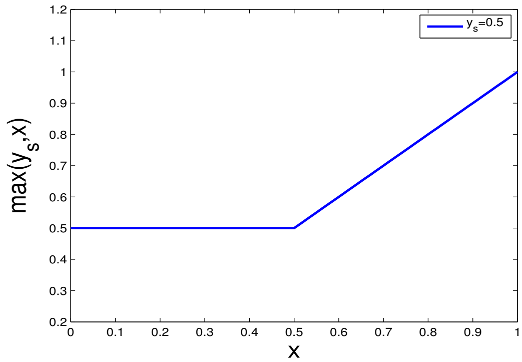
Without loss of generalization, let $x = \mathbb{E}_{y_r} \left[y_r | x, e \right]$ be the posterior mean of the

formulation this is equivalent to the case where x and y_r are independent.

[5] As noted before, we normalize the outside value for the agent to 0 since it is observationally equivalent to the fixed cost c

risky project.[6] The individual will choose the risky project if $x > y_s$, thus the payoff of the second period is given by $\max\{x; y_s\}$. Note this is a convex function of x.

Figure 1: **Utility in second period**



The value of experimentation is defined as the ex ante expectation of the utility in the second period, that is

$$U(e) = \mathbb{E}_x [\max\{x; y_s\}]$$

From the previous properties we can prove the following lemma:

Proposition 1. *The value of experimentation* $U(e)$ *is greater than* $\max\{\mu_0; y_s\}$*, and is differentiable, strictly increasing, and bounded in* e

The intuition of the proposition is as follows. The expected posterior mean $\mathbb{E}_x[x]$ is just the prior mean μ_0, which means that the learning process is a Martingale and assures that with enough signals or with a sufficiently large precision we would eventually learn. Therefore, as e approaches infinity, the sufficient statistic perfectly reveals the unknown return, i.e it is a known transformation of the unknown return. In other words, the limit conditional distribution $\lim_{e\to\infty} f(x|y_r, e)$ will be degenerate at y_r. Hence, the unconditional distribution of the posterior mean converges to the

[6] Since signals are ordered, the posterior mean will be a monotone transformation of the signal. Thus the distribution of the posterior mean will be a transformation of the distribution of the sufficient statistic. Hence, we can let $\underline{x} = 0$ and $\overline{x} = \overline{y}$

prior distribution. Moreover, the distribution $f(x|e)$, second order stochastically dominates the distribution $f(x|e')$, for all $e < e'$. Since the utility is convex in x, then the individual prefers a higher e.

If at time 0 the individual chooses to experiment, then she will choose precision e to maximize:

$$\max_{e} U(e) - C(e) \tag{1}$$

Unfortunately we cannot assure this is a concave problem since $U(e)$ may be convex for low values of precision (Moscarini and Smith, 2002). However, since the option value is bounded and the cost is strictly convex, we can assure the existence of a solution to the problem. The next proposition characterizes the solution.

Proposition 2. *A solution to problem (1) exists. If the solution is interior, the optimal precision e^* is characterized by*

$$C_e(e^*) = U_e(e^*) \tag{2}$$

Moreover, if $\int_0^y F(x|e)\,dx$ is strictly concave in e, then condition (2) is also sufficient and the maximum is unique.

The strict concavity condition implies that even though increasing the precision generates a mean preserving spread on the distribution, such spread becomes smaller as the precision becomes larger. Such condition was also suggested by Szalay (2009) to motivate the use of the first order approach in a procurement problem. The next lemma shows how the optimal precision depends on the prior belief.

Proposition 3. *Suppose the principal decides to hire an agent to acquire information. If $\int_0^y F(x|e)\,dx$ is concave in e then the optimal precision e^**

1. *Achieves a unique maximum when $y_s = \mu_0$*
2. *Is strictly increasing in y_s as long as $y_s < \mu_0$*
3. *Is strictly decreasing in y_s as long as $y_s > \mu_0$.*

At the beginning of first period a precision will choose to hire an agent if

$$U(e^*) - C(e^*) \geq \max\{y_s; \mu_0\}$$

Let the maximized objective function of the principal be denoted by $V(y_s, c) = \max\{U(e^*) - C(e^*); y_s; \mu_0\}$. The next proposition characterizes when the principal decides to hire an agent.

Proposition 4. *The function $V(y_s, c)$ is nondecreasing and convex in y_s and the principal decides to experiment, $e^* > 0$, when $c \leq \hat{c}$ and $y_s \in (a_c, b_c) \subseteq (0, \overline{y})$, where $\mu_0 \in (a_c, b_c)$. Moreover, such interval is decreasing in c, that is $(a_c, b_c) \subset (a_{c'}, b_{c'})$ for any $c < c' < \hat{c}$, with $(a_0, b_0) = (0, \overline{y})$ and $a_{\hat{c}} = \mu_0 = b_{\hat{c}}$.*

Even if beliefs are relatively pessimistic the individual decides to acquire information because of the potential gain represented by the value of experimentation. Indeed the precision chosen is increasing in the beliefs when they are pessimistic and achieves a maximum when ex ante the two projects have the same expected mean. When beliefs start to be optimistic, precision decreases, until a point where the agent does not have any more incentives to acquire information. The lower is the fixed cost c, the greater is the interval over which the principal decides to hire the agent. Furthermore, if there is no fixed cost, the principal will always decide to hire an agent to collect information.

2.3 Constrained Efficiency

Now suppose the principal does not observe the precision, nor the information gathered by the individual. Also assume individuals have limited liability as described before. In this context a fixed wage will not induce any effort from the agent. Therefore the principal must provide incentives to the agent by imposing more risk in her payoff, and by making sure he chooses the project that is more convenient to the principal. The first best is no longer obtained since the payoff for the principal is no longer convex and thus it will not require as much effort as before, thus there will be less acquired information in the second best. The constrained efficient problem for the principal can be stated as

$$\max_{e,w(y_r),w(y_s)} \mathbb{E}_x \left[\max_{j_x \in \{s,r\}} \mathbb{E}_{y_{j_x}} \left[y_{j_x} - w\left(y_{j_x}\right) | x, e \right] \right] \tag{3}$$

subject to

$$\mathbb{E}_x \left[\mathbb{E}_{y_{j_x}} \left[w\left(y_{j_x}\right) | x, e \right] \right] - C\left(e\right) \geq 0 \tag{4}$$

$$e \in \arg\max \mathbb{E}_x \left[\mathbb{E}_{y_{j_x}} \left[w\left(y_{j_x}\right) | x, e \right] \right] - C\left(e\right) \tag{5}$$

$$\mathbb{E}_{y_{j_x}} \left[w\left(y_{j_x}\right) | x, e \right] \geq \mathbb{E}_{y_{-j_x}} \left[w\left(y_{-j_x}\right) | x, e \right] \text{ for all } x \tag{6}$$

$$0 \leq w\left(y_j\right) \leq y_j \text{ for } j = r, s \tag{7}$$

Equation (4) is the same individual rationality constraint of before. Equation (5) is the incentive compatibility constraint that ensures the agent will choose the suggested precision. Equation (6) is another incentive compatibility constraint to make sure the agent chooses the project that is more convenient to the principal, where y_{-j_x} denotes the project the individual has not chosen. The last equation represents the limited liability constraint.

This problem is hard to solve because each incentive compatibility constraint involves a continuum of restrictions. However, the second incentive constraint can be reduced to only one constraint when optimal wages for the risky project are monotone nondecreasing. Since signals are ordered, a posterior implied by a signal first order stochastically dominates any posterior generated by any less favorable signal. Therefore a less favorable signal implies that the expected wage is lower. Since distributions are continuous in x, there must exists a cutoff x_c such that the expected wage when the risky project is chosen given such signal is equal to the wage when the safe project is chosen. Any more (less) favorable signal than x_c implies the agent will choose the risky (safe) project and that the constraint will not be binding.

The first incentive constraint can also be reduced to one equation using the first order approach, as it is traditional in the contract theory literature. That is, provided some conditions that assure the concavity of the agents payoff with respect to the precision, we can replace the constraint (5) for its first order condition. Let us for now assume that wages are nondecresing, and in the next proposition I will prove that is the case. Following Rogerson (1985), the doubly relaxed program is given by:

$$\max_{e,x_c,w(y_r),w(y_s)} \int_{x_c}^{\overline{x}} \int_{\underline{y}}^{\overline{y}} (y_r - w(y_r)) f(x, y_r|e)\, dy_r dx + (y_s - w(y_s)) F(x_c|e) \tag{8}$$

subject to

$$\int_{x_c}^{\overline{x}} \int_{\underline{y}}^{\overline{y}} w(y_r) f(x, y_r|e)\, dy_r dx + w(y_s) F(x_c|e) - C(e) \geq 0 \tag{9}$$

$$\int_{x_c}^{\overline{x}} \int_{\underline{y}}^{\overline{y}} w(y_r) f_e(x, y_r|e)\, dy_r dx + w(y_s) F_e(x_c|e) - C_e(e) \geq 0 \tag{10}$$

$$\int_{\underline{y}}^{\overline{y}} w(y_r) f(y_r|x_c, e)\, dy_r = w(y_s) \tag{11}$$

$$0 \leq w(y_j) \leq y_j, \quad \text{for } j = r, s \tag{12}$$

Equation (10) is the incentive compatibility constraint using the first order approach. The inequality assures the multiplier associated with it is positive when the constraint is binding. Equation (11) is the constraint that makes sure the agent takes the desired decision of the principal. Let λ, δ, and ϕ be the Lagrange multipliers for the first three constraints. Since the problem is linear on the wages, the optimal wages are determined by a bang-bang solution that is bounded by the limited liability constraint. After rearranging the derivative with respect to wages we obtain

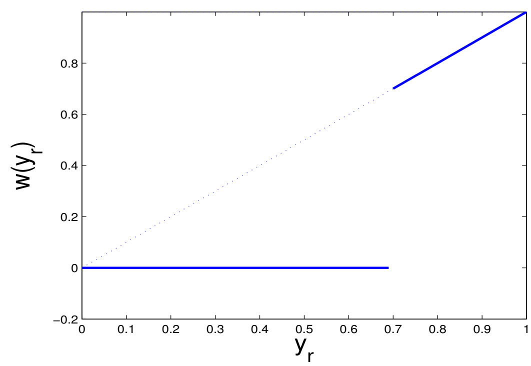
$$-1 + \lambda - \delta \frac{F_e(x_c|y_r, e)}{1 - F(x_c|y_r, e)} - \phi \frac{f(x_c|y_r, e)}{(1 - F(x_c|y_r, e)) f(x_c|e)} \tag{13}$$

Whenever this condition is positive, the wage will be set to the upper bound; if it is negative, then the optimal wage is zero. Using the structure of the signals and the latter equation we can indeed prove the that wages are nondecreasing in the following proposition.

Proposition 5. *The optimal wage schedule $w(y_r)$ is monotone and is characterized by a cutoff z such that*

$$w(y_r) = \begin{cases} y_r & \text{if } y_r \geq z, \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

Figure 2: **Optimal Contract**



The cutoff z is given by the value of y_r such that condition (13) is equal to zero. Wages are thus monotone if for any greater (lower) y_r the derivative is positive (negative). Thus showing that the derivative satisfies the single crossing condition implies that wages are monotone. The intuition for the proof has two components that work in the same direction.

As it is common in moral hazard problems, the first component is a likelihood ratio $\frac{F_e(x_c|y_r,e)}{1-F(x_c|y_r,e)}$. However, this ratio is particular to this problem since it describes the probability of undertaking the project when the individual increases the precision of the signal. Given the properties of the distribution, the numerator is zero in the interior only if $x_c = y_r$ and positive (negative) if y_r is greater (smaller) than x_c. Hence the ratio satisfies the single crossing property.

The second component is the hazard ratio $\frac{f(x_c|y_r,e)}{1-F(x|y_r,e)}$ which is characteristic of hidden information problems. Given the MLRP condition of the signals with respect to the return, this hazard ratio is monotone decreasing with respect to the return (see Athey (2002)). This in turn implies that the conditional truncated distribution of the signals first order stochastically dominates truncated distributions conditional on lower returns. Thus the observation of a higher return suggests it was more likely the agent observed a signal greater than the cutoff.

Recall that higher effort induces a second order stochastically dominated unconditional distribution, that is a distribution with more weight in the tails. Therefore, by rewarding the agent only for high returns, the principal is encouraging her to exert a greater effort. On the other hand, higher returns increase the probability of the realization of more favorable signals. Thus the optimal wage schedule creates the incentive for the agent to adopt the risky technology after observing favorable signals.

Given that wages were proved to be monotone nondecreasing, we are ready to show that the first order approach is valid. In other to do that, we first need to prove that the agent's utility is concave in effort, and thus the first order condition yields the effort that maximizes such utility. Then, we need to show that constraint (10) is indeed binding.

Proposition 6. *Let* (e^*, x_c^*, w^*) *be the solution to the doubly relaxed problem, if* $F(x|y,e)$ *is convex in* e *then* (e^*, x_c^*, w^*) *is also a solution for the constrained efficient problem.*

This condition over the conditional cdf of the sufficient statistic is a particular case of the condition found by Sinclair-Desgagné (1994) in the context of contracts with multisignals. Although our contract only depends directly on the return, it indirectly depends on the sufficient statistic since the risky project will only be chosen if the unobserved signal is greater than a cutoff. Thus his conditions apply to our case. Jointly with the condition stated in Proposition 5, it requires that such probability is decreasing in the number of signals but at an increasing rate.

2.4 Optimal Monotone Contract

The optimal contract found in the previous subsection is not continuous. In particular, the payment for the principal is not monotone since any return greater than the threshold will yield him zero profit. As argued by Innes (1990), this type of contracts could be manipulated by either the principal or the agent if any of them can affect the return before the contract is paid. For example, the principal would have incentive to sabotage the risky project by burning profits in excess of the threshold. Similarly, the agent would have an incentive to inflate profits by borrowing money and "revealing" a higher apparent profit to the principal.

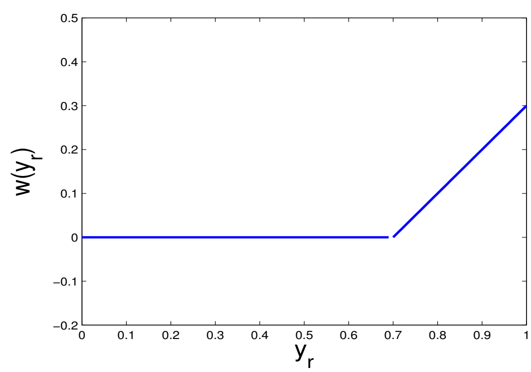
In order to prevent this behavior, a monotonicity constraint must be imposed, thus modifying the limited liability constraint:

$$w\left(y_r - \epsilon\right) \leq w\left(y_r\right) \leq w\left(y - \epsilon\right) + \epsilon$$

The same argument in the preceding section applies, and the optimal monotone contract will be option like, with an strike price z_0 greater than the safe return:

$$w\left(y_r\right) = \max\left\{0, y_r - z_0\right\} \tag{15}$$

Figure 3: **Optimal Monotone Contract**



3 Team Problem with Multiple Agents

Now suppose there is a population of N risk neutral agents who must choose between a risky project and a safe project. Let the agents appropriate the whole return of the project they choose. All agents share the same prior beliefs and can acquire information about the risky project by exerting costly effort. Assume the information gathered becomes public and other agents can use it to update beliefs. Since information is now a public good there will be free-riding in equilibrium and there will be less information than in the first best. The purpose of this section is to derive the optimal contract designed by a social planner who does not observed the precision chosen by each agent, nor the information gathered by the agents. Assume also that contracts must satisfy budget balance and that agents have limited liability.

3.1 Model

The utility functions for each agent and the returns for each project are given as before. Of particular importance will be the fixed cost incur by experimenters since it will generate a natural partition of the population between experimenters and non-experimenters. The fixed cost will also assure the existence of an equilibrium when the size of the population goes to infinity.

Assume that the sufficient statistic obtained from the information acquisition of all agents is publicly observed by the agents but not observed (or understood) by the principal. In this scenario agents decide whether to experiment and obtain costly extra signals or use the available signals from others. Thus, the outside option is no longer zero.

I consider again the case of two stages. In the first stage the individuals decide simultaneously whether to acquire information or not. If an agent i decides to experiment, she must also choose how many signals to acquire, e_i at a cost $C\left(e_i\right)$. Assume signals acquired by different individuals are independent and identically distributed with pdf $f\left(x_k|y_r\right)$. At the end of the period each individual i observes the overall sufficient statistic x with pdf given by $f\left(x|y_r, \sum_{j=1}^{N} e_i\right)$. At the beginning of the second stage each agent updates her beliefs and decides which project to pursue. Since all signals are public, in equilibrium everybody will take the same decision in the final period.

3.2 Equilibrium

Suppose first that each agent appropriates the return from the chosen project and thus the payoff for agent i is given by $y - C(e_i)$. In order to simplify the analysis I will focus on equilibria where experimenters choose the same level of precision. Let the number of experimenters be denoted by n. It is important to note that not every agent will be necessarily an experimenter in equilibrium, a nonexperimenter will set $e_i = 0$. Let e_{-i} be the sum of precision chosen by all individuals except i. Therefore each agent is willing to solve the following problem:

$$\max_{e_i} E_x \left[\max_{j_x \in \{s,r\}} \mathbb{E}_{y_{j_x}} \left[y_{j_x} | x, e_i + e_{-i} \right] \right] - C(e_i)$$

Following the analysis of the previous section, we will define an option value of experimentation for an agent i that this time will depend on the overall precision chosen by all the experimenters:

$$U(e_i + e_{-i}) = \mathbb{E}_x \left[\max_{j_x \in \{s,r\}} \mathbb{E}_{y_{j_x}} \left[y_{j_x} | x, e_i + e_{-i} \right] \right]$$

Therefore, the ex-ante utility for a non-experimenter is given by $U(e_{-i})$. On the other hand, the ex-ante utility for experimenters is given by $\max_{e_i} \{U(e_i + e_{-i}) - C(e_i)\}$. Assume first that agents also observe the overall precision, hence the appropriate concept for equilibrium is a Subgame Perfect Nash equilibrium (SPE).[7] The number of equilibria depends on the properties of the cost function and the distribution of the

[7] Alternatively we can assume that the overall precision is unobservable, then the appropriate concept is a Perfect Bayesian Nash equilibrium (PBE). However, there exists a PBE with the same payoffs and actions on the equilibrium path as the ones in the SPE we are interested in. Thus we use this equilibrium concept for the sake of simplicity.

posterior expected mean. Assume for the rest of the paper that $\int_0^y F(x|e_i + e_{-i})\,dx$ is strictly concave to obtain a unique solution (recall Proposition (2)).

Definition 7. *A experimenter-symmetric SPE is defined by the number of people experimenting, n^*, and the symmetric precision of the signals e^*, such that nobody has incentives to deviate, that is*

- *Non-experimenters do not want to deviate:*

$$U(n^*e^*) \geq U(e_i + n^*e^*) - C(e_i) \;\; \textit{for any } e_i$$

- *Experimenters do not want to deviate:*

$$U(n^*e^*) - C(e^*) \geq U((n^* - 1)e^*)$$

and

$$e^* = \arg\max_{e_i} \{U(e_i + (n^* - 1)e^*) - C(e_i)\}$$

In this experimenter-symmetric equilibrium the overall precision will be n^*e^*. These two variables will play the same role in the learning process since an increase in either one will have the same effect (in terms of elasticities) on the value of experimentation. In fact, this function will have the same properties with respect the overall precision ne as the ones described in the previous section.

As before, we will use backward induction to solve for the agent's behavior. In a symmetric equilibrium where n people experiment in the first period, an experimenter will choose the optimal precision e^* such that

$$C_e(e^*) = U_{e_i}(e_i + (n^* - 1)e^*)$$

The optimal precision inherits the same properties as the ones found in the previous section. However, we now have an interesting relationship between precision, overall precision, and the number of experimenters as summarized in the following lemma.

Lemma 8. *The optimal precision e^* is decreasing in n, but the overall precision ne^* is increasing in n.*

The intuition behind this lemma is that although n and e have the same relative effect on the value of experimentation, the convexity of the cost induces an imperfect complementarity between the two components of the overall precision. If the cost of the precision were to be linear, the overall precision would remain unchanged since an increase in the number of signals will be exactly offset by the decrease in the precision. The convexity of the cost function implies that effort becomes less responsive to changes in n.

The monotonic response of the overall precision to n implies that $U(n^*e^*) > U((n^*-1)e^*)$ since the value of experimentation is monotone increasing in the overall precision. However, these increments will become smaller after some threshold because of the concavity of such function. Therefore, the equilibrium number of experimenters n^* is given by the equation[8]

$$U(n^*e^*) - C(e^*) \geq U((n^*-1)e^*)$$

Lemma 9. *In a experimenter-symmetric SPE, every agent will experiment if $N < \hat{N}$, and the number of experimenters will be independent of N as long as $N > \hat{N}$*

The concavity and boundedness of the option value of experimentation jointly with the fixed cost of experimentation implies that the number of experimenters will be finite in equilibrium even as the size of the population goes to infinity. On the other hand, the optimal precision is set such that its marginal cost is greater than the average cost. However, as the total amount of signals acquired increases (the total number of experimenters increase), the optimal precision converges to the minimum

[8] In differentiable terms this condition is equivalent to $C(e^*) \geq U_n(n^*e^*)$, with equality if $n > 0$

efficient scale, which ensures its finiteness. Note the similarity of the problem solved by a firm facing perfect competition with fixed costs.

3.3 First Best

Suppose now that there exists a social planner who wants to maximize the ex-ante total welfare of this economy. Given that n agents are experimenting, ex-ante total welfare is then defined as:

$$W = NU\left(\sum_{i=1}^{n} e_i\right) - \sum_{i=1}^{n} C\left(e_i\right)$$

Note that in the aggregate, the number of experimenters are associated with a linear cost, whereas the precision of the signals have a convex one. Note also that the social planner must consider the externality generated by the signals by multiplying the value of experimentation by the number of agents in the economy. Given a number of experimenter n, the social planner chooses e^{FB} such that

$$C_e\left(e^{FB}\right) = NU_e\left(ne^{FB}\right)$$

The social planner will also increase the number of experimenters n^{FB} as long as[9]

$$C\left(e^{FB}\right) \leq N\left[U\left(n^{FB}e^{FB}\right) - U\left(\left(n^{FB}-1\right)e^{FB}\right)\right]$$

Lemma 10. *In the first best, the number of experimenters goes to infinity as N goes to infinity, but its proportion $\frac{n^{FB}}{N}$ goes to zero. The first best precision remains finite and converges to the minimum efficient scale.*

Note again how N increases the marginal value of experimentation, this implies

[9]In differentiable terms this condition is equivalent to $C\left(e^{FB}\right) \geq NU_n\left(n^{FB}e^{FB}\right)$, with equality if $n^{FB} > 0$

that in the first best there is more experimentation n than in equilibrium. Moreover, the number of experimenters grows without bound as the size of the population goes to infinity. However, n does not increase as fast as N because of the concavity and boundedness of $U(\cdot)$. Note also that the central limit theorem implies that the speed of convergence of the learning process is $\sqrt{ne}$ and thus is not optimal to increase n as fast as N.

On the other hand, while an increase in N has a direct positive effect on e pushing it to infinity, there is also an indirect effect of N coming through n which is growing bigger and decreases e. It turns out that these effects approximately offset and the first best precision converges to the point where average cost is minimized, which is close to the same precision chosen in equilibrium.[10]

Because of the greater marginal benefit and the linearity of the costs associated to the number of experimenters, a social planner decides to choose a greater n than the one obtained in equilibrium. On the other hand, because of the convexity of the cost associated to the precision, the optimal precision remains finite and close to the equilibrium one. In other words, the social planner decides to increase n and maintain fixed e since their relative effect on the option value of experimentation is the same but is more costly to increase e.

The first best can be implemented under budget balance if the individual precision is observed. For example, by distributing the surplus from adopting the new technology among the experimenters when they choose the first best precision, but not when they deviate, the first best is implemented. To see this note that if experimenters follow the suggested first best precision their payoff is given by

$$\frac{N}{n^{FB}} U\left(n^{FB} e^{FB}\right) - C\left(e^{FB}\right)$$

whereas the payoff for nonexperimenters or experimenters that deviate from the first best precision is 0. From the conditions obtained for the first best we know that

[10] The discreteness of n is what prevents them to be equal, but as N grows large the choice of n resembles the case of a continuum number of experimenters.

the payoff for obedient experimenters will be greater than 0 and thus they will not want to deviate. Note also that because the suggested wages are nondecreasing the agents will be willing to adopt the best technology.

3.4 Constrained Efficiency

Suppose there exists a social planner that wants to maximize ex-ante total welfare as before. Assume that the principal does not observe the precision chosen by experimenters, nor the aggregate statistic. Although the first assumption is natural, the second is not given that agents can observe the statistic. However, one can interpret this assumption as if agents are experts who can interpret the signals and relate them to the returns, whereas the principal lacks such degree of expertise. This is observed in scenarios such as CEOs hiring chemists to develop a new product, and that is the main reason why principals must hire an agent for this purpose.

In addition assume both agents have limited liability as before and that there must be budget balance. Under this restriction, the wages suggested in the last subsection do not implement the first best since agents would have incentives to choose a lower precision because they do not internalize the social gains since they are split among all the experimenters. Assume that the social planner only observes the final output and designs a wage schedule for experimenters and nonexperimenters as a function of the observed returns. Let $w(y)$ and $v(y)$ be the wages for an experimenter and a nonexperimenter when the observed return is y.

The timing is as follows. First the social planner offers the menu of contracts. Then each agent chooses a contract. Agents who decided to experiment then choose the precision of their signals. At the end of the stage every agent observes the aggregate sufficient statistic. Then agents update beliefs given the reported information on how many people experimented, their precision, and the observed signals. Then agents take a definitive decision over which technology to use. After the final output is realized the social planner pays the promised contract.

The planner solves the following problem:

$$\max_{n,\{e_i\}_{i=1}^{n},w(y_r),w(y_s),v(y_r),v(y_s)} NU\left(\sum_{i=1}^{n} e_i\right) - \sum_{i=1}^{n} C\left(e_i\right) \tag{16}$$

subject to

$$\mathbb{E}_x\left[\mathbb{E}_{y_{jx}}\left[w\left(y_{jx}\right)|e_i + e_{-i}\right]\right] - C\left(e_i\right) \geq \mathbb{E}_x\left[\mathbb{E}_{y_{jx}}\left[v\left(y_{jx}\right)|e_{-i}\right]\right] \tag{17}$$

$$\mathbb{E}_x\left[\mathbb{E}_{y_{jx}}\left[v\left(y_{jx}\right)|e_{-i}\right]\right] \geq \mathbb{E}_x\left[\mathbb{E}_{y_{jx}}\left[w\left(y_{jx}\right)|e + e_{-i}\right]\right] - C\left(e\right) \text{ for any } e \tag{18}$$

$$e \in \arg\max \mathbb{E}_x\left[\mathbb{E}_{y_{jx}}\left[w\left(y_{jx}\right)|e + e_{-i}\right]\right] - C\left(e\right) \tag{19}$$

$$\mathbb{E}_{y_{jx}}\left[w\left(y_{jx}\right)|x, \sum_{i=1}^{n} e_i\right] \geq \mathbb{E}_{y_{-jx}}\left[w\left(y_{-jx}\right)|x, \sum_{i=1}^{n} e_i\right] \text{ for all } x \tag{20}$$

$$\mathbb{E}_{y_{jx}}\left[v\left(y_{jx}\right)|x, \sum_{i=1}^{n} e_i\right] \geq \mathbb{E}_{y_{-jx}}\left[v\left(y_{-jx}\right)|x, \sum_{i=1}^{n} e_i\right] \text{ for all } x \tag{21}$$

$$w\left(y_j\right), v\left(y_j\right) \geq 0 \text{ for } j = r, s \tag{22}$$

$$nw\left(y_j\right) + \left(N - n\right) v\left(y_j\right) \leq N y_j \text{ for } j = r, s \tag{23}$$

Constraint (17) is the individual rationality constraint for experimenters that prevents them from not exerting any effort and free ride the public information. Constraint (18) is the individual rationality constraint for the nonexperimenter and it states that the individual is better off by not investing in information rather than acquiring a costly extra signal. The incentive compatibility constraint (19) assures the individual is willing to invest in the prescribed precision.

Lemma 11. *If (17) and (19) are satisfied, then (18) is satisfied.*

Constraints (20) and (21) are similar to constraint (6) and assures both experimenters and nonexperimenters will choose the project chosen by the principal. Finally, constraints (22) and (23) are the limited liability and budget balance constraints. The latter constraint will be binding since we are maximizing total welfare, thus we can substitute the wage for a nonexperimenter by $v\left(y_j\right) = \frac{Ny_j - nw(y_j)}{N-n}$. Using this expression and constraints (20) and (21), all agents will choose the risky project if:

$$0 \leq \mathbb{E}_{y_r}\left[w\left(y_r\right)|x, \sum_{i=1}^{n} e_i\right] - w\left(y_s\right) \leq \frac{N}{n}\left(\mathbb{E}_{y_r}\left[y_r|x, \sum_{i=1}^{n} e_i\right] - y_s\right)$$

Therefore, assuming wages are monotone nondecreasing, there is no distortion in the chosen optimal project and the constraint will only bind at an x_c such that $\mathbb{E}_{y_r}\left[y_r|x_c, e\right] = y_s$. This means the cutoff is such that the posterior mean of the risky project is equal to the return of the safe project. Using (23) into the remaining constraints and substituting (19) by the first order approach,[11] the principal's problem is simplified to:

$$\max_{n,\{e_i\}_{i=1}^{n},w(y_r),w(y_s)} NU\left(\sum_{i=1}^{n} e_i\right) - \sum_{i=1}^{n} C\left(e_i\right) \tag{24}$$

subject to

$$\begin{aligned}\int_{x_c}^{\overline{x}}\int_{\underline{y}}^{\overline{y}} w\left(y_r\right) f\left(x, y_r|e_i + e_{-i}\right) dy_r dx + w\left(y_s\right) F\left(x_c|e_i + e_{-i}\right) - C\left(e\right) \quad \geq \\ \int_{x_c}^{\overline{x}}\int_{\underline{y}}^{\overline{y}} \frac{Ny_r - nw\left(y_r\right)}{N-n} f\left(x, y_r|e_{-i}\right) dy_r dx + \frac{Ny_s - nw\left(y_s\right)}{N-n} F\left(x_c|e_{-i}\right)\end{aligned} \tag{25}$$

$$\int_{x_c}^{\overline{x}}\int_{\underline{y}}^{\overline{y}} w\left(y_r\right) f_e\left(x, y_r|e_i + e_{-i}\right) dy_r dx + w\left(y_s\right) F_e\left(x_c|e_i + e_{-i}\right) - C_e\left(e_i\right) \geq 0 \tag{26}$$

$$\mathbb{E}_{y_r}\left[w\left(y_r\right)|x_c, \sum_{i=1}^{n} e_i\right] = w\left(y_s\right) \tag{27}$$

$$0 \leq w\left(y_j\right) \leq \frac{N}{n} y_j \text{ for } j = s, r \tag{28}$$

Let λ, δ, and ϕ be the Lagrange multipliers for the first three constraints. The

[11] The first order approach is again valid if $F\left(x|y_r, e\right)$ is convex in e and wages are monotone nondecresing, see Proposition 6

problem is again linear on the wages. Rearranging the derivative with respect to wages we obtain

$$
\begin{gathered}
1+\left(\frac{n-1}{N-n+1}\right)\left(\frac{1-F\left(x_c|y_r,e_{-i}\right)}{1-F\left(x_c|y_r,e_i+e_{-i}\right)}\right) \\
-\frac{\delta}{\lambda}\left(\frac{F_{e_i}\left(x_c|y_r,e_i+e_{-i}\right)}{1-F\left(x_c|y_r,e_i+e_{-i}\right)}\right)-\frac{\phi}{\lambda}\left(\frac{f\left(x_c|y_r,e_i+e_{-i}\right)}{1-F\left(x_c|y_r,e_i+e_{-i}\right)}\right)
\end{gathered}
\tag{29}
$$

As mentioned in the previous section, the last two components of the equation are increasing in y_r and represent a monotone likelihood ratio and a monotone hazard ratio. But this time there is another component that is decreasing in y_r. The ratio $\frac{1-F(x_c|y_r,e_{-i})}{1-F(x_c|y_r,e_i+e_{-i})}$ is the same likelihood ratio but expressed in discrete terms, and is decreasing given that the numerator has a lower precision. The rationale for having an element decreasing in the returns is that by rewarding the experimenter for extreme outcomes, the experimenter is having incentives to become a nonexperimenter by reducing the overall precision and increasing the probability of being rewarded. The next proposition gives plausible sufficient conditions to obtain monotone wages by

Lemma 12. *If $1-F\left(x|y_r,e\right)$ is at least as logconcave as $1-F\left(x|e\right)$ in e, or $N\geq\hat{N}$, then optimal wages $w\left(y_r\right)$ are characterized by a cutoff z such that*

$$
w\left(y_r\right)=\begin{cases}\frac{N}{n}y_r & \text{if } y_r\geq z, \\ 0 & \text{otherwise.}\end{cases}
\tag{30}
$$

Given such conditions, the optimal contract suggests that the experimenter is encouraged to increase the variance of the posterior mean by increasing her payoff in realizations that are much better than the returns from the safe project. On the contrary, the total output will be split among non-experimenters when the risky return is not significantly better than the safe return. The intuition for having the logconcavity condition is that if the rate at which the conditional survival function decreases is lower than the one of the unconditional distribution, then the effect of the discrete rate is always smaller than the effect of the marginal rate by concavity. On the other hand, as the population grows, the relative effect decreases since the ratio of experimenters decreases and thus their relative reward is bigger compared to the one for nonexperimenters. Therefore the incentives to become nonexperimenters disappear.

4 Implementation

The analyzed problem can be understood from the perspective of optimal taxation when individuals generate information externalities. The contract suggests that the planner should subsidize experimenters when new projects are adopted and generate sufficiently high returns, otherwise they should be heavily taxed.

In the context of technology adoption, the existence of farmer cooperatives can be used to implement this contract. For example, National Cereal Boards play an important role on the coordination of farmers within a country. In Kenya this organism has played the role of creditors, promoters of research and market development, and regulators (Raikes, 1994). Although in recent decades their policies have been oriented towards a free market, the board used to lobby in the government for the determination of prices. An institution like this one could potentially reward farmers differently according to their willingness to try new technologies.

Another example of such organizations is the National Federation of Coffee Growers of Colombia.[12] The federation groups more than 500 thousand colombian families that produce coffee. Its mission is to represent the interests of coffee growers, create social programs to improve the quality of life of the producers, investment in research and knowledge transfer as well as in promotion and advertising, and the commercialization of coffee.

Within this last objective is what is considered by them their most significant service called the Purchase Guarantee Policy. This policy involves the setting of a

[12] http://www.federaciondecafeteros.org/particulares/en/

minimum price at which the coffee should be sold. If no buyers are willing to acquire the product at this price, they commit to purchasing it. This price is public and constitutes a reference point for the market. Hence, the federation's role could be used to differentiate the price paid to experimenting growers who increase the speed of adoption of new technologies given this is in its interests.

The proposed contract can also be implemented within firms. The level of innovation or the adoption of better technologies or practices are reflected in the value of the firm. Therefore stock options and profit sharing strategies posses similar properties as our contract.

Stock option programs give workers the right to buy company's shares at a fixed price for a given period of time. These will be only exercised if the market price is higher than the strike price originally agreed to. Usually stock options are used as a long- term motivator and the employee is constrained on exercising the option after certain time. Similarly, the firm could constrain the option exercise until the market price crosses a threshold, and thus implementing the proposed contract.

Likewise, profit sharing is also used as a long-term motivator where individuals are entitled to a percentage of the profits of a firm after a given period. To implement the contract the firm could set a threshold on the profits such that the workers can only claim her share if profits are greater than such level.

The optimal contract might be also interpreted as a patent policy to encourage innovation. It suggests that patents should only be given if it is shown that the new technology is significantly better than the previous one, and not for marginal improvements. However, this result cannot be interpreted as a restriction on the use of new technology as often happens with patents. In other words, the optimal contract does not allocate the property rights of the new technology. On the contrary, it encourages the adoption of the new technology by all the population, while rewarding innovators with the surplus they generated, suggesting an optimal pricing policy.

5 Conclusions

I first studied the problem faced by a principal who must hire an agent that can acquire information about a new project and must also take the decision of whether to pursue it or not. I show how the optimal contract resembles restricted stock. When the optimal contract is constrained to be continuous then it resembles an option with a strike price greater than the return of the safe project.

I also analyzed the problem faced by innovators who can acquire costly information before choosing between a known project and an unknown one, but such information is also observed but others for free. In this context, information is a public good and thus a free-riding problem arises. In equilibrium there will be less experimenters than in the first best since agents do not internalize the social benefits of experimentation.

The first best level of experimentation can be implemented when the number of signals acquired by the agents is observed; however, this is not necessarily the case when such investment is not observed. I derive the optimal contract when the amount or precision of the revealed information is unobserved and experimentation cannot be enforced. The optimal contract suggest that experimenters should be given the whole surplus if new technology is significantly better that the previous one. The intuition for this result is that experimenters must increase the number of signals acquired to increase the probability of being rewarded with the surplus.

The conclusions of the model are robust when there is a finite set of possible projects. However, the same framework does not apply when the decision belongs to a continuum as in Malcomson (2009) since the principal will not be able to offer a contract contingent on the chosen project. An interesting extension of the proposed model is to assume that individuals in the society only observe whether the firms decide to undertake the project or not. The observability of such an action is a noisy signal of the acquired information by each agent. In this context herding behavior may arise and the provision of incentives potentially differs to the one proposed here.

References

S. Athey. Monotone comparative statics under uncertainty. *The Quarterly Journal of Economics*, 117(1):187, 2002.

D. Blackwell. Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. July 31-August 12, 1950. Statistical Laboratory of the University of California, Berkeley.*, volume 1, pages 93–102, Berkeley, CA, 1951. University of California Press.

A. Bonatti and J. Horner. Collaborating. *The American Economic Review*, 101(2): 632–663, 2011.

H. Chade and N. Kovrijnykh. Information acquisition, moral hazard, and rewarding for bad news. 2011.

J.S. Demski and D.E.M. Sappington. Delegated expertise. *Journal of Accounting Research*, pages 68–89, 1987.

J.L. Doob. *Stochastic processes*. Wiley publications in statistics. Wiley, 1953. ISBN 9780471218135. URL `http://books.google.com/books?id=KvJQAAAAMAAJ`.

B. Ebersberger, O. Marsili, T. Reichstein, and A. Salter. Fortune favours the brave:: The distribution of innovative returns in finland, the netherlands and the uk. *Structural Change and Economic Dynamics*, 19(4):357–362, 2008.

F. Ederer. *Launching a Thousand Ships: Incentives for Parallel Innovation*. Working Paper, 2008.

E. Feess and M. Walzl. Delegated expertise–when are good projects bad news? *Economics Letters*, 82(1):77–82, 2004.

A.D. Foster and M.R. Rosenzweig. Learning by doing and learning from others: Human capital and technical change in agriculture. *The Journal of Political Economy*, 103(6):1176–1209, 1995.

D. Gromb and D. Martimort. Collusion and the organization of delegated expertise. *Journal of Economic Theory*, 137(1):271–299, 2007.

R.D. Innes. Limited liability and incentive contracting with ex-ante action choices* 1. *Journal of economic theory*, 52(1):45–67, 1990.

R.A. Lambert. Executive effort and selection of risky projects. *The Rand Journal of Economics*, pages 77–88, 1986.

J.M. Malcomson. Principal and expert agent. *The BE Journal of Theoretical Economics*, 9(1):17, 2009.

G. Manso. Motivating innovation. *The Journal of Finance*, 66(5):1823–1860, 2011.

O. Marsili and A. Salter. ÔinequalityÕof innovation: skewed distributions and the returns to innovation in dutch manufacturing. *Economics of Innovation and New Technology*, 14(1-2):83–102, 2005.

P.R. Milgrom. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, pages 380–391, 1981.

G. Moscarini and L. Smith. The optimal level of experimentation. *Econometrica*, 69 (6):1629–1644, 2001.

G. Moscarini and L. Smith. The law of large demand for information. *Econometrica*, pages 2351–2366, 2002. ISSN 0012-9682.

N. Persico. Information acquisition in auctions. *Econometrica*, 68(1):135–148, 2000.

P. Raikes. Business as usual: Some veal Ôfood markets in kenya. *Sociologia Ruralis*, 34(1):26–44, 1994.

W.P. Rogerson. The first-order approach to principal-agent problems. *Econometrica: Journal of the Econometric Society*, pages 1357–1367, 1985.

M. Rothschild and J.E. Stiglitz. Increasing risk: I. a definition. *Journal of Economic theory*, 2(3):225–243, 1970.

B. Sinclair-Desgagné. The first-order approach to multi-signal principal-agent problems. *Econometrica: Journal of the Econometric Society*, pages 459–465, 1994.

D. Szalay. Contracts with endogenous information. *Games and Economic Behavior*, 65(2):586–625, 2009.

A Appendix

Proof for Proposition 1. First, to show that $U(e) > \max\{\mu_0, y_s\}$ integrate by parts the value of experimentation to obtain

$$\begin{aligned} U(e) &= y_s + \int_{y_s}^{\overline{y}} (1 - F(x|e))\, dx \\ &= y_s + \int_0^{\overline{y}} (1 - F(x|e))\, dx - \int_0^{y_s} (1 - F(x|e))\, dx \\ &= \mu_0 + \int_0^{y_s} F(x|e)\, dx \end{aligned} \tag{31}$$

Thus the value of experimentation is greater than $\max\{\mu_0, y_s\}$. The value of experimentation is also differentiable because the conditional distribution $f(x|y_r, e)$ is assumed differentiable. To prove that is strictly increasing in the precision, I will prove that the distribution $f(x|e)$ second order stochastically dominates the distribution $f(x|e')$ whenever $e < e'$. Let x, x' and x'' be the sufficient statistics for the first e signals, the e' signals, and the additional $e' - e$ signals, respectively. Note that by independence $f(x'|e') = f(x|e)\, f(x''|e' - e)$. Define the corresponding conditional means as:

$$\mu = \int_0^{\overline{y}} y f(y|x, e)\, dy$$

$$\mu' = \int_0^{\overline{y}} y f(y|x', e')\, dy$$

First note that by the law of iterated expectations

$$\mathbb{E}_x\left[x\right] = \mu_0 = \mathbb{E}_{x'}\left[x'\right]$$

Therefore the sequence of posterior means is a Martingale. Now, using independence we know that

$$\begin{aligned}
\mathbb{E}_{x'}\left[x'|x\right] &= \int_0^{\overline{y}} \int_0^{\overline{y}} yf\left(y|x',e'\right) f\left(x''|e'-e\right) dydx'' \\
&= \int_0^{\overline{y}} \int_0^{\overline{y}} yf\left(y,x''|x,e'\right) dydx'' \\
&= \int_0^{\overline{y}} yf\left(y|x,e\right) dy \\
&= x
\end{aligned}$$

Therefore x' is a mean preserving spread of μ. Rothschild and Stiglitz (1970) show that this is equivalent to having

$$\int_0^a F\left(x|e\right) dx \leq \int_0^a F\left(x|e'\right) dx$$

for all $a \in [0, \overline{y}]$.

Using (31) we obtain $U\left(e\right) \leq U\left(e'\right)$. On the other hand, the fact that the posteriors are Martingale imply that when the number of independent signals becomes large enough, the posterior mean will converge almost surely to the true y_r (Doob, 1953). Formally, we have that $\lim_{e\rightarrow\infty} f\left(x|y_r,e\right) = 1$ if $x = y_r$ and 0 otherwise. Thus the limit unconditional distribution of the posterior is given by

$$
\begin{aligned}
\lim_{e\to\infty} f(x|e) &= \lim_{e\to\infty} \int_0^{\overline{y}} f(x|y_r, e)\, g(y_r)\, dy_r \\
&= \int_0^{\overline{y}} \lim_{e\to\infty} f(x|y_r, e)\, g(y_r)\, dy_r \\
&= g(x)
\end{aligned}
$$

where the second line is obtained using uniform convergence, which in turn is obtained from the almost surely convergence and the differentiability of the distribution (Ascoli's theorem). Thus the value of experimentation is bounded by $\mu_0 + \int_0^{y_s} G(y)\, dy < \infty$ □

Proof for Proposition 2. Define a compact domain for e where the upper and lower bound are given by the largest and the smallest e such that $C(e) = U(e)$, respectively. If there is no such e, the individual is better off by not experimenting ($e = 0$). Since the objective function is differentiable we know a maximum exists using Weierstrass Theorem. Moreover, using the Intermediate Value Theorem we know an interior optimum is characterized by

$$
U_e(e) - C_e(e) = 0
$$

The latter condition is also sufficient if the problem is concave, which is the case when $U_{ee}(e) = \int_0^{y_s} F_{ee})x|e)\, dx \leq 0$ since the cost is convex by assumption. The maximum is unique if the latter inequality is strict. □

Proof for Proposition 3. Suppose the agent decides to experiment, and thus is in an interior solution. Using the implicit function theorem we know

$$
\frac{\partial e^*}{\partial y_s} = -\frac{F_e(y_s|e)}{\int_0^{y_s} F_{ee}(x|e)\, dx - C_{ee}(e)}
$$

The numerator is strictly negative, then the sign of the derivative will be determined by the denominator. The ordering of the signals implies that the conditional distribution $f(x|y_r, e)$ is logsupermodular, thus the unconditional cdf $F(x|e)$ is also logsupermodular (see Athey (2002)). Therefore for every e, all the unconditional cdfs will cross uniquely at μ_0. To see this just consider the case when $e = 0$ and the distribution is degenerate at μ_0. Hence, $F_e(\mu_0|e) = 0$. Also note that by the second order stochastic dominance $F_e(y_s|e) > 0$ whenever $y_s < \mu_0$, and $F_e(y_s|e) < 0$ whenever $y_s > \mu_0$. Hence the optimal effort achieves a maximum when $y_s = \mu_0$, and is increasing (decreasing) for smaller (greater) y_s.

□

Proof for Proposition 4. Using the envelope theorem we obtain

$$\frac{\partial}{\partial y_s}\left(U(e^*) - C(e^*)\right) = F(y_s|e^*) \geq 0$$

which implies that the value is nondecreasing in y_s since the other alternatives are also nondecreasing in y_s. The second derivative is given by

$$\frac{\partial^2}{\partial^2 y_s}\left(U(e^*) - C(e^*)\right) = f(y_s|e^*) + F_e(y_s|e^*)\frac{\partial e}{\partial y_s}$$

The first element is always positive since it is a density function. The second term is also positive by the single crossing property and Proposition (3). Thus the function is strictly convex in y_s. Since the other alternatives are also convex in y_s, then the maximum of convex functions is also convex.

In the presence of fixed costs, $U(e^*) - C(e^*)$ as a function of y_s will cross at most once each of the outside options. It could cross once the constant μ_0 from below since $U(e^*) - C(e)$ is increasing in y_s. It could cross once y_s from above since its first derivative with respect to y_s are between 0 and 1. When there is no fixed cost, $c = 0$, $U(e) = \mu_0$ and $e^* = 0$ at $y_s = 0$. On the other hand, when $y_s = \overline{y}$, then

$U(e) = y_s$ and again $e^* = 0$. Therefore the principal always prefers to hire an agent for any interior y_s.

Since $U(e^*) - C(e^*)$ is linear in c, there exists a $\hat{c}$ such that $U(e^*) - C(e^*) = \mu_0 = y_s$. Thus, for any $c < \hat{c}$, there exists $a_c, b_c \in (0, \overline{y})$ such that $U(e^*) - C(e) > \mu_0$ for any $y_s > a_c$, and $U(e^*) - C(e) > y_s$ for any $y_s < b_c$. Obviously it must be the case that $\mu_0 \in (a_c, b_c)$. Note that a_c and b_c are increasing and decreasing in c, respectively, precisely because the function crosses from below and above each of the corresponding outside options. Finally, for any $c > \hat{c}$, the interval is empty and the principal never experiments.

$\square$