Experimentation and Project Selection: Screening and Learning*

Renato Gomes
Toulouse School of Economics
renato.gomes@tse-fr.eu

Daniel Gottlieb
The Wharton School, University of Pennsylvania
dgott@wharton.upenn.edu

Lucas Maestri
Toulouse School of Economics
lucas.maestri@tse-fr.eu

January 10, 2013

Abstract

We study optimal contracting in a setting that combines experimentation and adverse selection. In our leading example, an entrepreneur (agent) is better informed than the investor (principal) about both the quality the project (risky arm's distribution) and the entrepreneur's outside option (payoff of the safe arm). The investor's profit-maximizing mechanism can be uniquely implemented with a menu of *equity contracts* with different provisions on *control rights*. In each of these contracts, the investor offers a fixed initial payment (*seed money*) in exchange for a fixed share of the total revenue (*equity*), and a termination clause that specifies the critical number of failures before the project is aborted. We obtain necessary and sufficient conditions for the investor to extract all the rents from the entrepreneur's expertise on project quality. Our model has implications for the design of contracts to finance innovative activities.

JEL Classification: D21, J33, M12, O31, O34

Keywords: Adverse selection, experimentation, bandit problem, multi-dimensional screening, entrepreneurship

*We acknowledge funding support from the Mack Center for Technological Innovation at the Wharton School. We thank Niko Matouschek, Humberto Moreira, and Jean Tirole for helpful comments and suggestions.

1 Introduction

Entrepreneurs with great business ideas typically need external assistance to launch their startup dreams. The same is true, though, of those with not-so-great ideas. Unfortunately, telling good and bad ideas apart is often a hard task: it takes time and costly experimentation. How can an investor with limited expertise in the entrepreneur's business decide which ideas to finance? How does asymmetric information on project quality affect the amount of experimentation? What contractual solutions arise to mitigate adverse selection in project financing?

Questions like these arise in many economic settings that combine *experimentation* and *asymmetric information.* Consider a manager who wants to incentivize his employees to engage in innovative (and innately risky) activities. The employees may have better information about the true value of their ideas and can, therefore, assess their chances of success better than the manager. How can the manager screen employees with good ideas from those who are simply innovation-prone? How should the manager design in-house "innovation programs"? How should employees be rewarded for their innovative ventures?

In order to answer these questions, we introduce asymmetric information in an important class of Bayesian learning models known as two-armed bandit problems.[1] There is a safe arm with a known payoff distribution and a risky arm that may be either "good" or "bad". Revenues are always zero if the risky arm is bad. If the risky arm is good, large revenues come at geometrically distributed random times. Both the principal (investor) and the agent (entrepreneur) *learn* about the risky arm from the output history.

There is *asymmetric information* regarding the primitives of the bandit problem played by the agent. Namely, the agent has private information about the probability that the risky arm is good (*ex-ante quality of the project*), and about the payoff from playing the safe arm (*entrepreneur's opportunity cost*). Ideally, were information complete, the investor would select entrepreneurs with high-quality projects (generating high revenues once financed) and low opportunity costs (requiring small rents to participate). We follow a mechanism design approach to derive the financing contracts that maximize the investor profits under asymmetric information, and identify the contractual features that optimally screen the entrepreneur's (two-dimensional) private information.

Multi-dimensional screening problems such as the one considered here often pose considerable challenges. This paper sidesteps these difficulties by following a simple two-step procedure: In the first step, we consider an auxiliary (one-dimensional) screening problem in which the quality of the project is known by both parties, while the opportunity cost remains private information of the entrepreneur (we call it the *knowledgeable investor's problem*). The solution to this problem displays fairly intuitive features: The entrepreneur relinquishes the project's *control rights* to the investor, who offers a menu of "tenure" plans to the entrepreneur. Each tenure plan guarantees the continuation of

[1] Two-armed bandit problems were first introduced in the economics literature by Rothschild (1974).

the project until a critical period *regardless of the output realizations.* After this critical period, the project continues to be financed if and only if the entrepreneur generated high revenues for at least one period. Importantly, by offering tenure plans of different durations, the investor can effectively screen the entrepreneur's opportunity costs. While entrepreneurs with high opportunity costs choose short-term plans, those with low opportunity costs choose longer tenure plans.

In the second step, we reintroduce asymmetric information about the project's quality and study whether the solution to the knowledgeable investor's problem is implementable under two-dimensional asymmetric information. The answer to this question relies on the use of menus of output-contingent contracts, which make the entrepreneur's payments a function of the total revenues accrued to the project. One class of output-contingent contracts, labeled *equity-based contracts*, are particularly simple and highly employed in project financing. Such contracts specify the entrepreneur's remuneration as the sum of a fixed part (*seed money*) and a variable part which is linear in the total revenue accrued to the project (*equity share*). The investor is able to effectively screen the entrepreneur's private information about the quality of the project by offering a menu of equity contracts with different combinations of seed money and equity shares. Entrepreneurs with high quality projects have a comparative advantage at selecting "high-powered" contracts (that is, contracts with large equity shares but less seed money), while entrepreneurs with low quality projects select "low-powered" contracts.

We identify a necessary and sufficient condition for the solution of the knowledgeable investor problem to be implementable when the investor does not know the quality of the project. Expressed in terms of primitives, this condition imposes restrictions on the semi-elasticity of experimentation, which measures how the participation of entrepreneurs with a given opportunity cost changes as payments are increased by one dollar.[2] The necessary and sufficient condition requires the inverse of this semi-elasticity to be concave.

Intuitively, the inverse of the semi-elasticity of experimentation determines the rate of growth of informational rents. The concavity condition implies that the *marginal informational rent* associated with the entrepreneur's opportunity cost is decreasing. As a result, the investor is willing to finance entrepreneurs at a positive and increasing rate, which leads to expected payments that are increasing and convex in the quality of the project. Convexity of expected payments is the crucial condition for the implementability of the solution of the knowledgeable investor problem when project quality is not observable. Because convexity implies that the lower envelope of the entrepreneur's expected payments is increasing in project quality, we can construct equity contracts with increasing equity shares (in the sense that entrepreneurs with projects of greater quality are assigned larger equity shares), thereby satisfying the monotonicity condition that guarantees incentive compatibility. This semi-elasticity condition is satisfied by a wide range of distributions, including power function

[2] The semi-elasticity of experimentation coincides with the reverse hazard rate of the distribution of opportunity costs.

distributions (which we use to illustrate our results in closed-form).

One surprising corollary of this implementability result is that the entrepreneur's private information about the quality of the project can be elicited at *no cost to the investor*. Intuitively, by tailoring the seed money and the equity share to different levels of project quality, equity contracts render the entrepreneur's expertise about his own project irrelevant regarding payments. As a result, all rents obtained by the entrepreneur stem from his private information about opportunity costs alone (that is, his options *outside of the partnership*).

We qualify the results above by studying alternative settings where information about the quality of the project cannot be elicited at zero cost, such as when the inverse semi-elasticity of experimentation is not concave, or the entrepreneur is subject to a binding limited liability constraint. In both settings, the optimal mechanism can still be implemented by a menu of equity contracts. However, because the monotonicity condition for implementability is violated, we need to apply an approach analogous to the "ironing principle". As a result, a positive mass of types receive the same contract – i.e., "bunching" occurs at the optimum. In the "extreme" cases where (i) the inverse of the semi-elasticity is strictly convex, or (ii) there are types with sufficiently low opportunity costs, "complete bunching" occurs and the optimal mechanism offers the *same equity-based contract to all types.*

Perhaps surprisingly, in both settings where the quality of the project cannot be elicited at zero cost, a positive mass of types obtain *strictly lower payoffs* when they are privately informed about both ex-ante quality and opportunity costs, than when they have private information on their opportunity costs only. Therefore, some types actually receive *negative information rents* from their private information on the ex-ante quality of the risky arm. This result is an interesting consequence of the multi-dimensionality of our environment (in standard one-dimensional screening models, agents always get non-negative rents from their private information).

Related Literature

Our paper is related to the literature on incentives for experimentation. In settings with a single agent, Bergemann and Hege (1998, 2005) and Hörner and Samuelson (2012) analyze dynamic hidden action models in which the principal and the agent learn about the value of the investment over time. In turn, Bonatti and Hörner (2011) study experimentation in teams with unobservable actions. In a related setting, Campbell, Ederer, Spinnewijn (2012) analyze the role of information sharing for experimentation in teams.

Assuming that the principal can commit to long-term contracts, Manso (2012) shows that contracts that foster experimentation greatly differ from standard pay-for-performance contracts. Ederer (2011) extends the framework of Manso to a setting where a principal faces multiple agents, and tests the theory with a laboratory experiment. In turn, Garfagnini (2012) studies optimal dynamic delegation of power to an agent who can take hidden actions. Klein (2012) analyzes how a principal should optimally motivate an agent who is able to choose one among several experimentation policies. Gerardi and Maestri (2012) study optimal sequential testing when there are agency issues. In an environment with moral hazard and adverse selection, Canidio and Legros (2012) analyze how the presence of career concerns shapes optimal experimentation contracts. Halac, Kartin and Liu (2012) demonstrate that, when there is moral-hazard, optimal experimentation contracts induce under experimentation by low-ability agents. The presence of moral hazard is a central element in all papers above. In contrast, our model focuses on adverse-selection aspects of the agency conflict.

The study of experimentation under asymmetric information naturally leads to a multi-dimensional screening environment. Solving multi-dimensional screening problems is often challenging, as one cannot determine from the outset the direction in which incentive constraints bind (this is known as the *integrability condition* – see Rochet and Stole (2003) and references therein). In our setting, the nature of the bandit problem played by the agent allows us to sidestep this issue by applying the two-step procedure described previously.

The remainder of the paper is organized as follows. Section 2 introduces the framework. Section 3 solves for the two benchmark cases of symmetric information (first best) and asymmetric information about opportunity costs only (knowledgeable investor). Section 4 identifies conditions for the solution of the knowledgeable investor problem to be implementable under two-dimensional asymmetric information. Then, it characterizes the optimal mechanism and discusses its unique implementation by a menu of equity-based contracts. Section 5 characterizes the optimal mechanism in environments where the solution of the knowledgeable investor problem is not implementable under two-dimensional asymmetric information. Section 6 concludes.

2 Model

One investor (principal) and one entrepreneur (agent) populate a dynamic economy with T periods, indexed by $t \in \{1, \ldots T\}$. At each period, the entrepreneur has the option of taking a "regular job" (*safe arm*) or launching a risky (but potentially highly profitable) project (*risky arm*). Undertaking the risky project requires an outlay (to cover the operational costs) of $K > 0$ per period.

The entrepreneur is not able to launch the risky project without the investor's assistance. Such assistance can take several forms: Entrepreneurs may, for example, need to have access to assets (or technology) held by the investor in order to develop the project; they may also be financially constrained, and therefore need the investor's financial support to cover (at least part of) the project's operational costs.

The relationship between the investor and the entrepreneur is plagued by asymmetric information. The entrepreneur is better informed than the investor about the distribution of returns of the project (risky arm), captured by the parameter θ (to which we refer as the *quality of the project*) and about his payoff in the regular job (safe arm), denoted by c (to which we refer as the entrepreneur's *opportunity cost*).[3]

Geometric Bandit Returns. The quality of the project θ captures the ex-ante the probability that the project is *good* (as opposed to being *bad*). A bad project generates zero revenue in each period with probability one. A good project produces revenue $\triangle > 0$ at geometrically distributed random times. Accordingly, in each period, the project's revenue equals $\triangle$ with probability λ and zero with complementary probability.[4] Therefore, the expected per-period profit (i.e., revenues net of costs) from a good project is

$$\lambda \cdot \triangle - K,$$

which we assume to be strictly positive.

We refer to the event that the project generates revenue $\triangle$ as a *success,* and we refer to the event that it generates no revenue as a *failure.* The project's output at period t is denoted by $x_t \in \{s, f, \emptyset\} \equiv X$, where $x_t = s$ ($x_t = f$) denotes a success (failure), and $x_t = \emptyset$ means that the project is inactive at period t. If the project is inactive, it produces a (riskless) cash flow normalized to zero.

Learning. An important feature of our model is that both the investor and the entrepreneur are able to learn about the quality of the project from its output history. We refer to the sequence of outputs $x^t \equiv (x_1 \ldots, x_t) \in X^t$ as the project's *output history* at period t.

The entrepreneur and the investor update beliefs according to Bayes' rule. A success immediately reveals that the project is good, whereas a failure is only partial information about the quality of the project. An entrepreneur with initial belief θ who observes failures in all periods $t \in \{1, \ldots, t-1\}$ believes that the project is good with probability

$$\theta_t(\theta) \equiv \frac{\theta \cdot (1-\lambda)^{t-1}}{\theta \cdot (1-\lambda)^{t-1} + (1-\theta)}.$$

Thus, in each period before a success is observed, either a new failure occurs, which reduces the posterior probability to $\theta_{t+1}(\theta)$, or a success occurs, which shifts the posterior probability to 1. The entrepreneur's belief that the project is good after t observations is a random variable with support on $\{\theta_t(\theta), 1\}$. We denote this random variable by $\tilde{\theta}_t(\theta)$.

Importantly, the output history is public (i.e., observed by the investor and the entrepreneur) and contractible (i.e., financing contracts can be made contingent on the output history – more on

[3]There is no loss of generality in assuming that the safe arm produces a deterministic payoff. Because there is no learning on the part of the agent about this arm, one can always replace the safe arm's random payoff by its expectation.

[4]Keller, Rady, and Cripps (2005) consider the continuous-time version of this stochastic process, and label it the *exponential bandit* process, as the time before a success follows an exponential distribution. We adapt their nomenclature by noting that the geometric distribution is the discrete-time analog of the exponential distribution and, therefore, refer to this process as a *geometric bandit.* Two-armed geometric bandits are adopted, among many others, by Bergemann and Hege (1998, 2005), Hörner and Samuelson (2011), and Gerardi and Maestri (2012).

this below). For future reference, and with a slight abuse of notation, let $\mathbf{S}(x^t)$ and $\mathbf{F}(x^t)$ denote the number of successes and failures associated with the output history x^t at period t.

Asymmetric Information. The quality of the project θ and the opportunity cost c are private information of the entrepreneur. From the investor's perspective, the quality θ is a draw from a distribution G with support $[0, 1]$, while the opportunity cost c is a draw from the distribution H with support $[\underline{c}, \overline{c}] \subseteq \mathbb{R}_{++}$. Both distributions are absolutely continuous, with differentiable densities g and h. The parameter λ, which is the probability that a good project produces a success at any given period, is common knowledge between the investor and entrepreneur.

We make two technical assumptions. First, as is usually assumed in the bandit literature, the distributions G and H are independent. Second, as is standard in mechanism design, we require the distribution H to be log-concave. To rule out uninteresting cases, we further assume that it is inefficient for entrepreneurs with the highest opportunity cost to undertake the project even when the project is known to be good, and that it is efficient for entrepreneurs with the lowest opportunity cost to undertake the project if the project is known to be good: $\overline{c} \geq \lambda \cdot \triangle - K \geq \underline{c}$.

The entrepreneur's private information at $t = 1$ is described by the pair $\tau \equiv (\theta, c) \in \Gamma \equiv [0, 1] \times [\underline{c}, \overline{c}]$. Because the output of the project is publicly observable, the pair $\tau \equiv (\theta, c)$ fully summarizes the entrepreneur's private information conditional on any output history. We therefore refer to τ as the entrepreneur's *type*.

Financing Mechanisms

We assume that the investor makes a take-it-or-leave-it offer to the entrepreneur in the form of a *mechanism*. A mechanism consists of an action plan ϕ and a payment rule p. An *action plan* is a collection of functions $\phi_t : \Gamma \times X^t \to [0, 1]^t$, $t = 1, ..., T$. The function $\phi_t(\tau|x^{t-1})$ specifies, for each type τ, the probability of continuing the project at period t when the output history up to that period is x^{t-1}. If the action plan ϕ, following an output history x^{t-1}, determines that the project shall be terminated in period t, the entrepreneur is assumed to play the safe arm (reaping a per-period payoff of c) from period t on. We assume that parties cannot renege the recommendation of an action plan after agreeing to follow it (e.g., because of vesting provisions or large severance payments).[5]

A *payment rule* is a function $p : \Gamma \times X^T \to \mathbb{R}$ that maps the entrepreneur's type τ and the final output history x^T into a payment to the entrepreneur. For simplicity, and with no loss of generality, we evaluate payments in units of the initial period. Although this normalization may seem counterintuitive (payments are conditional on the final history), it allows us to compare payments between financing ventures that last for different numbers of periods. A mechanism is therefore described by the pair $\mathcal{M} = (\phi, p)$.

[5]Kaplan and Stromberg (2003) document the widespread use of vesting provisions and severance payments in venture capital financing.

For a given project of quality θ, an action plan ϕ uniquely defines a probability measure over the space of final output histories X^T, which we denote by $\lambda[\phi(\tau)|\theta]$. Define the expected payment of an entrepreneur that reports type $\hat{\tau}$ when the project's quality is θ as

$$\tilde{P}(\hat{\tau};\theta) \equiv \mathbb{E}^{\lambda[\phi(\hat{\tau})|\theta]}\left[p(\hat{\tau},\tilde{x}^T)\right],$$

and denote by $P(\tau) \equiv \tilde{P}(\tau;\theta)$ the expected payments associated with truthful reporting. Analogously, the expected discounted duration of a project of quality θ for an entrepreneur that reports type $\hat{\tau}$ is

$$\tilde{\Phi}(\hat{\tau};\theta) \equiv \mathbb{E}^{\lambda[\phi(\hat{\tau})|\theta]}\left[\sum_{t=1}^{T}\delta^t \cdot \phi_t(\hat{\tau}|\tilde{x}^{t-1})\right],$$

and $\Phi(\tau) \equiv \tilde{\Phi}(\tau;\theta)$ is the expected discounted duration associated with truthful reporting.

Equipped with this notation, we can write the expected payoff of an entrepreneur of type $\tau = (\theta, c)$ who reports his type truthfully under the mechanism $\mathcal{M} = (\phi, p)$ as

$$U(\tau|\mathcal{M}) \equiv P(\tau) - \Phi(\tau) \cdot c. \tag{1}$$

By the Revelation Principle, we can restrict attention to truthful direct-revelation mechanisms. A mechanism $\mathcal{M} = (\phi, p)$ is *incentive compatible* (IC) if the entrepreneur cannot benefit from deviating to a non-truthful reporting strategy:

$$\tilde{P}(\hat{\tau};\theta) - \tilde{\Phi}(\hat{\tau};\theta) \cdot c \leq U(\tau|\mathcal{M}) \tag{IC}$$

for all $\tau = (\theta, c)$ and $\hat{\tau}$ in Γ.

A mechanism $\mathcal{M}$ is *individually rational* (IR) if, for each $\tau = (\theta, c) \in \Gamma$, the entrepreneur is better-off by accepting the investor's mechanism rather than exerting his outside option from the outset (i.e., playing the safe arm by himself and obtaining c per period):

$$U(\tau|\mathcal{M}) \geq 0. \tag{IR}$$

A mechanism $\mathcal{M} = (\phi, p)$ that satisfies the IC and IR constraints is said to be *feasible.* An action plan ϕ is said to be *implementable* if there exists a payment rule p such that $\mathcal{M} = (\phi, p)$ is feasible. A mechanism $\mathcal{M}$ is *optimal* if it maximizes the investor's expected profit,

$$\Pi(\mathcal{M}) \equiv \mathbb{E}^{\theta,c}\left\{\mathbb{E}^{\lambda[\phi(\tau)|\theta]}\left[\sum_{t=1}^{T}\left(\delta^t \cdot \phi_t(\tau|\tilde{x}^{t-1}) \cdot \left(\tilde{\theta}_t\left(\theta\right) \cdot \lambda \cdot \triangle - K\right)\right) - p(\tau, x^T)\right]\right\}, \tag{2}$$

within the class of feasible mechanisms.

3 Preliminaries

We will now consider two benchmarks that lay the ground to our characterization of the optimal mechanism. We start with the first-best benchmark.

3.1 First-Best Benchmark

In the first best, the investor observes the entrepreneur's type $\tau = (\theta, c)$. Accordingly, we can disregard the incentive constraints. We say that a mechanism is *first-best optimal* if it maximizes the investor's expected profit (2) subject to the individual rationality constraint (IR).

Because the investor wants to minimize payments to the entrepreneur, it is immediate that the IR constraint binds for every type $\tau \in \Gamma$ in any first-best optimal mechanism. Then, payments after every final output history have to equal the total rents foregone by the entrepreneur by not playing the safe arm:

$$p(\tau, x^T) = \sum_{t=1}^{T} \delta^t \cdot \phi_t(\tau|\tilde{x}^{t-1}) \cdot c.$$

After plugging the expression above into the objective (2), the principal's problem can be recast as that of choosing, for each type $\tau \in \Gamma$, an action plan $\phi(\tau)$ that maximizes

$$\mathbb{E}^{\lambda[\phi(\tau)|\theta]} \left[\sum_{t=1}^{T} \left(\delta^t \cdot \phi_t(\tau|\tilde{x}^{t-1}) \cdot \left(\tilde{\theta}_t\left(\theta\right) \cdot \lambda \cdot \triangle - K - c \right) \right) \right]. \tag{3}$$

As the objective function (3) reveals, the principal faces a two-arm bandit problem with (i) a risky arm that yields positive revenues at geometrically distributed random times (in case the project is good) but costs $K + c$ at each round, and (ii) a safe arm with a zero deterministic payoff.

For each type $\tau = (\theta, c)$, the geometric bandit problem described above is the discrete-time analogue of the exponential bandit considered by Keller, Rady, and Cripps (2005). Analogously to its continuous-time version, the principal's payoff (3) is maximized by an action plan with the following threshold structure: If a success is observed at some period t, the project is continued until the final period T. If only failures occur, the project is terminated at some critical period $k(\tau)$. An action plan with this structure is called a *tenure action plan*:

Definition 1 (Tenure Action Plans) *An action plan is said to be a tenure action plan if for every $\tau \in \Gamma$ there exists a number $k(\tau) \in \{1, \ldots, T\}$ such that*

1. *the project is financed up to period $k(\tau)$ irrespective of its output history: $\phi_t(\tau|\cdot) = 1$ for all $t \leq k(\tau)$;*
2. *the project is terminated in case of $k(\tau)$ consecutive failures starting from $t = 1$: $\phi_t(\tau|x^{t-1}) = 0$ for all $t > k(\tau)$ if $\mathbf{F}(x^{k(\tau)}) = k(\tau)$; and*
3. *the project is financed up to period T if at least one success occurs in some period $t \leq k(\tau)$: $\phi_t(\tau|x^{t-1}) = 1$ for all $t \leq T$ if $\mathbf{S}(x^{k(\tau)}) \geq 1$.*

In this case, we say that type τ bears $k(\tau)$ failures.

For a heuristic derivation of the first-best optimal action plan, fix the type $\tau = (\theta, c)$ and assume that $k-1$ failures were observed in the first $k-1$ periods. Consider the follow continuation strategy:

"Continue the project in period k. If a success happens in that period, continue until period T. Otherwise, terminate the project."

The expected payoff from this strategy from the perspective of period k is:

$$V_k(\theta, c) \equiv \underbrace{\theta_k(\theta) \cdot \lambda \cdot \triangle - K - c}_{flow\ payoff\ in\ period\ k} \quad + \quad \mathbf{1}_{[k<T]} \cdot \underbrace{\theta_k(\theta) \cdot \lambda \cdot \delta \cdot \left[\frac{1-\delta^{T-k}}{1-\delta} \cdot (\lambda \cdot \triangle - K - c) \right]}_{option\ value\ from\ a\ good\ project\ in\ period\ k+1}, \tag{4}$$

where $\mathbf{1}_{[k<T]}$ is the indicator function that equals one if $k < T$ and zero if $k = T$. The first term is the expected flow payoff from continuing the project at period k. The second term is the expected option value of arriving at period $k+1$ knowing that the project is good.

Note that the expected payoff $V_k(\theta, c)$ is strictly decreasing in c. Therefore, for every $k \in \{1, \ldots, T\}$, we can define the *threshold opportunity cost* $v_k^{FB}(\theta)$ as the opportunity cost that equates the payoff from continuing the project at period k (and then following the strategy above) to zero (which is continuation payoff from terminating the project in period $k-1$ and playing the safe arm). That is, the threshold opportunity cost $v_k^{FB}(\theta)$ is the solution to the following *Gittins formula*:

$$V_k(\theta, v_k^{FB}(\theta)) = 0. \tag{5}$$

The next Lemma describes the first-best optimal action plan, which follows from Gittins (1979).

Lemma 1 (First Best) *The first-best optimal action plan ϕ^{FB} is a tenure action plan. For every $k \in \{1, \ldots, T\}$, consider the threshold functions $v_k^{FB}(\cdot)$ defined according to the Gittins formula (5). An entrepreneur with type $\tau = (\theta, c)$ bears $k^{FB}(\tau) \in \{1, \ldots, T\}$ failures if and only if*

$$v_{k^{FB}(\tau)+1}^{FB}(\theta) < c \leq v_{k^{FB}(\tau)}^{FB}(\theta).$$

Figure 1 illustrates the first-best optimal action plan when $T = 3$, $[\underline{c}, \overline{c}] = [1, 2]$, $\lambda = \frac{1}{2}$, $K = 1$, $\triangle = 6$ and $\delta \approx 1$.

As a stepping stone to characterize the optimal mechanism, we will now consider the intermediary situation where the entrepreneur has private information only about the payoff of the safe arm (the opportunity cost c).

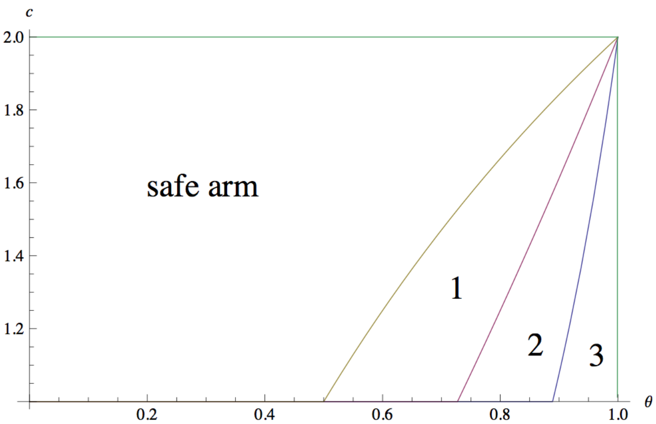

Figure 1: The first-best optimal action plan when $[\underline{c}, \bar{c}] = [1, 2]$ under the following parameterization: $T = 3$, $\lambda = \frac{1}{2}$, $K = 1$, $\triangle = 6$ and $\delta \approx 1$. Areas 1, 2 and 3 refer to the set of types that play the risky arm at $t = 1$ and bear, respectively, 1, 2 and 3 failures according to ϕ^{FB}.

3.2 Knowledgeable Investor Benchmark

Consider an investor who observes the ex-ante quality θ but does not observe the entrepreneur's opportunity cost c. In this environment, the associated entrepreneur's incentive constraints are one-dimensional since they only concern truthful revelation of the opportunity cost. Accordingly, we say that a mechanism $\mathcal{M} = (\phi, p)$ satisfies *incentive compatibility on the safe arm* (ICS) if no type (θ, c) benefits from reporting some type $(\theta, \hat{c})$:

$$\tilde{P}(\theta, \hat{c}; \theta) - \tilde{\Phi}(\theta, \hat{c}; \theta) \cdot c \leq U(\tau|\mathcal{M}) \tag{ICS}$$

for all $\tau = (\theta, c) \in \Gamma$ and $\hat{c} \in [\underline{c}, \bar{c}]$.

A mechanism $\mathcal{M}$ that satisfies individual rationality (IR) and incentive compatibility on the safe arm (ICS) is said to be *feasible for a knowledgeable investor* (or *i-feasible*, for short). A mechanism $\mathcal{M}$ is *optimal for the knowledgeable investor* (or *i-optimal*, for short) if it maximizes the investor's expected profit (2) within the class of i-feasible mechanisms. We denote an i-optimal mechanism by $\mathcal{M}^i = \left(\phi^i, p^i\right)$.

Let $\sigma(c) \equiv \frac{h(c)}{H(c)}$ denote the *semi-elasticity of experimentation.* This semi-elasticity measures the percentage change in the mass of entrepreneurs with opportunity cost c that play the risky arm at $t = 1$ as the expected per-period rents increase by one dollar. Because the distribution H is log concave, the semi-elasticity of experimentation decreases as c increases. The next lemma characterizes the i-optimal mechanism.

Lemma 2 (Knowledgeable Investor Optimum) *The i-optimal mechanism employs a tenure ac-*

tion plan ϕ^i described by the threshold functions $v_k^i(\theta)$ that solve the following Virtual Gittins formula:

$$V_k\left(\theta, v_k^i(\theta) + \frac{1}{\sigma\left(v_k^i(\theta)\right)}\right) = 0. \tag{6}$$

An entrepreneur with type $\tau = (\theta, c)$ bears $k^i(\tau) \in \{1, \ldots, T\}$ failures if and only if

$$v_{k^i(\tau)+1}^i(\theta) < c \leq v_{k^i(\tau)}^i(\theta),$$

and receives an expected payment of

$$P^i(\tau) = \Phi_1(\theta) \cdot v_1^i(\theta) + \sum_{k=2}^{k^i(\tau)} \left(\Phi_k(\theta) - \Phi_{k-1}(\theta)\right) \cdot v_k^i(\theta), \tag{7}$$

where $\Phi_k(\theta)$ is the expected discounted duration of a project of quality θ for an entrepreneur type that bears k failures:

$$\Phi_k(\theta) = \left(\frac{1-\delta^k}{1-\delta}\right) + \mathbf{1}_{[k<T]} \cdot \delta^k \cdot \theta \cdot \left(1 - (1-\lambda)^k\right) \cdot \left(\frac{1-\delta^{T-k}}{1-\delta}\right). \tag{8}$$

The presence of asymmetric information on c implies that the investor has to pay informational rents to screen the entrepreneur's opportunity costs. The lemma above shows that the knowledgeable investor optimum can be obtained from the first-best action plan after replacing the cost c by its virtual counterpart $c + \frac{1}{\sigma(c)}$, where the inverse of the semi-elasticity of experimentation captures the informational rents obtained by the entrepreneur. This leads to the Virtual Gittins formula (6), which equates the expected flow payoff from continuing the project at period k to the expected option value of arriving at period $k+1$ with a good project, *after adjusting for the entrepreneur's informational rents*.

Importantly, for each quality θ, the tolerance for failures $k^i(\theta, c)$, which is decreasing in c, serves to screen entrepreneurs with different opportunity costs. Entrepreneurs with *higher* opportunity costs sign contracts with the investor that tolerate a *smaller* number of failures before termination. It follows from (6) that no entrepreneur with opportunity cost greater than c^{max}, where c^{max} solves $\lambda \cdot \triangle - K = c^{max} + \frac{1}{\sigma(c^{max})}$, will ever play the risky arm, regardless of the project quality θ. Because $\bar{c} \geq \lambda \cdot \triangle - K \geq \underline{c}$, it follows that $c^{max} \in [\underline{c}, \bar{c})$. We refer to $[\underline{c}, c^{max}]$ as the *relevant cost range*.

The formula (7) directly computes the payments needed to implement the action plan ϕ^i. In order to prevent upward deviations on reported costs, an entrepreneur with type $\tau = (\theta, c)$ has to receive the foregone rents that would arise from such an inflated cost report. As it is clear from the formula, these foregone rents take into account the distinct expected durations associated with different cost reports (since reporting a higher cost leads to contracts that bear a weakly smaller number of failures).

Figure 2 illustrates the i-optimal action plan when $c \sim U[1,2]$ under the following parameterization: $T = 3$, $\lambda = \frac{1}{2}$, $K = 1$, $\triangle = 6$ and $\delta \approx 1$.

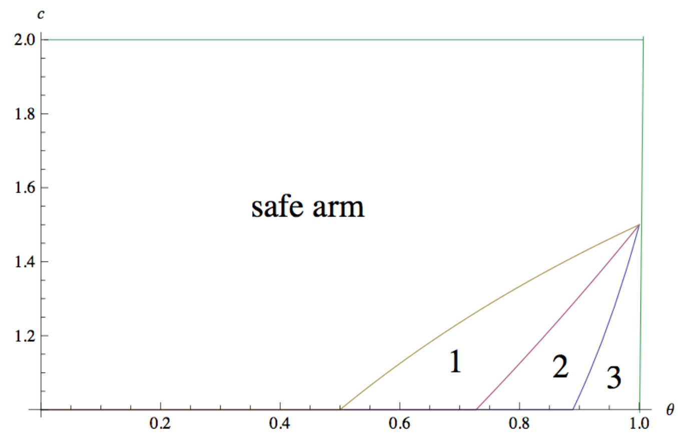

Figure 2: The i-optimal action plan when $c \sim U[1,2]$ under the following parameterization: $T = 3$, $\lambda = \frac{1}{2}$, $K = 1$, $\triangle = 6$ and $\delta \approx 1$. Areas 1, 2 and 3 refer to the set of types that play the risky arm at $t = 1$ and bear, respectively, 1, 2 and 3 failures according to ϕ^i.

Because there is double-sided commitment, the actual timing of payments is immaterial in the knowledgeable investor benchmark. One payment rule that implements the i-optimal action rule ϕ^i is to offer a single lump-sum payment worth of $P^i(\tau)$ monetary units (in the perspective of period $t = 1$) once the project is terminated. This rule is attractive since it dispenses with commitment on the side of the agent. More generally, any flow of lump-sum payments worth of $P^i(\tau)$ monetary units (in the perspective of period $t = 1$) implements the i-optimal action plan ϕ^i.

4 Linear Contracts and the Irrelevance of Expertise

We now investigate the investor's optimal mechanism in the more realistic scenario where the entrepreneur is privately informed about both his opportunity cost c (safe arm) and the quality of the project θ (risky arm). As will become clear shortly, the form of payments offered by the investor plays a fundamental role in screening the entrepreneur's two-dimensional private information. The following definition introduces an important class of payment rules.

Definition 2 (Linear Contracts) *A payment rule p takes the form of a* menu of linear contracts $\{a(\tau), b(\tau)\}_{\tau\in\Gamma}$ *if for every type $\tau = (\theta, c) \in \Gamma$, the payments received by the entrepreneur take the form*

$$p(\tau, x^T) = a(\tau) + \sum_{\{t: x_t = s\}} \delta^{t-1} \cdot b(\tau).$$

That is, payments to the entrepreneur are linear in the number of successes. We refer to $a(\tau)$ as the entrepreneur's initial payment and to $b(\tau)$ as the success bonus.

We will study when it is possible to elicit information about the ex-ante quality of project at no additional cost to the investor. The following definition formalizes this property:

Definition 3 (Irrelevance of Expertise) *We say that the entrepreneur's expertise (about the quality of the project) is irrelevant if the action plan ϕ^i associated with the i-optimal mechanism is implementable under asymmetric information over the two-dimensional type (θ, c).*

The following assumption plays an important role in the analysis of this section. We will discuss its interpretation in detail after the proof of Proposition 1.

Condition 1 (C) *Concavity: The inverse of the semi-elasticity of experimentation, $\frac{1}{\sigma(c)}$, is weakly concave on the relevant cost range, i.e., at each $c \in [\underline{c}, c^{max}]$.*

The next proposition establishes that condition C is both necessary and sufficient for the entrepreneur's expertise to be irrelevant. Moreover, under condition C, the tenure action plan ϕ^i can be implemented by a menu of linear contracts.

Proposition 1 (Optimal Mechanism) *The entrepreneur's expertise is irrelevant if and only if Condition C holds. In this case, the optimal action plan coincides with the i-optimal action plan: $\phi^* = \phi^i$. Moreover, ϕ^* is implementable by a menu of linear contracts p^*, with initial payments and success bonuses given by*

$$a^*(\tau) = P^i(\tau) - b^*(\tau) \cdot \Lambda_{k^i(\tau)}(\theta) \qquad \text{and} \qquad b^*(\tau) = \frac{\frac{\partial}{\partial\theta}\left[P^i(\tau)\right]}{\frac{d}{d\theta}\left[\Lambda_{k^i(\tau)}(\theta)\right]}, \tag{9}$$

where $\Lambda_k(\theta)$ is the expected discounted number of successes of an entrepreneur who bears exactly k failures with a project of quality θ:

$$\Lambda_k(\theta) = \theta \cdot \lambda \cdot \left(\frac{1-\delta^k}{1-\delta}\right) + \mathbf{1}_{[k<T]} \cdot \delta^k \cdot \theta \cdot \lambda \cdot \left(1 - (1-\lambda)^k\right) \cdot \left(\frac{1-\delta^{T-k}}{1-\delta}\right). \tag{10}$$

Proof of Proposition 1. First note that the knowledgeable investor problem is a relaxation of the investor problem (when there is two-dimensional asymmetric information on (θ, c)). As a consequence, the optimal action plan ϕ^* coincides with the i-optimal action plan ϕ^i provided that the latter is implementable (in the sense of the IC).

We structure this proof in two steps. In the first step, we will show that there exists a menu of linear contracts that implements ϕ^i if and only if condition C holds. In the second step, we will show that if there is no menu of linear contracts that implements ϕ^i, then ϕ^i is not implementable.

Step 1 *There exists a unique menu of linear contracts that implements ϕ^i if and only if condition C holds.*

By ICS, it is clear that any payment rule that implements ϕ^i leads to expected payments $P^i(\tau)$. By defining the initial payment $a^*(\tau)$ according to $a^*(\tau) = P^i(\tau) - b^*(\tau) \cdot \Lambda_{k^i(\tau)}(\theta)$, we assure by construction that

$$\begin{aligned} P^*(\tau) &\equiv \mathbb{E}^{\lambda[\phi^*(\tau)|\theta]}\left[a^*(\tau) + \textstyle\sum_{\{t:x_t=s\}} \delta^{t-1} \cdot b^*(\tau)\right] \\ &= \mathbb{E}^{\lambda[\phi^i(\tau)|\theta]}\left[a^*(\tau) + \textstyle\sum_{\{t:x_t=s\}} \delta^{t-1} \cdot b^*(\tau)\right] \\ &= P^i(\tau) - b^*(\tau) \cdot \Lambda_{k^i(\tau)}(\theta) + b^*(\tau) \cdot \Lambda_{k^i(\tau)}(\theta) \quad = \quad P^i(\tau), \end{aligned}$$

where the equality in the second line follows from $\phi^* = \phi^i$. We will now derive the unique success bonus $b^*(\tau)$ that, together with $a^*(\tau)$, implements the action plan ϕ^i.

To this end, consider a mechanism $\mathcal{M} = (\phi, p)$ such that ϕ is a tenure action plan. Let us introduce the following incentive constraints regarding non-truthful reporting about the risky arm (denoting $\tau = (\theta, c)$):

$$\tilde{P}(\hat{\tau};\theta) - \Phi_{k(\hat{\tau})}(\theta) \cdot c \leq U(\tau|\mathcal{M}) \quad \forall \tau, \hat{\tau} \in \Gamma \;\; with \;\; \theta \neq \hat{\theta} \;\; and \;\; k(\hat{\tau}) = k(\tau), \tag{ICR$_1$}$$

and

$$\tilde{P}(\hat{\tau};\theta) - \Phi_{k(\hat{\tau})}(\theta) \cdot c \leq U(\theta, c|\mathcal{M}) \quad \forall \tau, \hat{\tau} \in \Gamma \;\; with \;\; \theta \neq \hat{\theta} \;\; and \;\; k(\hat{\tau}) \neq k(\tau). \tag{ICR$_2$}$$

Condition ICR$_1$ rules out non-truthful reports on θ (and potentially on c as well) such that $k(\hat{\tau}) = k(\tau)$ (i.e., non-truthful reports that lead to bearing the same number of failures as the truthful report). In turn, condition ICR$_2$ rules out non-truthful reports on θ (and potentially on c as well) such that $k(\hat{\tau}) \neq k(\tau)$ (i.e., non-truthful reports that lead to bearing a number of failures different from that implied by the truthful report). Clearly, for a tenure action plan ϕ, the mechanism $\mathcal{M} = (\phi, p)$ satisfies condition IC if and only if it satisfies conditions ICS, ICR$_1$ and ICR$_2$.

By construction, $\phi^* = \phi^i$, and, by the choice of $a^*(\tau)$, $P^*(\tau) = P^i(\tau)$. Lemma 2 then implies that the optimal mechanism $\mathcal{M}^* = (\phi^*, p^*)$ satisfies condition ICS, in which case the entrepreneur's payoff is $U(\theta, c|\mathcal{M}^*) = P^*(\tau) - \Phi_{k^*(\tau)}(\theta) \cdot c = P^i(\tau) - \Phi_{k^i(\tau)}(\theta) \cdot c$. Note that $k^i(\tau)$ has zero partial derivative with respect to θ almost everywhere (and is not differentiable along the path $(\theta, v^i_{k^i(\tau)}(\theta))$). Therefore,

$$\frac{\partial U(\theta, c|\mathcal{M}^*)}{\partial \theta} = \frac{\partial}{\partial \theta}\left[P^i(\tau)\right] - \frac{d}{d\theta}\left\{\Phi_{k^i(\tau)}(\theta)\right\} \cdot c \qquad almost \;\; everywhere. \tag{11}$$

We will now turn to condition ICR$_1$. Observe that the entrepreneur's payoff satisfies strict increasing differences in the success bonus and the project's quality. Building on this observation, the following lemma employs standard techniques to characterize the set of mechanisms that satisfy condition ICR$_1$. In particular, the next lemma delivers an alternative envelope formula that depends explicitly on the schedule of bonuses $b(\tau)$. Coupled with (11), this new envelope formula pins down the unique candidate schedule of bonuses that implement $\phi^* = \phi^i$.

Lemma 3 *Consider a mechanism $\mathcal{M} = (\phi, p)$, where ϕ is a tenure action plan and $p = \{a(\tau), b(\tau)\}_{\tau \in \Gamma}$ is a menu of linear contracts. The mechanism $\mathcal{M}$ satisfies condition ICR$_1$ if and only if the following conditions jointly hold:*

1. *For every $\hat{k} \in \{1, ..., T\}$, the success bonus $b(\tau)$ restricted to support $\{\tau : k(\tau) = \hat{k}\}$ is weakly increasing in θ,*

2. *The envelope formula holds almost everywhere:*

$$\frac{\partial U(\theta, c|\mathcal{M})}{\partial \theta} = b(\tau) \cdot \frac{d}{d\theta} \left[\Lambda_{k(\tau)}(\theta)\right] \cdot \triangle - \frac{d}{d\theta} \left\{\Phi_{k(\tau)}(\theta)\right\} \cdot c. \tag{12}$$

Proof of Lemma 3. See appendix. Q.E.D.

Setting the envelope formula (12) evaluated at $\mathcal{M}^*$ equal to equation (11) pins down the optimal bonus $b^*(\tau)$ described in the statement of the current proposition. To conclude that the optimal mechanism $\mathcal{M}^*$ satisfies condition ICR$_1$, it only remains to be shown that the bonus $b^*(\tau)$ satisfies the monotonicity condition 1 from Lemma 3. The next lemma establishes that condition C is both a necessary and sufficient condition for the latter property to hold.

Lemma 4 *The following statements are equivalent:*

1. *For any $\hat{k} \in \{1, ..., T\}$, the bonus $b^*(\tau)$ restricted to the support $\{\tau : k^i(\tau) = \hat{k}\}$ is weakly increasing in θ.*

2. *For any $\hat{k} \in \{1, ..., T\}$, $P^i(\tau)$ restricted to the support $\{\tau : k^i(\tau) = \hat{k}\}$ is weakly convex in θ.*

3. *Condition C holds.*

Proof of Lemma 4. Formula (10) computes the expected discounted number of successes of an entrepreneur who bears exactly k failures with a project of quality θ, $\Lambda_k(\theta)$. Note that, for any k, $\Lambda_k(\theta)$ is a linear function of θ, in which case $\frac{d}{d\theta}\left[\Lambda_k(\theta)\right]$ is a constant. Therefore, by the definition of $b^*(\tau)$ in equation (9), it follows that conditions 1 and 2 are equivalent.

We will now establish the equivalence between conditions 2 and 3. To clarify ideas, let us start with the simpler case where $T = 1$. By equation (7), the payments received by an active type (i.e., $k^*(\tau) = 1$) equal $P^*(\tau) = P^i(\tau) = v_1^i(\theta)$. As a consequence, $P^i(\tau)$ is weakly convex in θ if and only if $v_1^i(\theta)$ is weakly convex. In turn, the Virtual Gittins formula (6) implies that $v_1^i(\theta)$ is given by

$$\theta \cdot \lambda \cdot \triangle - K = v_1^i(\theta) + \frac{1}{\sigma\left(v_1^i(\theta)\right)}. \tag{13}$$

Because the left-hand side of the equation above is linear, it follows that $v_1^i(\theta)$ is weakly convex if and only if $\frac{1}{\sigma(c)}$ is weakly concave in every $c \in [\underline{c}, c^{max}]$.

Now consider $T > 1$ and fix some $\hat{k} \in \{1, ..., T\}$. Analogously to the case where $T = 1$, the weak concavity of $\frac{1}{\sigma(c)}$ in every $c \in [\underline{c}, c^{max}]$ is both a necessary and sufficient condition for $P^i(\tau)$ to be weakly convex in θ. To establish this claim, we differentiate $P^i(\tau)$ with respect to θ and employ the Virtual Gittins formula (6). The details of this computation can be found in the appendix. Q.E.D.

In order to complete Step 1, it only remains to be shown that the mechanism $\mathcal{M}^* = (\phi^*, p^*)$ satisfies condition ICR$_2$. To see why this is true, fix some type $\tau = (\theta, c)$ that bears $k^*(\tau)$ failures, and consider his payoff from misreporting some type $\hat{\tau} = (\hat{\theta}, \hat{c})$ that bears a different number of failures $k^*(\hat{\tau}) \neq k^*(\tau)$. It follows from the thresholds in the Virtual Gittins formula (6) that for any type $\tau = (\theta, c)$ either (i) there exists some $\tilde{c}$ such that $k^*(\hat{\tau}) = k^*(\theta, \tilde{c})$, or (ii) there exists some $\tilde{\theta}$ such that $k^*(\hat{\tau}) = k^*(\tilde{\theta}, c)$ (or both). Assume for a moment that (i) is true. Then

$$\begin{aligned} U(\theta, c|\mathcal{M}^*) &\geq \tilde{P}^*(\theta, \tilde{c}; \theta) - \Phi_{k^*(\hat{\tau})}(\theta) \cdot c \\ &= P^*(\theta, \tilde{c}) - \Phi_{k^*(\hat{\tau})}(\theta) \cdot c \\ &\geq \tilde{P}^*(\hat{\theta}, \tilde{c}; \theta) - \Phi_{k^*(\hat{\tau})}(\theta) \cdot c \\ &= \tilde{P}^*(\hat{\tau}; \theta) - \Phi_{k^*(\hat{\tau})}(\theta) \cdot c, \end{aligned}$$

where the inequality in the first line follows from ICS, the equality from the first to the second line follows from the invariance of $P^*(\cdot)$ with respect to c within the support $\{\check{\tau} : k^*(\hat{\tau}) = k^*(\check{\tau})\}$, the inequality from the second to the third line follows from ICR$_1$, and the equality from the third to the fourth line follows from the invariance of $P^*(\cdot)$ with respect to c within the support $\{\check{\tau} : k^*(\hat{\tau}) = k^*(\check{\tau})\}$. The case where (ii) is true is analogous (we just invoke ICS and ICR$_1$ in the opposite order). This completes Step 1.

Step 2 *If there is no menu of linear contracts that implements ϕ^i, then ϕ^i is not implementable.*

The proof of this claim establishes that for any payment rule p, there exists a menu of linear contracts that leads to the same expected payments as p. The details appear in the appendix. Q.E.D.

As highlighted in the proof above, linear contracts make the entrepreneur's remuneration contingent on the number of successes observed. By doing so, these contracts satisfy an increasing differences property between the entrepreneur's success bonus and the quality of the project. This property enables the investor to tailor the initial payments $a^*(\tau)$ and the success bonuses $b^*(\tau)$ so as to screen the entrepreneur's private information about the quality of the project.

More remarkably, Proposition 1 shows that linear contracts elicit θ leaving no "further" informational rents to the entrepreneur (on top of the informational rents associated with the payoff of the safe arm c). This result relies on two facts. First, the entrepreneur's expertise is irrelevant if and only if the investor is able to design an incentive-compatible menu of linear contracts that generates the same expected payments $P^i(\tau)$ as the i-optimal mechanism. Second, by virtue of the increasing differences property in $(\theta, b^*(\tau))$, incentive compatibility requires $b^*(\tau)$ to be weakly increasing in θ across any two types τ and $\hat{\tau}$ that bear the same number of failures. As a consequence, expertise is irrelevant if and only if the expected payments $P^i(\tau)$ can be generated by a menu of linear contracts with success bonuses that are monotonic in θ.

The key to verify this monotonicity property is condition C, which requires that the inverse of the semi-elasticity $\sigma(c)$ to be weakly concave. This condition implies that the *marginal informational rents* associated with entrepreneur's opportunity costs are decreasing in c. Then, as the quality of the project increases, the investor is willing to add a positive and increasing number of entrepreneurs (with high opportunity costs) to the pool of active types, which makes the expected payments $P^i(\tau)$ an increasing and convex function of the project quality θ. Because convex functions have increasing increments, the convexity of $P^i(\tau)$ is both a necessary and sufficient condition for the existence a menu of linear contracts which success bonuses are weakly increasing. Therefore, condition C makes the problem of screening the risky arm "compatible" with the i-optimal solution, resulting on the irrelevance of the entrepreneur's expertise.

Under condition C, the optimal mechanism displays an important robustness property: It is invariant to the distribution of qualities $G(\theta)$. The next example illustrates these properties for an important family of distributions.

Example 1 (Optimal Mechanisms with Power Function Distributions) *Let the distribution of the entrepreneur's opportunity costs have the power form:*

$$H(c) = \left(\frac{c - \underline{c}}{\bar{c} - \underline{c}}\right)^\eta, \qquad \textit{where} \qquad \eta > 0.$$

For these distributions, $\frac{1}{\sigma(c)} = \frac{c - \underline{c}}{\eta}$*, condition C holds, and the entrepreneur's expertise is irrelevant. The optimal action plan is implementable by a menu of linear contracts with initial payments and success bonuses:*

$$a^*(\tau) = \frac{\Phi_{k^*(\tau)}(\theta)}{\eta + 1} \cdot (\underline{c} - \eta \cdot K) \qquad \textit{and} \qquad b^*(\tau) = \triangle \cdot \frac{\eta}{\eta + 1}.$$

The optimal payment rule under power function distributions features a menu of linear contracts with identical success bonuses *and initial payments that increase as the entrepreneur agrees to bear a higher number of failures before terminating the project.*

The parameter η measures the "elasticity of experimentation net of the lowest opportunity cost" in the sense that it represents the percent change in the mass of entrepreneurs that engage in experimentation as the rents promised by the investor (net of the lowest opportunity cost) increases by one percent: $\frac{h(c)(c-\underline{c})}{H(c)}$.[6] *The optimal success bonus offered to all types is then increasing in the elasticity parameter η. In the perfectly inelastic case, where $\eta \to 0$, the investor gets the entire cash flow from*

[6]This notion of elasticity obviously coincides with the standard experimentation elasticity, $\frac{h(c)c}{H(c)}$, when $\underline{c} \to 0$.

the project (and the entrepreneur's payment equals the initial payment $a^*(\tau) > 0$). In the perfectly elastic case, where $\eta \to +\infty$, the entrepreneur gets the entire cash flow from the project (and receives a negative initial payment $a^*(\tau) < 0$). Figure 2 describes the optimal action plan $\phi^* = \phi^i$ when $[\underline{c}, \bar{c}] = [1, 2]$ and $\eta = 1$ (uniform distribution) under the following parameterization: $T = 3$, $\lambda = \frac{1}{2}$, $K = 1$, $\triangle = 6$ and $\delta \approx 1$.

By comparing Figures 1 and 2 (which describe the first-best and the optimal mechanisms in the uniform case), one can easily notice that (i) relative to the first-best, the optimal mechanism induces a strictly smaller set of types to play the risky arm, and (ii) under the first-best, tolerance for failures is weakly higher for all active types. The next corollary to Proposition 1 shows that these properties hold beyond the uniform case.

Corollary 1 (Distortions) *Assume that condition C holds. Relative to the first-best, the optimal action plan features:*

1. ***Under-Experimentation:*** $\{\tau : k^*(\tau) \geq 1\} \subset \{\tau : k^{FB}(\tau) \geq 1\}$,
2. ***Early Termination:*** $k^*(\tau) \leq k^{FB}(\tau)$ *for all* τ *such that* $k^*(\tau) \geq 1$.

Both under-experimentation and early termination are natural consequences of the asymmetric information between the investor and entrepreneur: As the later is entitled to informational rents, the investor is less willing to undertake the risky project (and, when undertaking it, the tolerance for failures is reduced) relative to the first-best benchmark.

Limited Liability and Equity-Based Contracts

Entrepreneurs often face severe financial constraints, especially at early stages. In order to incorporate this important feature of entrepreneurial financing, we introduce in this section the requirement that the contract offered by the investor must assign non-negative total payments in any final output history. We refer to his requirement as the *limited liability* condition (LL):

$$p(\tau, x^T) \geq 0 \quad \forall \tau \in \Gamma \;\; \textit{and} \;\; \forall x^T \in X^T. \tag{LL}$$

Throughout this subsection, we say that a mechanism is *optimal* if it maximizes the investor's profits subject to feasibility *and* the limited liability constraint LL. The next definition presents an important class of linear contracts that satisfy condition LL.

Definition 4 (Equity-Based Contracts) *A payment rule p takes the form of a menu of equity-based contracts $\{a(\tau), b(\tau)\}_{\tau \in \Gamma}$ if for every type $\tau = (\theta, c) \in \Gamma$, the payments received by the entrepreneur take the form*

$$p(\tau, x^T) = \alpha(\tau) + \sum_{\{t : x_t = s\}} \delta^{t-1} \cdot \beta(\tau) \cdot \triangle,$$

where $\alpha(\tau) \geq 0$ and $\beta(\tau) \in (0,1)$. In other words, equity-based contracts specify (i) an initial transfer (seed money) $a(\tau) \geq 0$ paid by the investor to the entrepreneur; and (ii) a revenue (or equity) share $\beta(\tau) \in (0,1)$ that determines how the investor and the entrepreneur share revenues accrued to the firm.

Equity-based contracts are remarkably simple and bear close resemblance to the actual practice of venture capital financing (see, for example, Kaplan and Stromberg (2002)). They are a subclass of the class of linear contracts, with the additional requirement that initial payments have to be positive and success bonuses are a fraction between 0 and 1 of accrued revenues.

The analysis that follows extends Proposition 1 to settings with limited liability on the part of the entrepreneur. First, the next lemma derives an (easily verifiable) necessary and sufficient condition under which the i-optimal action plan ϕ^i is optimal (that is, maximizes the investor's profit subject to IC and LL).

Lemma 5 (Verifying Limited Liability) *Assume that Condition C holds. The i-optimal action plan ϕ^i is optimal if and only if $\frac{P^i(\theta,c)}{\theta}$ is weakly decreasing in θ for every c in the relevant cost range. This condition is equivalent to $P^i(1,c) \geq \frac{\partial}{\partial\theta}\left[P^i(1,c)\right]$ for every $c \in [\underline{c}, c^{max}]$.*

The property that $\frac{P^i(\theta,c)}{\theta}$ is weakly decreasing in θ for every $c \in [\underline{c}, c^{max}]$ is equivalent to saying that the elasticity of the total payment $P^i(\tau)$ with respect to the project quality is less than one. That is, a one percentage point increase in the quality of the project leads to an increase in payments of less than one percentage point. It turns out that by requiring payments per quality to be decreasing, this property guarantees that the entrepreneur receives non-negative seed money at the optimum.

The next proposition derives the unique menu of equity contracts that implements the i-optimal action plan when limited liability holds.

Proposition 2 (Limited Liability) *Assume that Condition C holds. When the limited liability constraint is not binding at the optimum, the i-optimal action plan ϕ^i is uniquely implementable by a menu of equity-based contracts with equity shares $\beta^*(\tau) = \frac{b^*(\tau)}{\triangle}$ and initial transfers $\alpha^*(\tau) = a^*(\tau)$, where the linear menu $\{a^*(\tau), b^*(\tau)\}_{\tau\in\Gamma}$ is described in (9).*

Although not directly expressed in terms of primitives, the slackness of the limited liability constraint can be easily verified for the i-optimal mechanism in many examples of interest.

Example 2 (Equity-Based Contracts with Power Function Distributions) *Consider the case of Power Function distributions as described in Example 1. It can be easily checked that the limited liability constraint is not binding (or equivalently, that $\frac{P^i(\theta,c)}{\theta}$ is weakly decreasing in θ) if and only if*

$$\underline{c} \geq \eta \cdot K. \tag{14}$$

In words, the limited liability constraint is slack provided that the project's operational costs are sufficiently small, or the lowest opportunity cost of the entrepreneur is sufficiently high. In this case, the optimal action plan is implementable by a menu of equity-based contracts with equity shares $\beta^*(\tau) = \frac{\eta}{\eta+1}$ *and seed money* $\alpha^*(\tau) = a^*(\tau)$.

Equity-based contracts with power distributions can be interpreted as offering a constant equity share and an amount of seed money that is increasing in the contract's tolerance for failures. As in Example 1, the equity share is increasing in the elasticity of experimentation (adjusted for the lower opportunity cost). In particular, when the opportunity cost is uniformly distributed, the optimal contract features fifty-fifty revenue sharing: $\beta^*(\tau) = \frac{1}{2}$.

The results above identify necessary and sufficient conditions for the entrepreneur's expertise to be irrelevant, and characterize the optimal mechanisms in these cases. The next section relaxes these conditions and analyzes the optimal mechanism when the entrepreneur's expertise is not irrelevant.

5 Contracting When Expertise is Relevant

In Section 4, we characterized the optimal mechanism when the inverse of the semi-elasticity of experimentation is weakly concave (condition C), in which case $\phi^{\hat{i}}$ is implementable. In Section 4, we introduced limited liability and provided a necessary and sufficient condition for it to be slack at the optimum. In both cases, the entrepreneur's expertise is irrelevant and the investor can implement the i-optimal action plan through of a menu of linear contracts (in the latter case, equity-based contracts).

This section considers environments where these conditions fail, so that the entrepreneur's expertise about the quality of the project affects the optimal mechanism. Subsection 5.1 characterizes the optimal mechanism when the inverse of the semi-elasticity of experimentation is not concave in the relevant range, while subsection 5.2 presents the optimal mechanism when the entrepreneur's limited liability constraint is binding at the optimum.

A complete analysis of the optimal dynamic mechanism for arbitrary distributions and utility functions that are not quasi-linear (as in the case where LL binds) is beyond the scope of this paper. In this section, we restrict attention to projects that last for a single period ($T = 1$). This single-period assumption adds tractability to the model and highlights the trade-offs that arise when condition C or quasi-linearity of preferences no longer hold. We also restrict attention to deterministic action plans throughout this section: $\phi_t : \Gamma \times X \to \{0, 1\}$. Deterministic action plans rule out stochastic policies regarding whether to undertake the project or not.

5.1 When Concavity Fails

We study two specific forms of non-concavity of the inverse of the semi-elasticity of experimentation. Other forms can be treated in a similar fashion. Condition CC requires that $\frac{1}{\sigma(c)}$ is first locally concave and then locally convex:

Condition 2 (CC) *Concavity-Convexity: There exists* $\hat{c} \in (\underline{c}, c^{max})$ *such that the inverse of the semi-elasticity of experimentation,* $\frac{1}{\sigma(c)}$*, is concave for* $c \in [\underline{c}, \hat{c}]$*, and strictly convex for* $c > \hat{c}$*.*

In turn, condition CX requires that $\frac{1}{\sigma(c)}$ is strictly convex in the relevant range.

Condition 3 (CX) *Convexity: The inverse of the semi-elasticity of experimentation,* $\frac{1}{\sigma(c)}$*, is strictly convex at every* $c \in [\underline{c}, c^{max}]$*.*

The next proposition characterizes the optimal mechanism under each of these two conditions. In both cases, the optimal mechanism can be implemented by a menu of linear contracts. However, due to the non-concavity of $\frac{1}{\sigma(c)}$, the i-optimal action plan ϕ^i is no longer implementable. In order to restore incentive compatibility, the optimal mechanism is an "ironed version" (in the sense of Myerson (1981)) of ϕ^i. As a result, the optimal mechanism $\hat{\mathcal{M}} = (\hat{\phi}, \hat{p})$ features "bunching," i.e., all types in some interval receive the same contract.

Proposition 3 (Optimal Mechanism with Non-Concave σ) *Consider the static model ($T = 1$). Then,*

1. *If condition CC holds, the optimal action plan is implementable by a menu of linear contracts* $\hat{p} = \left\{\hat{a}(\tau), \hat{b}(\tau)\right\}_{\tau \in \Gamma}$ *that features bunching at the top: There exists* $\hat{\theta}_0 \in (0,1)$ *such that*

$$(\hat{a}(\tau), \hat{b}(\tau)) = (a^*(\tau), b^*(\tau)) \quad if \quad \theta \leq \hat{\theta}_0, \qquad and \qquad (\hat{a}(\tau), \hat{b}(\tau)) = (\hat{a}, \hat{b}) \quad if \quad \theta > \hat{\theta}_0.$$

2. *If condition CX holds, the optimal action plan is implementable by a menu of linear contracts* $\hat{p} = \left\{\hat{a}(\tau), \hat{b}(\tau)\right\}_{\tau \in \Gamma}$ *that exhibits full bunching:*

$$(\hat{a}(\tau), \hat{b}(\tau)) = (\hat{a}, \hat{b}) \qquad for \;\; all \qquad \tau \in \Gamma.$$

Figure 3 illustrates the results form Proposition 3. We sketch its proof below. The formal derivation appears in the appendix.

Proof of Proposition 3 (Sketch). Consider a deterministic action plan $\phi : (\theta, c) \mapsto \{0, 1\}$ and a payment rule with expected payments $P(\theta, c)$. Incentive compatibility on the safe arm (ICS) requires that $\phi(\theta, c)$ is weakly decreasing in c for every θ, in which case there exists a threshold

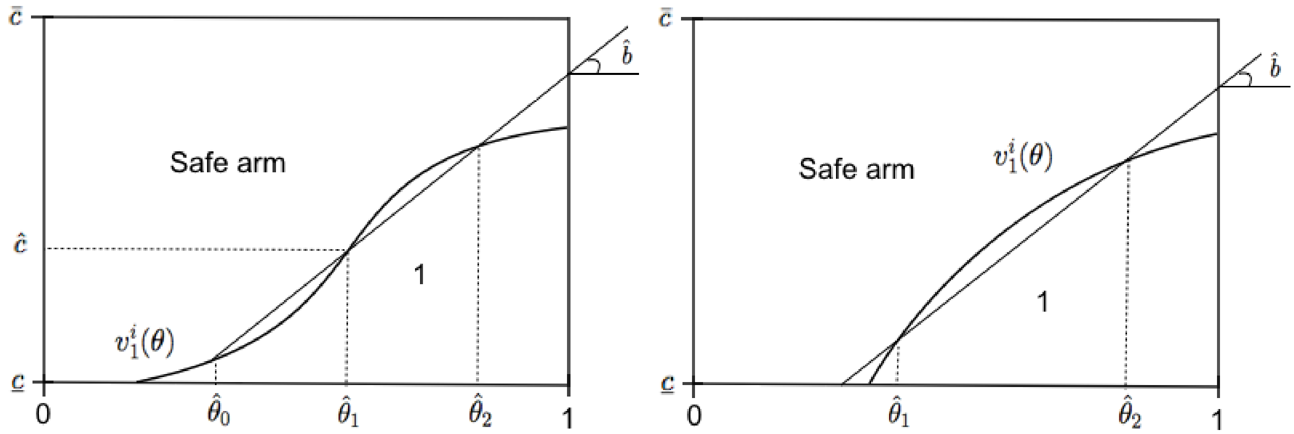

Figure 3: The optimal action plans when the inverse semi-elasticity is concave-convex (left-side panel) and convex (right-side panel).

function $v(\theta)$ such that the entrepreneur undertakes the project if and only if $c \leq v(\theta)$. Moreover, IR implies that $P(\theta, c) = v(\theta)$, in which case $U(\theta, c|\mathcal{M}) = v(\theta) - c$ for all (θ, c) such that $c \leq v(\theta)$.

In turn, incentive compatibility on the risky arm (ICR_1) implies that a menu of linear contracts satisfies ICR_1 if and only if $b(\tau)$ is weakly increasing in θ, and that, by the envelope theorem, $b(\tau) = \frac{\partial U(\theta, c|\mathcal{M})}{\partial \theta} = v'(\theta)$. As a consequence, a deterministic mechanism is feasible if and only if its threshold function $v(\theta)$ is convex.

Consider first the case where condition CX holds. Because $\frac{1}{\sigma(c)}$ is strictly convex, the i-optimal action plan $v_1^i(\theta)$ derived in equation (13) is strictly concave, and therefore not implementable. The investor program therefore consists in finding the best convex function $\hat{v}_1(\theta)$ that "irons" $v_1^i(\theta)$. As the right-side panel of Figure 3 illustrates, this is accomplished by a linear curve. Because $\hat{v}_1(\theta)$ is linear, the optimal revenue share schedule $\hat{b}(\tau) = \hat{v}'(\theta)$ is constant, and the resulting menu of venture capital contracts exhibits full bunching.

Consider now the case where condition CC holds. Denote $\hat{\theta}_1 \equiv \left(v_1^i\right)^{-1}(\hat{c})$. Because $\frac{1}{\sigma(c)}$ is weakly concave for $c \in [\underline{c}, \hat{c}]$ and strictly convex for $c \in (\hat{c}, \bar{c}]$, the i-optimal action plan $v_1^i(\theta)$ derived in equation (13) is weakly convex for $\theta \in [0, \hat{\theta}_1]$ and strictly concave for $\theta \in (\hat{\theta}_1, 1]$. As before, the investor program consists in finding the best convex function $\hat{v}_1(\theta)$ that "irons" $v_1^i(\theta)$. As the left-side panel of Figure 3 illustrates, this is accomplished by following the i-optimal action plan "at the bottom", i.e., $\hat{v}_1(\theta) = v_1^i(\theta)$ for all $\theta \leq \hat{\theta}_0 < \hat{\theta}_1$, and linearizing the threshold function "at the top", i.e., $\hat{v}_1(\theta) = \hat{a} + \hat{b} \cdot \lambda \cdot \theta \cdot \triangle$ for all $\theta > \hat{\theta}_0$. The linearity of $\hat{v}_1(\theta)$ at the top generates bunching in the menu of venture capital contracts for all $\theta > \hat{\theta}_0$. In particular, notice that the bunching region starts before $\hat{\theta}_1$ (which is the inflection point of $v_1^i(\theta)$). Q.E.D.

As the proof sketch above makes clear, the non-concavity of $\frac{1}{\sigma(c)}$ makes the incentive compatibility constraint associated with the risky arm binding. In order to satisfy this constraint, the investor has to modify the menu of linear contracts relative to the one from Proposition 1. As a result, new "informational rents" arise, now associated with the ex-ante quality of the project θ. Interestingly, and unlike one-dimensional adverse selection models, these rents are non-monotone and can even be negative. This is subject of our next corollary.

Corollary 2 (When Expertise Hurts due to Non-Concavity) *Consider the static model (*$T = 1$*) and assume that either condition CC or CX holds. Let* $\hat{P}(\theta, c)$ *denote the expected payments associated with the optimal mechanism. There exist* $0 < \hat{\theta}_1 < \hat{\theta}_2 < 1$ *such that*

$$\begin{aligned} \hat{P}(\theta, c) &\geq P^i(\theta, c) \qquad if \qquad \theta \in [0, \hat{\theta}_1] \cup [\hat{\theta}_2, 1], \\ \hat{P}(\theta, c) &< P^i(\theta, c) \qquad if \qquad \theta \in (\hat{\theta}_1, \hat{\theta}_2). \end{aligned}$$

Because $\hat{P}(\theta, c) = \hat{v}(\theta)$ for all active types (i.e., those types for which $c \leq \hat{v}(\theta)$), Figure 3 represents the schedule of expected payments (as a function of θ) of the i-optimal mechanism $\mathcal{M}^i$ and of the optimal mechanism (under two-dimensional private information) $\hat{\mathcal{M}}$. Whether condition CC or CX holds, (i) entrepreneurs with projects of intermediate quality (i.e., $\theta \in (\hat{\theta}_1, \hat{\theta}_2)$) receive *negative* informational rents from holding private information on θ (expertise), while entrepreneurs with projects of low or high quality (i.e., $\theta \in [0, \hat{\theta}_1] \cup [\hat{\theta}_2, 1]$) receive (weakly) positive informational rents from their expertise on θ. That is, entrepreneurs with projects of intermediate quality receive strictly lower rents when the ex-ante quality of the project is their private information than when it is commonly known. Thus, unlike in one-dimensional screening models, private information can make the agent strictly worse (even when the dimensions of private information are uncorrelated, as c and θ are here). Interestingly, Corollary 2 implies that entrepreneurs with projects of intermediate quality would benefit from being able to provide hard evidence regarding project quality.

5.2 When Limited Liability Binds

Limited liability prevents the investor from charging strictly positive payments back from the entrepreneur in any terminal outcome history. As established in Proposition 5, whenever $\frac{P^i(\theta,c)}{\theta}$ fails to be weakly decreasing in θ, there is no menu of equity-based contracts satisfying LL that implements ϕ^i. As a consequence, the investor's choice of which mechanism to offer is affected by the entrepreneur's expertise.

The next proposition characterizes the optimal mechanism when LL binds and condition C holds. As before, the optimal mechanism can be implemented by a menu of equity-based contracts. Recall that, under condition C, the i-optimal expected payment $P^i(\tau)$ is convex in θ. This implies that the menu p^* from Proposition 1 is such that the output-contingent part of the entrepreneur's expected remuneration, $b^*(\tau) \cdot \triangle \cdot \lambda \cdot \theta$, increases faster on θ than its fixed component, $a^*(\theta)$. As a consequence,

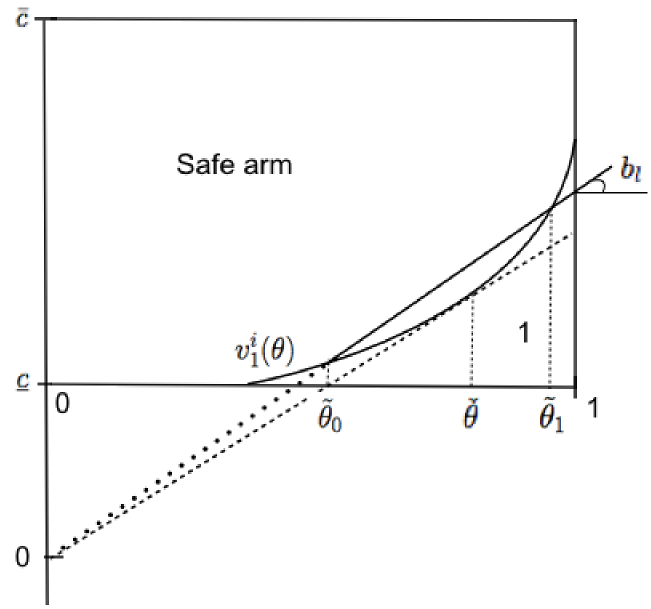

Figure 4: The optimal action plan when limited liability binds.

p^* violates LL only for projects with high enough quality θ. In order to restore LL "at the top", the optimal mechanism is an "ironed version" of ϕ^i. As a result, the optimal menu of venture capital contracts assigns the same contract to all entrepreneurs with a project of high enough quality.

Proposition 4 (Optimal Mechanism when LL binds) *Consider the static model ($T = 1$) and assume that condition C holds. The optimal action plan is implemented by a menu of equity-based contracts $p_l = \{a_l(\tau), b_l(\tau)\}_{\tau\in\Gamma}$. It exhibits bunching at the top if and only if $\mathcal{M}^* = (\phi^*, p^*)$ violates limited liability. In the latter case, there exists $\tilde{\theta}_0 \in [0, 1)$ such that*

$$(a_l(\tau), b_l(\tau)) = (a^*(\tau), b^*(\tau)) \quad if \quad \theta \leq \tilde{\theta}_0, \quad and \quad (a_l(\tau), b_l(\tau)) = (a_l, b_l) \quad if \quad \theta > \tilde{\theta}_0.$$

Moreover, if $\underline{c}$ is sufficiently low, the optimal menu of venture capital contracts exhibits full bunching, i.e.,

$$(a_l(\tau), b_l(\tau)) = (a_l, b_l) \qquad for \;\; all \qquad \tau \in \Gamma.$$

Figure 4 illustrates the results form Proposition 4. The optimal mechanism is an ironed version of ϕ^i designed to satisfy LL for all projects of high enough θ. In particular, the optimal bunching region starts before $\check{\theta}$, the threshold quality above which p^* violates LL. As $\underline{c}$ decreases, the bunching area expands, and, for $\underline{c}$ low enough, the optimal mechanism is such that all types are offered the same contract (full bunching).

The next corollary investigates the informational rents that the entrepreneur's expertise generates when the LL constraint binds. As in Corollary 2, these rents are non-monotone in θ, and are actually negative for high enough θ.

Corollary 3 (When Expertise Hurts Because LL Binds) *Consider the static model ($T = 1$), and assume that condition C holds and that $\mathcal{M}^* = (\phi^*, p^*)$ violates limited liability. Denoting by $P_l(\theta, c)$ the expected payments associated with the optimal mechanism, there exists $0 < \tilde{\theta}_1 < 1$ such that*

$$P_l(\theta, c) \geq P^i(\theta, c) \qquad if \qquad \theta \in [0, \tilde{\theta}_1],$$
$$P_l(\theta, c) < P^i(\theta, c) \qquad if \qquad \theta \in (\tilde{\theta}_1, 1].$$

Intuitively, in order to satisfy LL, the optimal menu of equity-based contracts p_l has to decrease the revenue share of high quality projects, and increase their seed money, relative to p^*. In order to do so in an incentive-compatible way, p_l increases the revenue shares of intermediate types, therefore expanding the mass of entrepreneurs with intermediate quality who undertake the risky project. In order to reduce the costs of "over-experimentation" for intermediate types, the investor reduces the mass of entrepreneurs with high quality projects who try the risky arm. For this reason, expertise in quality actually hurts the entrepreneurs with high enough θ, leading to negative informational rents from expertise.

6 Conclusion

This paper studies optimal contracting in settings that combine experimentation and asymmetric information. To understand the implications of adverse selection for experimentation, we build a model that embeds a geometric two-arm bandit problem into a principal-agent framework. The agent has private information about his payoff from playing the safe arm (which we interpret as his outside option, or opportunity cost of experimentation), as well as about the stochastic process that governs the risky arm. We construct a two-step procedure to solve for the principal's profit-maximizing mechanism: In the first stage, we consider a relaxed (one-dimensional screening) problem in which private information is confined to opportunity costs. Having solved this relaxed program, we derive a necessary and sufficient condition for the solution of this program to be implementable in the setting with two-dimensional private information. When it is not implementable, the optimal mechanism consists of an ironed version of the solution of the relaxed program.

The use of output-contingent contracts is essential to screen the agent's private information about the risky arm. The principal offers contracts with different tolerance levels in order to screen the agent's opportunity costs, and offers different "powers" to screen the agent's information about the quality of the risky arm. The optimal mechanism can be implemented via a menu of linear contracts and, when the agent's limited liability constraints do not bind, via a menu of equity-based contracts.

Throughout the analysis, we favored the venture-capital-financing interpretation of our model. In this interpretation, the principal is an investor and the agent is an entrepreneur who is privately informed about both the project's prospects and his outside options. The mechanism that maximizes the investor's payoff can be implemented as a menu of contracts that feature (i) an initial payment (seed money) to the entrepreneur in exchange for an equity share, and (ii) a termination clause that specifies the maximum number of failures before the project is aborted.

More broadly, our model can be applied to diverse situations where the principal wants to select which agents shall engage in innovative activities. Consider the problem of designing contracts for research and development (R&D). In this interpretation, the agent conducting research (who may be an employee of an organization or a different organization) is better informed about the quality of the project and his opportunity cost of engaging in experimentation. In turn, the principal, who holds the property rights of possible innovations, designs contracts intended to encourage the *right* innovations at the *lowest* costs.

Another natural interpretation of our model pertains the problem of an academic recruiter who wants to select the right candidates for assistant professorships. The recruiter can offer menus of tenure-track contracts (with various tenure clocks) and different combinations of publication prizes and baseline salary. The choice of a tenure clock screens the candidate's outside options (e.g., jobs outside academia), while the choice of baseline salary/publication prize screens the candidate's assessment of the quality of his own research.

We show that the agent's expertise about the risky arm is irrelevant if and only if the inverse of the semi-elasticity of experimentation is concave. When this concavity condition fails, a positive mass of entrepreneurs actually get negative rents from their expertise. These results have puzzling implications for the incentives for information acquisition. An interesting avenue for future research is to study optimal contracting when information acquisition is endogenous.

Appendix

The proofs omitted in the main text are presented below.

Proof of Lemma 2. The knowledgeable investor chooses a mechanism $\mathcal{M} = (\phi, p)$ to maximize expected profits (2) subject to incentive compatibility on the safe arm (ICS). Note that the entrepreneur's payoff (1) satisfies strictly decreasing differences in (Φ, c), where Φ is the expected duration of the action plan. Therefore, we can use standard arguments to prove that:

Lemma 6 ***(ICS)*** *A mechanism $\mathcal{M} = (\phi, p)$ satisfies incentive compatibility on the safe arm (ICS) if and only if the following conditions jointly hold:*

1. *For every $\theta \in [0, 1]$, the expected discounted duration of the project, $\Phi(\tau)$, is weakly decreasing in c,*

2. *For every $\theta \in [0, 1]$, the expected payment of type (θ, c), $P(\tau)$, satisfies the envelope formula:*

$$P(\tau) = \Phi(\tau) \cdot c + \int_c^{\overline{c}} \Phi(\theta, \hat{c}) d\hat{c}. \tag{15}$$

Plugging the envelope formula above into the objective function (2) leads to:

$$\mathbb{E}^c \left\{ \mathbb{E}^{\lambda[\phi(\tau)|\theta]} \left[\sum_{t=1}^{T} \left(\phi_t(\tau, \tilde{x}^{t-1}) \cdot \left(\tilde{\theta}_t \cdot \triangle - K - c - \frac{1}{\sigma(c)} \right) \right) \right] \right\}. \tag{16}$$

Let us for now ignore the monotonicity constraint on $\Phi(\tau)$ and consider the relaxed problem $\mathcal{P}^r$ of choosing an action plan ϕ to maximize (16) pointwise in c:

$$\mathcal{P}^r: \qquad \max_{\phi} \quad \mathbb{E}^{\lambda[\phi(\tau)|\theta]} \left[\sum_{t=1}^{T} \left(\phi_t(\tau, \tilde{x}^{t-1}) \cdot \left(\tilde{\theta}_t \cdot \triangle - K - c - \frac{1}{\sigma(c)} \right) \right) \right].$$

The objective function above is similar to the first-best objective function, the only difference being that the opportunity cost c is replaced by its virtual counterpart, $c + \frac{1}{\sigma(c)}$. We can therefore apply Lemma 1 to conclude that the solution of problem $\mathcal{P}^r$ is the tenure action plan ϕ^i according to which an entrepreneur with type $\tau = (\theta, c)$ bears exactly $k \in \{1, \ldots, T\}$ failures if and only if $v^i_{k+1}(\theta) < c \leq v^i_k(\theta)$, where the functions $v^i_k(\theta)$ are determined by equation (6).

Consider now the types $\tau = (\theta, c)$ and $\hat{\tau} = (\theta, \hat{c})$ such that $\hat{c} > c$. Because H is log concave, $\frac{1}{\sigma(c)}$ is strictly increasing in c. Therefore, under ϕ^i, an entrepreneur with type τ bears a weakly greater number of failures than an entrepreneur with type $\hat{\tau}$. This implies that the expected discounted duration $\Phi(\tau)$ is weakly decreasing in c for every $\theta \in [0, 1]$. We can therefore conclude that the solution of the knowledgeable investor's problem employs the action plan ϕ^i. Evaluating the the envelope formula (15) at the action plan ϕ^i leads to (7). Q.E.D.

Proof of Lemma 3. Condition ICR_1 is equivalent to saying that for every $\hat{k} \in \{1, ..., T\}$

$$(a(\tau), b(\tau)) \in \arg\max_{(\hat{a},\hat{b})} \left\{\hat{a} + \hat{b} \cdot \triangle \cdot \Lambda_k(\theta) - \Phi_k(\theta) \cdot c\right\} \quad s.t. \quad (\hat{a}, \hat{b}) \in \left\{(a(\tau), b(\tau)) : k(\tau) = \hat{k}\right\}.$$

Notice that, because $\Lambda_k(\theta)$ is strictly increasing in θ, the objective above satisfy strict increasing differences in $(\hat{b}, \theta)$. Milgrom's Constraint Simplification theorem (Milgrom (2004), page 105) then implies that $(a(\tau), b(\tau))$ is a maximizing schedule if and only if the envelope condition (12) holds and $b(\tau)$ with support $\{\tau : k(\tau) = \hat{k}\}$ is weakly increasing in θ. Q.E.D.

Proof of Lemma 4. With a slight abuse of notation, let us write $\Phi_0^*(\theta) = 0$. Therefore, from equation equation (7), it follows that the expected payments $P^i(\tau)$ restricted to the support $\{\tau : k^i(\tau) = \hat{k}\}$ is weakly increasing in θ for every $\hat{k}$ provided that for all $k \in \{1, \ldots, K\}$

$$\frac{d^2}{d\theta^2} \left\{(\Phi_k(\theta) - \Phi_{k-1}(\theta)) \cdot v_k^i(\theta)\right\} \geq 0. \tag{17}$$

Consider the Virtual Gittins formula (6) at some $k \in \{1, \ldots, K\}$. Multiplying both sides by $\theta \cdot (1-\lambda)^{t-1} + (1-\theta)$ leads to

$$\theta \cdot (1-\lambda)^{k-1} \cdot \left(\lambda \cdot \triangle - K - v_k^i(\theta) - \frac{1}{\sigma\left(v_k^i(\theta)\right)}\right) - (1-\theta) \cdot \left(K + v_k^i(\theta) + \frac{1}{\sigma\left(v_k^i(\theta)\right)}\right)$$

$$= \mathbf{1}_{[k<T]} \cdot \theta \cdot (1-\lambda)^{k-1} \cdot \lambda \cdot \delta \cdot \left[\frac{1-\delta^{T-k}}{1-\delta} \cdot \left(\lambda \cdot \triangle - K - v_k^i(\theta) - \frac{1}{\sigma\left(v_k^i(\theta)\right)}\right)\right].$$

Rearranging the equation above leads to

$$\theta \cdot (1-\lambda)^{k-1} \cdot \left(\lambda \cdot \triangle - K - \mathbf{1}_{[k<T]} \cdot \lambda \cdot \delta \cdot \left[\frac{1-\delta^{T-k}}{1-\delta} \cdot (\lambda \cdot \triangle - K)\right]\right) - (1-\theta) \cdot K \tag{18}$$

$$= \left(\theta \cdot (1-\lambda)^{k-1} \cdot \left[1 + \mathbf{1}_{[k<T]} \cdot \lambda \cdot \delta \cdot \frac{1-\delta^{T-k}}{1-\delta}\right] + (1-\theta)\right) \left(v_k^i(\theta) + \frac{1}{\sigma\left(v_k^i(\theta)\right)}\right)$$

$$= (\Phi_k(\theta) - \Phi_{k-1}(\theta)) \cdot \left(v_k^i(\theta) + \frac{1}{\sigma\left(v_k^i(\theta)\right)}\right),$$

where the last equality follows from (8).

Notice that the first line of (18) in affine in θ. Moreover, the expression for $\Phi_k(\theta) - \Phi_{k-1}(\theta)$ is also affine in θ and is strictly positive for every θ. The next lemma establishes a useful mathematical fact.

Lemma 7 *Consider an interval $(\theta_1, \theta_2) \in \mathbb{R}_+$ and let $A_1, A_2, B_1,$ and B_2 be scalars such that $B_1 + B_2 \cdot \theta > 0$ for all $\theta \in (\theta_1, \theta_2)$. Let the function $\zeta : \mathbb{R}_+ \to \mathbb{R}_+$ be twice continuously differentiable with $\zeta'(x) \geq 0$ for all x. Let $x(\theta)$ be implicitly defined by*

$$A_1 + A_2 \cdot \theta = (B_1 + B_2 \cdot \theta) \cdot (x(\theta) + \zeta(x(\theta))). \tag{19}$$

Then $\frac{d^2}{d\theta^2}\left[(B_1 + B_2\theta)\, x(\theta)\right] \gtreqless 0$ if and only if $\zeta''(x(\theta)) \lesseqgtr 0$.

Proof of Lemma 7. Let us first compute the following derivatives:

$$\begin{aligned}
\frac{d}{d\theta}\left[\left(B_1+B_2\cdot\theta\right)x\left(\theta\right)\right] &= \left(B_1+B_2\cdot\theta\right)\cdot x'\left(\theta\right)+B_2\cdot x\left(\theta\right)\\
\frac{d}{d\theta}\left[\left(B_1+B_2\cdot\theta\right)\cdot\zeta\left(x\left(\theta\right)\right)\right] &= \left(B_1+B_2\cdot\theta\right)\cdot\zeta'\left(x\left(\theta\right)\right)\cdot x'\left(\theta\right)+B_2\cdot\zeta\left(x\left(\theta\right)\right)\\
\frac{d^2}{d\theta^2}\left[\left(B_1+B_2\cdot\theta\right)\cdot x\left(\theta\right)\right] &= \left(B_1+B_2\cdot\theta\right)\cdot x''\left(\theta\right)+2\cdot B_2\cdot x'\left(\theta\right)\\
\frac{d^2}{d\theta^2}\left[\left(B_1+B_2\cdot\theta\right)\cdot\zeta\left(x\left(\theta\right)\right)\right] &= \left(B_1+B_2\cdot\theta\right)\cdot\zeta''\left(x\left(\theta\right)\right)\cdot\left[x'\left(\theta\right)\right]^2\\
&\quad+\left(B_1+B_2\cdot\theta\right)\cdot\zeta'\left(x\left(\theta\right)\right)\cdot x''\left(\theta\right)+2\cdot B_2\cdot\zeta'\left(x\left(\theta\right)\right)\cdot x'\left(\theta\right).
\end{aligned}$$

From the third and fourth equalities above it follows that

$$\frac{d^2}{d\theta^2}\left[\left(B_1+B_2\cdot\theta\right)\cdot\zeta\left(x\left(\theta\right)\right)\right]=\left(B_1+B_2\cdot\theta\right)\cdot\zeta''\left(x\left(\theta\right)\right)\cdot\left[x'\left(\theta\right)\right]^2+\zeta'\left(x\left(\theta\right)\right)\cdot\frac{d^2}{d\theta^2}\left[\left(B_1+B_2\cdot\theta\right)\cdot x\left(\theta\right)\right]. \tag{20}$$

In turn, because of (19), we know that

$$\frac{d^2}{d\theta^2}\left[\left(B_1+B_2\cdot\theta\right)\cdot\left(x\left(\theta\right)+\zeta\left(x\left(\theta\right)\right)\right)\right]=0. \tag{21}$$

Putting (20) and (21) together leads to

$$\frac{d^2}{d\theta^2}\left[\left(B_1+B_2\cdot\theta\right)\cdot x\left(\theta\right)\right]=-\left[\frac{\left(B_1+B_2\cdot\theta\right)\cdot\left[x'\left(\theta\right)\right]^2}{\left(1+\zeta'\left(x\left(\theta\right)\right)\right)}\right]\cdot\zeta''\left(x\left(\theta\right)\right),$$

proving the result. Q.E.D.

From Lemma 7 it then follows that condition C, which guarantees the concavity of $\frac{1}{\sigma(\cdot)}$ in the relevant cost range, implies that (17) holds for all $k\in\{1,\ldots,K\}$. Therefore, condition C implies that statement 2 holds.

Finally, if condition C fails for some $c\in[\underline{c},c^{max}]$, it follows from Lemma 7 that $P^i(\tau)$ restricted to the support $\{\tau:k^i(\tau)=1\}$ is not weakly increasing in θ. This establishes the equivalence between statements 2 and 3. Q.E.D.

Proof of Step 2 of Proposition 1: The result follows from the following lemma.

Lemma 8 *For any deterministic mechanism (ϕ,p), where ϕ is a tenure action plan, there exists a menu of linear contracts $p_l:\Gamma\times X^T\to\mathbb{R}$ such that $P(\hat{\tau};\theta)=P_l(\hat{\tau};\theta)$ for any $\hat{\tau}\in\Gamma$ and any $\theta\in[0,1]$.*

Proof. Define the subset of terminal nodes $B\equiv\{x^T:x_t\neq s\ \forall t\in\{1,\ldots,T\}\}$. Take a payment rule $p:\Gamma\times X^T\to\mathbb{R}$ and, without any loss of generality, set $p(\hat{\tau},x^T)=0$ to all terminal nodes x^T that are not reachable with positive probability under the mechanism (ϕ,p). It follows by definition that

$$\tilde{P}(\hat{\tau};\theta)=\theta\cdot\sum_{x^T}\left\{p(\hat{\tau},x^T)\cdot\left(\lambda^{\mathbf{S}(x^T)}\cdot(1-\lambda)^{\boldsymbol{F}(x^T)}\right)\right\}+(1-\theta)\cdot\sum_{x^T\in B}\left\{p(\hat{\tau},x^T)\right\}$$

$$= \theta \cdot \left[\sum_{x^T} \left\{ p(\hat{\tau}, x^T) \cdot \left(\lambda^{\boldsymbol{S}(x^T)} \cdot (1-\lambda)^{\boldsymbol{F}(x^T)} \right) \right\} - \sum_{x^T \in B} \left\{ p(\hat{\tau}, x^T) \right\} \right] + \sum_{x^T \in B} \left\{ p(\hat{\tau}, x^T) \right\}$$

$$= \Lambda_{k(\hat{\tau})}(\theta) \cdot \left[\frac{\sum_{x^T} \left\{ p(\hat{\tau}, x^T) \cdot \left(\lambda^{\boldsymbol{S}(x^T)} \cdot (1-\lambda)^{\boldsymbol{F}(x^T)} \right) \right\} - \sum_{x^T \in B} \left\{ p(\hat{\tau}, x^T) \right\}}{\frac{d}{d\theta} \left[\Lambda_{k(\hat{\tau})}(\theta) \right]} \right] + \sum_{x^T \in B} \left\{ p(\hat{\tau}, x^T) \right\},$$

where the last equality follows from the fact that $\frac{\Lambda_{k(\hat{\tau})}(\theta)}{\frac{d}{d\theta}\left[\Lambda_{k(\hat{\tau})}(\theta)\right]} = \theta$ for any tenure action plan ϕ (with tolerance for failures described by $k(\cdot)$).

Now consider the menu of linear contracts $p_l : \Gamma \times X^T \to \mathbb{R}$ with initial payments $a(\hat{\tau})$ and success bonuses $b(\hat{\tau})$. By definition,

$$\tilde{P}_l(\hat{\tau}; \theta) = a(\hat{\tau}) + \Lambda_{k(\hat{\tau})}(\theta) \cdot b(\hat{\tau}).$$

Now set

$$a(\hat{\tau}) = \sum_{x^T \in B} \left\{ p(\hat{\tau}, x^T) \right\}$$

and

$$b(\hat{\tau}) = \frac{\sum_{x^T} \left\{ p(\hat{\tau}, x^T) \cdot \left(\lambda^{\boldsymbol{S}(x^T)} \cdot (1-\lambda)^{\boldsymbol{F}(x^T)} \right) \right\} - \sum_{x^T \in B} \left\{ p(\hat{\tau}, x^T) \right\}}{\frac{d}{d\theta} \left[\Lambda_{k(\hat{\tau})}(\theta) \right]}.$$

Note from equation (10) that $\frac{d}{d\theta}\left[\Lambda_{k(\hat{\tau})}(\theta)\right]$, for each tolerance level $k(\hat{\tau})$, is a constant that does not depend on θ. This implies that the payment rule p_l is well-defined. By construction, $\tilde{P}(\hat{\tau}; \theta) = \tilde{P}_l(\hat{\tau}; \theta)$ for any $\hat{\tau} \in \Gamma$ and any $\theta \in [0, 1]$, as we wanted to show. Q.E.D.

By virtue of Lemma 8, a deterministic mechanism (ϕ, p), where ϕ is a tenure action plan, is feasible if and only if there exists a deterministic mechanism (ϕ, p_l), where p_l is a menu of linear contracts, that is also feasible. This concludes the proof of Step 2. Q.E.D.

Proof of Corollary 1. The function $V_k(\theta, c)$, defined in (4), is strictly decreasing in c for every $k \in \{1, \ldots, K\}$. Therefore

$$0 = V_k \left(\theta, v_k^i(\theta) + \frac{1}{\sigma\left(v_k^i(\theta)\right)} \right) < V_k\left(\theta, v_k^i(\theta)\right),$$

what implies that $v_k^{FB}(\theta) > v_k^i(\theta)$ for all $k \in \{1, \ldots, K\}$. The result then follows from Proposition 1. Q.E.D.

Proof of Lemma 5. The menu $\{a^*(\tau), b^*(\tau)\}_{\tau \in \Gamma}$ satisfies condition LL if and only if $a^*(\tau) \geq 0$ for every $\tau \in \Gamma$. After substituting the formula for $b^*(\tau)$ into the formula for $a^*(\tau)$, we obtain that $a^*(\tau) \geq 0$ if and only if

$$P^i(\tau) - \frac{\partial}{\partial\theta}\left[P^i(\tau)\right] \cdot \theta \geq 0, \tag{22}$$

which is equivalent to property that $\frac{P^i(\theta,c)}{\theta}$ is weakly decreasing (after applying the quotient rule for differentiation). As established by Lemma 4, $P^i(\tau)$, restricted to the support $\{\tau : k^i(\tau) = \hat{k}\}$, is weakly convex in θ for any $\hat{k} \in \{1, ..., T\}$. It immediately follows from convexity that $P^i(1, c) \geq \frac{\partial}{\partial\theta}\left[P^i(1, c)\right]$ for every $c \in [\underline{c}, c^{max}]$ is necessary and sufficient for (22) to hold for every $\theta \in [0, 1]$. Q.E.D.

Proof of Proposition 2. Under condition C and when limited liability is slack, the mechanism with action plan ϕ^* and payments $\{a^*(\tau), b^*(\tau)\}_{\tau\in\Gamma}$ is optimal. Moreover, by construction,

$$a^*(\tau) + b^*(\tau) \cdot \Lambda_{k^i(\tau)}(\theta) = \alpha^*(\tau) + \beta^*(\tau) \cdot \triangle \cdot \Lambda_{k^i(\tau)}(\theta).$$

Now note that

$$0 < \underline{c} \leq a^*(\tau) + b^*(\tau) \cdot \Lambda_{k^i(\tau)}(\theta) = P^i(\tau) < \triangle \cdot \Lambda_{k^i(\tau)}(\theta),$$

where the last inequality follows from the i-optimality of ϕ^i. Because $a^*(\tau) \geq 0$, we can conclude that $b^*(\tau) < \triangle$, or equivalently, $\beta^*(\tau) \in (0, 1)$. Q.E.D.

Proofs of Section 5: Propositions 3 and 4

Notation: *For any line $A+B\theta$ we write $P(A + B\theta)$ for the projection of the line $A+B\theta$ into $[\underline{c}, \bar{c}]$. Formally,*

$$P(A + B\theta) := \begin{cases} A + B\theta \text{ if } A + B\theta \in [\underline{c}, \bar{c}] \\ \underline{c} \text{ if } A + B\theta < \underline{c} \\ \bar{c} \text{ if } A + B\theta > \bar{c}. \end{cases}$$

Lemma 9 *There exists an optimal deterministic contract under limited liability.*

Proof of Lemma 9. Let (a, b) be any implementable contract subject to the limited liability constraint. Notice that (a, b) induces a unique experimentation curve $v : \Theta \to [\underline{c}, \bar{c}]$, according to which type $\tau = (\theta, c)$ experiments if and only if $c \leq v(\theta)$. First, notice that any curve v is clearly uniformly bounded by 1. Second, it is straightforward to check that the limited liability constraint implies an upper bound K_1 on $|b|$. For each θ let $\partial v(\theta)$ be the set of sub-gradients of v at θ. Clearly $\sup |\partial v(\theta)| \leq \sup |b| \leq K_1$. This enables us to work with a bounded subspace[7] of $C[0, 1]$, consisting of (uniformly bounded by 1) convex functions with Lipschitz constant K_1. From Arzela-Ascoli Theorem this space is sequentially compact. Since the subspace of convex functions is closed and the principal's profit is continuous the Lemma follows. Q.E.D.

For the remainder of this subsection, consider the optimal contract $(a^*(\theta), b^*(\theta))$ when the limited liability is ignored. Accordingly, $v^*(\theta)$ denotes the optimal experimentation curve. If $v^*(\theta) = \underline{c}$ then define $(a^*(\theta), b^*(\theta)) = 0$.

[7] Endow $C[0, 1]$ with the sup-norm.

Lemma 10 *Let v_1 and v_2 two implementable curves. If for every θ we have either $v_2(\theta) \leq v_1(\theta) \leq v^*(\theta)$ or $v_2(\theta) \geq v_1(\theta) \geq v^*(\theta)$, with a strict inequality for a subset Θ' with positive measure, then the profit from a contract leading to the curve v_1 is strictly higher than the profit from the curve v_2.*

Proof of Lemma 10. Fix $\theta \in \Theta$ and consider a cutoff $v(\theta)$ so that type (θ, c) experiments if and only if $c \leq v(\theta)$. In the proof of Proposition 1 it was shown that the principal's profit is (locally) strictly increasing in the cutoff $v(\theta)$ whenever $v(\theta) < v^*(\theta)$ and it is strictly decreasing if $v(\theta) > v^*(\theta)$. Q.E.D.

6.1 Limited Liability

Let $\left(a^L(\theta), b^L(\theta)\right)$ be an optimal deterministic contract under limited liability.

Proposition 4 follows from Lemma 11 below.

Lemma 11 *Assume that there exists θ_0 with $v^*(\theta_0) > 0$ such that $a^*(\theta) < 0$ for $\theta > \theta_0$ and $a^*(\theta) > 0$ for $\theta < \theta_0$.. There exists an optimal deterministic contract under limited liability $\left(a^L(\theta), b^L(\theta)\right)$. There exists $\theta^* < \theta_0$ such that $\left(a^L(\theta), b^L(\theta)\right) = (a^*(\theta), b^*(\theta))$ for all $\theta < \theta^*$. Furthermore $\left(a^L(\theta), b^L(\theta)\right)$ is constant for all $\theta > \theta^*$ and it is is discontinuous at θ^*.*

Proof of Lemma 11. Let v^L be the optimal experimentation curve under limited liability and $\left(a^L(\theta), b^L(\theta)\right)$ be an induced optimal contract[8]. Let $\hat{\Theta} \subset \Theta$ be the set of points such that v^L is differentiable.

The bulk of this proof of this Lemma is by contradiction. Lemma 10 will be essential. In some steps below, we will assume towards a contradiction that v^L has some properties which are inconsistent with the statement of the Lemma. We will then construct another curve v^{**} which is closer (satisfies the statement of Lemma 10) to the curve v^* and is implementable under limited liability. This will lead to the desired contradiction. We remind the reader that any implementable curve is convex and therefore continuous at any interior point.

Claim 1 $v^{L\prime}(\theta) \geq 0$ *for all* $\theta \in \hat{\Theta}$.

Proof There are three cases.

Case 1 $v^L(\theta) > v^*(\theta)$ *for all* $\theta \in int\Theta$.

Let $\hat{\theta} = \sup\left\{\theta : v^{L\prime}(\theta) < 0\right\}$. If $\hat{\theta} = 1$ consider the linear contract which induces every type $(\theta, c) : c \leq v^L(1)$ to experiment. It is implementable under limited liability, since this contract is of the form $(\hat{a}, 0)$ with $\hat{a} = v^L(1)$. From Lemma 10 this contract leads to a higher profit, a contradiction.

Assume $\hat{\theta} < 1$. Consider the alternative contract which induces the experimentation curve v^{**} : $v^{**}(\theta) = v^L(\theta)$ if $\theta \geq \hat{\theta}$ and $v^{**}(\theta) = v^L(\hat{\theta})$ otherwise. The new contract is implementable and from Lemma 10 leads to a contradiction.

[8] This holds [a.e.]. We omit this comment in the remaining of the proof.

Case 2 $v^L(\theta) = v^*(\theta)$ *for some* $\theta \in int\Theta$.

Let θ^* be the minimum θ such that $v^L(\theta) = v^*(\theta)$. Again, let $\hat{\theta} = \sup\left\{\theta : v^{L\prime}(\theta) < 0\right\}$.

Case 2.1 Assume first that $\hat{\theta} = 1$. Notice that in this case we have $v^L(\theta) > v^*(\theta)$ for all $\theta < \theta^*$ and $v^L(\theta) < v^*(\theta)$ for all $\theta > \theta^*$. Therefore the linear contract such that every type $(\theta, c) : c \leq v^L(\theta^*)$ experiments leads to a higher profit (Lemma 10), a contradiction.

Case 2.2 Assume that $\hat{\theta} < 1$.

Case 2.2.1 First, $\hat{\theta} \leq \theta^*$. Consider the alternative contract which induces the experimentation curve $v^{**} : v^{**}(\theta) = v^L(\theta)$ if $\theta \geq \hat{\theta}$ and $v^{**}(\theta) = v^L(\hat{\theta})$ otherwise. The new contract is implementable and leads to a higher profit (Lemma 10), a contradiction.

Case 2.2.2 Assume $\hat{\theta} > \theta^*$.

Consider the curve $\tilde{v}(\theta) = v^*(\theta^*) = v^L(\theta^*)$ for all θ. There are two cases:

Case 2.2.2.1 $\tilde{v}(\theta) \geq v^L(\theta)$ for all interior θ.

In this case the curve $v^{**} : v^{**}(\theta) = \tilde{v}(\theta)$ for all θ leads to a higher profit.

Case 2.2.2.2 $\tilde{v}\left(\hat{\theta}\right) = v^L\left(\hat{\theta}\right)$ for some interior $\hat{\theta}$.

The curve $v^{**} : v^{**}(\theta) = \tilde{v}(\theta)$ for all $\theta \leq \hat{\theta}$, $v^{**}(\theta) = v^L(\theta)$ elsewhere leads to a higher profit.

Case 3 $v^L(\theta) < v^*(\theta)$ *for all* $\theta \in int\Theta$.

This argument is analogous to the arguments above and its omitted for brevity.

Claim 2 *There exists* $\hat{\theta} < 1$ *such that* $\theta > \hat{\theta}$ *implies that* v^L *is linear.*

Proof There are two cases.

Case 1 $v^L(1) \geq v^*(1)$.

Notice that: a) by assumption $v^L(1) \geq v^*(1)$; b) for any contract $\left(\tilde{a}, \tilde{b}\right)$ we have $\tilde{b} = \tilde{v}'$ almost everywhere; c) $\theta > \tilde{\theta}$ we have $a^*(\theta) < 0$. Hence we conclude that there exists $\bar{\theta} \in int\Theta$ such that $\theta \in \left(\bar{\theta}, 1\right)$ implies that $v^L(\theta) > v^*(\theta)$. Consider the following two lines: Line 1: $A_1 + \partial v_+^L\left(\bar{\theta}\right)\theta$ where $A_1 + \left(\partial v_+^L\left(\bar{\theta}\right)\right)\bar{\theta} = v^L\left(\bar{\theta}\right)$; Line 2: $A_2 + \partial v_-^L(1)\,\theta$ where $A_2 + \partial v_-^L(1)\,1 = v^L(1)$. There exists $\hat{\theta} \in \left(\bar{\theta}, 1\right)$ such that $A_1 + \partial v_+^L\left(\bar{\theta}\right)\hat{\theta} = A_2 + \left(\partial v_-^L(1)\right)\hat{\theta}$. We use these 2 lines to construct a more profitable curve according to:

$$v^{**}(\theta) := \begin{cases} v^L(\theta) \text{ if } \theta < \bar{\theta} \\ A_1 + \partial v_+^L\left(\bar{\theta}\right)\bar{\theta} \text{ if } \theta \in \left(\bar{\theta}, \hat{\theta}\right) \\ P\left(A_2 + \partial v_-^L(1)\,\theta\right) \text{ elsewhere.} \end{cases}$$

Case 2 $v^L(1) < v^*(1)$.

In this case, there is an interval $\left(\bar{\theta}, 1\right)$ such that $v^L(\theta) < v^*(\theta)$ for all $\theta \in \left(\bar{\theta}, 1\right)$. Thus, for some $\theta' \in \left(\bar{\theta}, 1\right)$ consider the curve v^{**} defined by $v^{**}(\theta) = v^L(\theta)$ for all $\theta < \theta'$ and $v^{**}(\theta) := P\left(v^L(\theta') + \left(\frac{v^L(1) - v^L(\theta')}{1-\theta'}\right)(\theta - \theta')\right)$ otherwise. This curve is clearly closer to v^{**} in the sense of Lemma 10 . It suffices to show that this curve is implementable. For that, let $\left(a^L(1_-), b^L(1_-)\right)$ be the contract that was offered to type 1 under the putative optimal contract. Since v^L is convex we have $\partial v^L_-(1) \geq \left(\frac{v^L(1)-v^L(\theta')}{1-\theta'}\right)$. From $v^L_-(1) = a^L(1_-) + \partial v^L_-(1_-) = a^{**}(1_-) + b^{**}(1_-)$ we conclude that $a^L(1_-) \leq a^{**}(1_-)$. Hence the limited liability constraint holds. Q.E.D.

Define θ^* as the minimum θ such that $\theta', \theta'' > \theta$ implies $\left(a^L(\theta'), b^L(\theta')\right) = \left(a^L(\theta''), b^L(\theta'')\right)$. From Claim 2 it is well defined.

Claim 3 $v^L(\theta^*) = v^*(\theta^*)$.

Proof There are two cases.

Case 1 $v^L(\theta^*) > v^*(\theta^*)$.

Consider the line passing through $\left(\theta^*, v^L(\theta^*)\right)$ and $\left(1, v^L(1)\right) : A + B\theta$. Let $\check{\theta} \in \Re$ be such that $A + B\check{\theta} = \underline{c}$. Let $\dot{\theta} = \max\left\{\check{\theta}, 0\right\}$. There are two cases.

Case 1.1 $A + B\dot{\theta} \geq v^*\left(\dot{\theta}\right)$.

In this case letting $v^{**} : v^{**}(\theta) = \max\{A + B\theta, \underline{c}\}$ we obtain a more profitable curve.

Case 1.2 $A + B\dot{\theta} < v^*\left(\dot{\theta}\right)$.

Let $\ddot{\theta} \in \left(\dot{\theta}, \theta^*\right)$ be such that $A + B\ddot{\theta} = v^*\left(\ddot{\theta}\right)$. Consider the new curve v^{**} defined by $v^{**}(\theta) = P(A + B\theta)$if $\theta > \ddot{\theta}$ and $v^*(\theta)$ elsewhere. v^{**} is implementable and yields a profitable deviation.

Case 2 $v_L(\theta^*) < v^*(\theta^*)$.

There exists $\varepsilon > 0$ such that: i) for all $\theta \in [\theta^* - \varepsilon, \theta^* + \varepsilon]$ we have $v_L(\theta) < v^*(\theta)$; ii) For all $\theta \in (\theta^* - \varepsilon, \theta)$ the line $A + B\theta$ that passes through $(\theta^* - \varepsilon, v^L(\theta^* - \varepsilon))$ and $(\theta^* + \varepsilon, v^L(\theta^* + \varepsilon))$ is closer to v^* than v^L. Hence the following curve leads to a higher profit:

$$v^{**}(\theta) := \begin{cases} P(A + B\theta) & \text{if } \theta \in [\theta^* - \varepsilon, \theta^* + \varepsilon] \\ v^L(\theta) & \text{elsewhere.} \end{cases}$$

Claim 4 $v^L(\theta) = v^*(\theta)$ *for all* $\theta < \theta^*$.

Proof Notice that v^L is linear for all $\theta > \theta^*$ (Claim 2). We can write it as $A + B\theta$ for $\theta > \theta^*$.

Case 1 $\theta^* > \theta_0$. Then from the definition of θ_0 we have $B < v^{*\prime}(\theta^*)$.

Define $\mathcal{A} := \{\theta : A + B\theta = v^*(\theta),\ \theta < \theta^*\}$.

Case 1.1 $\mathcal{A} = \emptyset$.

In this case the contract that induces the experimentation curve $v^{**}(\theta) = P(A + B\theta)$ for all θ leads to a higher profit.

Case 1.2 $\mathcal{A} \neq \emptyset$.

Let $\theta_1 := \inf \mathcal{A}$. Clearly, the following curve is implementable and leads to a higher profit:

$$v^{**}(\theta) := \begin{cases} P(A + B\theta) & \text{if } \theta > \theta_1 \\ v^*(\theta) & \text{elsewhere.} \end{cases}$$

Case 2 $\theta^* \leq \theta_0$.

In this case we must have $B \geq v^{*\prime}(\theta^*)$, otherwise the contract defined by the curve v^{**}, which is $v^*(\theta) = P(v^*(\theta^*) + v^{*\prime}(\theta^*)(\theta - \theta^*))$ if $\theta > \theta^*$ and $v^*(\theta)$ elsewhere is a profitable deviation.

Therefore, suppose that $B \geq v^{*\prime}(\theta^*)$. Finally, notice that the contract induced by the curve

$$v^{**}(\theta) := \begin{cases} P(A + B\theta) \text{ if } \theta > \theta^* \\ v^*(\theta) \text{ elsewhere,} \end{cases}$$

is implementable and it is optimal among all contracts which are induce the curve $A + B\theta$ for $\theta > \theta^*$. This proves the Claim. Q.E.D.

The last claim shows that v^L is discontinuous at θ^*. From assumption A2 and Claim 5 it follows that $\theta^* \leq \theta_0$. Hence we will conclude the lemma showing that $\theta^* < \theta_0$ and that v^L is discontinuous at θ^*. First, it follows from assumptions A1 and A2 that if $\theta^* = \theta_0$ then the contract under limited liability for all $\theta > \theta_0$ is given by $(a^*(\theta_0), b^*(\theta_0))$ where $b^*(\theta_0) = v^{*\prime}(\theta_0)$ and $b^*(\theta_0)\theta_0 = v^*(\theta_0)$ (From assumption A1 $a^*(\theta_0) = 0$). We show that for some $\varepsilon > 0$ the contract that induces a curve that is identical to $v^*(\theta)$ until $\theta^* - \varepsilon$, passes through $(\theta^* - \varepsilon, v^*(\theta^* - \varepsilon))$ and has slope $v^{*\prime}(\theta^*) + \gamma(\varepsilon)$ for $\theta > \theta^* - \varepsilon$ achieves a higher profit, where the slope $v^{*\prime}(\theta^*) + \gamma(\varepsilon)$ is calculated such that the limited liability constraint binds. First, take $\tilde{\varepsilon} < \theta_0$. It is immediate that γ is differentiable on $[0, \tilde{\varepsilon}]$ and that $\gamma'(0) > 0$. Notice that for each ε in $[0, \tilde{\varepsilon}]$ the new curve $v^{**}(\varepsilon)$ will be identical to v^* for all $\theta < \theta_0 - \varepsilon$. For $\theta \in (\theta_0 - \varepsilon, 1)$ the curve passes through $(\theta_0 - \varepsilon, v^*(\theta_0 - \varepsilon))$ and has slope $v^{*\prime}(\theta_0) + \gamma(\varepsilon)$. That is, the contract for types $\theta > \theta_0 - \varepsilon$ is $a^{**}(\theta) = 0$, $b^{**}(\theta) = v^{*\prime}(\theta_0) + \gamma(\varepsilon)$. Thus, for any type $\theta > \theta_0 - \varepsilon$ the (conditional) probability that an agent experiments is $\mathrm{H}((v^{*\prime}(\theta_0) + \gamma(\varepsilon))\theta)$ and its expected payment is $(v^{*\prime}(\theta_0) + \gamma(\varepsilon))\theta$. Therefore, the derivative of the principal's profit at $\varepsilon = 0$ is:

$$\begin{aligned} &\frac{d}{d\varepsilon}\left[\begin{array}{c} \int_0^{\theta_0-\varepsilon} [\theta\alpha + (1-\theta)\beta - v^*(\theta)]\, H(v^*(\theta))\, d\theta \\ \int_{\theta_0-\varepsilon}^1 [\theta\alpha + (1-\theta)\beta - (v^{*\prime}(\theta^*) + \gamma(\varepsilon))\theta]\, H((v^{*\prime}(\theta_0) + \gamma(\varepsilon))\theta)\, d\theta \end{array}\right]_{\varepsilon=0} \\ = \;& \gamma'(0)\int_{\theta_0}^1 \left[\theta\alpha + (1-\theta)\beta - \left[v'^*(\theta_0)\theta + \frac{H(v^{*\prime}(\theta_0)\theta)}{h(v^{*\prime}(\theta_0)\theta)}\right]\right] h\left(v^{*\prime}(\theta_0)\theta\right) d\theta, \end{aligned}$$

where we used the fact that $v^*(\theta_0) = v^{*\prime}(\theta_0)\theta_0$. From the convexity of v^* and the monotonicity of the reverse hazard rate we have $v^{*\prime}(\theta_0)\theta + \frac{H(v^{*\prime}(\theta_0)\theta)}{h(v^{*\prime}(\theta_0)\theta)} < v^*(\theta) + \frac{H(v^*(\theta))}{h(v^*(\theta))}$ for $\theta > \theta^*$. Since $\theta\alpha + (1-\theta)\beta - \left[v^*(\theta)\theta + \frac{H(v^{*\prime}(\theta))}{h(v^*(\theta))}\right] = 0$ we conclude that the term above is positive, which concludes the proof. Q.E.D.

Lack of Concavity

Condition 2 *[CC] .Concavity-Convexity: There exists $\hat{c} \in (\underline{c}, c^{max})$ such that the inverse of the semi-elasticity of experimentation, $\frac{1}{\sigma(c)}$, is weakly concave for $\hat{c} \in (\underline{c}, c^{max})$, and strictly convex for $c > \hat{c}$.*

Let v^* represent the optimal curve when the principal is informed and let $(a^*(\theta), b^*(\theta))$ the correspondent optimal contract[9]. Let θ' be such that $\hat{c} = v^*(\theta')$.

[9]Notice that v^* may not be convex.

Proposition 3 follows directly from Lemma 12 below.

Lemma 12 *Suppose that condition [CC] holds There exists an optimal deterministic contract* $\left(\hat{a}(\theta), \hat{b}(\theta)\right)$*. There exists* $\theta^* < \theta'$ *such that all types* $\theta > \theta^*$ *choose the same contract. For all* $\theta < \theta^*$ *we have* $\left(\hat{a}(\theta), \hat{b}(\theta)\right) = (a^*(\theta), b^*(\theta))$*. Furthermore if If* $v^*(\theta^*) > \underline{c}$ *then* $\left(\hat{a}(\theta), \hat{b}(\theta\right)$ *is discontinuous at* θ^**.*

Proof of Lemma 12. The argument establishing existence is analogous to the one in Lemma 9 and is omitted. For the remaining of this proof let $\hat{v}$ be an optimal curve. Most of the proof of this Lemma is by contradiction. Lemma 10 will be essential. In some steps below, we will assume towards a contradiction that $\hat{v}$ has some properties which are inconsistent with the statement of the Lemma. We will then construct another curve v^{**} which is closer (satisfies the statement of Lemma 10) to the curve v^* and is implementable. This will lead to the desired contradiction. We remind the reader that any implementable curve is convex and therefore continuous at any interior point. The argument that shows that the optimal curve $\hat{v}$ is increasing everywhere is analogous to the one in Lemma 6 and is omitted.

For the remaining of this proof $\hat{v}$ denotes an experimentation curve induced by an optimal deterministic contract $\left(\hat{a}(\theta), \hat{b}(\theta)\right)$.

Claim 1 $\hat{v}$ *is linear in* $[\theta', 1]$.

There are three cases.

Case 1 $\hat{v}(\theta) > v^*(\theta)$ *for all* $\theta \in [\theta', 1)$.

Consider the line $A + B\theta$ such that $A + B\theta' = \hat{v}(\theta')$ and $B = \partial_+\hat{v}(\theta')$. There are two cases.

Case 1.1 $A + B\theta \geq v^*(\theta)$ *for all* $\theta \in [\theta', 1)$.

In this case the curve v^{**} : $v^{**}(\theta) = \hat{v}(\theta)$ if $\theta < \theta'$; $v^{**}(\theta) = P(A + B\theta)$ elsewhere, leads to a higher profit, a contradiction.

Case 1.2 $A + B\hat{\theta} = v^*\left(\hat{\theta}\right)$ *for some* $\hat{\theta} \in [\theta', 1)$.

In this case consider the line $A' + B'\theta$ which passes through $\left(\hat{\theta}, v^*\left(\hat{\theta}\right)\right)$ and $(1, \hat{v}(1))$. The following curve leads to a higher profit:

$$v^{**}(\theta) := \begin{cases} v^*(\theta) \text{ if } \theta < \theta' \\ A + B\theta \text{ if } \theta \in \left[\theta', \hat{\theta}\right] \\ P(A' + B'\theta) \text{ elsewhere.} \end{cases}$$

Case 2 $\hat{v}(\theta) < v^*(\theta)$ *for all* $\theta \in [\theta', 1)$.

In this case consider the line $A + B\theta$ which passes through $(\theta', \hat{v}(\theta'))$ and $(1, \hat{v}(1))$. The curve v^{**} such that $v^{**}(\theta) = \hat{v}(\theta)$ for all $\theta < \theta'$, $v^{**}(\theta) = P(A + B\theta)$ elsewhere leads to a higher profit.

Case 3 $\hat{v}(\theta) = v^*(\theta)$ *for some* $\theta \in [\theta', 1)$.

Consider the set: $\{\theta \in [\theta', 1) : \hat{v}(\theta) = v^*(\theta)\}$.

There are two cases.

Case 3.1 $\{\theta \in [\theta', 1) : \ \hat{v}(\theta) = \ v^*(\theta)\} = \{\theta_1\}$.

First assume that $\theta_1 = \theta'$. In this case the curve:

$$v^{**}(\theta) := \begin{cases} v^*(\theta) \text{ if } \theta < \theta' \\ P(v^*(\theta') + v'^*(\theta')(\theta - \theta')) \text{ elsewhere,} \end{cases}$$

leads to a higher profit. Hence assume that $\theta_1 > \theta'$.

Consider a line $A + B\theta$ passing through $(\theta_1, \hat{v}(\theta_1))$ with slope $B \in (v^{*\prime}(\theta_1), \hat{v}'(\theta_1))$. There are two cases.

Case 3.1.1 $A + B\theta \geq v^*(\theta)$ *for every* θ.

In this case the curve $v^{**} : v^{**}(\theta) = P(A + B\theta)$ leads to a higher profit, a contradiction.

Case 3.1.2 $A + B\hat{\theta} = v^*(\hat{\theta})$ *for some* $\hat{\theta}$.

The following curve leads to a higher profit:

$$v^{**}(\theta) := \begin{cases} v^*(\theta) \text{ if } \theta < \hat{\theta} \\ P(A + B\theta) \text{ elsewhere.} \end{cases}$$

Case 3.2 $\{\theta \in [\theta', 1) : \ \hat{v}(\theta) = \ v^*(\theta)\} = \{\theta_1, \theta_2\}$.

Without loss let $\theta_1 < \theta_2$. Consider the line $A + B\theta$ passing through $(\theta_1, \hat{v}(\theta_1))$ and $(\theta_2, \hat{v}(\theta_2))$. There are two cases.

Case 3.2.1 $A + B\theta \geq v^*(\theta)$ *for every* θ.

In this case the curve $v^{**} : v^{**}(\theta) = P(A + B\theta)$ leads to a higher profit.

Case 3.2.2 $A + B\hat{\theta} = v^*(\hat{\theta})$ *for some* $\hat{\theta}$.

The following curve leads to a higher profit:

$$v^{**}(\theta) := \begin{cases} v^*(\theta) \text{ if } \theta < \hat{\theta} \\ P(A + B\theta) \text{ elsewhere.} \end{cases}$$

Define $\theta^* = \inf\{\theta : \hat{v} \text{ is linear in } [\theta, 1]\}$. Notice that from Claim 1 $\theta^* \leq \theta'$.

Claim 2: *If* $\theta^* > 0$ *then* $\hat{v}(\theta^*) = v^*(\theta^*)$.

Proof: There are two cases.

Case 1 $\hat{v}(\theta^*) > v^*(\theta^*)$.

Consider the line passing through $\hat{v}(\theta^*)$ with slope $\partial_+\hat{v}(\theta^*) : A + B\theta$. There are two cases.

Case 1.1 $A + B\theta \geq v^*(\theta)$ *for all* $\theta < \theta^*$.

The curve $v^{**} : v^{**}(\theta) = P(A + B\theta)$ leads to a higher profit.

Case 1.2 $A + B\theta_1 = v^*(\theta_1)$ *for* $\theta_1 < \theta^*$.

The following curve leads to a higher profit:

$$v^{**}(\theta) := \begin{cases} \hat{v}(\theta) \text{ if } \theta < \theta_1 \\ P(A + B\theta) \text{ elsewhere.} \end{cases}$$

Case 2 $\hat{v}(\theta^*) < v^*(\theta^*)$.

Consider the set: $\left\{\tilde{\theta} < \theta^* : \hat{v}\left(\tilde{\theta}\right) = v^*\left(\tilde{\theta}\right)\right\}$. There are two cases.

Case 2.1 $\{\theta < \theta^* : \hat{v}(\theta) = v^*(\theta)\} = \emptyset$.

Let $\tilde{\theta} := \sup\{\theta < \theta^* : \hat{v}(\theta) \geq \underline{c}\}$. For each $\varepsilon > 0$ let $A_\varepsilon + B_\varepsilon\theta$ the curve passing through $\left(\tilde{\theta}, \hat{v}\left(\tilde{\theta}\right)\right)$ and $(\theta^* + \varepsilon, \hat{v}(\theta^* + \varepsilon))$.

It is straightforward to see that for a small $\varepsilon > 0$ the following curve is implementable and leads to a higher profit:

$$v^{**}(\theta) := \begin{cases} P(A_\varepsilon + B_\varepsilon\theta) \text{ if } \theta \in \left[\tilde{\theta}, \theta^* + \varepsilon\right] \\ \hat{v}(\theta) \text{ if } \theta > \theta^* + \varepsilon. \end{cases}$$

Case 2.2 $\{\theta < \theta^* : \hat{v}(\theta) = v^*(\theta)\} \neq \emptyset$.

Let $\hat{\theta} := \sup\{\theta < \theta^* : \hat{v}(\theta) = v^*(\theta)\}$. For each $\varepsilon > 0$ let $A_\varepsilon + B_\varepsilon\theta$ the curve passing through $\left(\tilde{\theta}, \hat{v}\left(\tilde{\theta}\right)\right)$ and $(\theta^* + \varepsilon, \hat{v}(\theta^* + \varepsilon))$.

It is straightforward to see that for a small $\varepsilon > 0$ the following curve is implementable and leads to a higher profit:

$$v^{**}(\theta) := \begin{cases} P(A_\varepsilon + B_\varepsilon\theta) \text{ if } \theta \in \left[\tilde{\theta}, \theta^* + \varepsilon\right] \\ \hat{v}(\theta) \text{ if } \theta > \theta^* + \varepsilon. \end{cases}$$

Claim 2 *$\hat{v}(\theta) = v^*(\theta)$ for all $\theta < \theta^*$.*

From Claim 1 we know that $\hat{v}(\theta) = A + B\theta$ if $\theta > \theta^*$. If $B \geq v'^*(\theta)$ the following curve is implementable and achieves the highest profit among all curves such that $\hat{v}(\theta) = A + B\theta$ if $\theta > \theta^*$:

$$v^{**}(\theta) := \begin{cases} v^*(\theta) \text{ if } \theta < \theta^* \\ \hat{v}(\theta) \text{ if } \theta \geq \theta^*. \end{cases}$$

Hence, suppose towards a contradiction that $B < v'^*(\theta)$. There are two cases:

Case 1 $\{\theta < \theta^* : A + B\theta = v^*(\theta)\} = \emptyset$.

In this case the following curve leads to a higher profit: $v^{**} : v^{**}(\theta) = P(A + B\theta)$ for every θ.

Case 2 $\{\theta < \theta^* : A + B\theta = v^*(\theta)\} \neq \emptyset$.

Let $\theta_1 = \sup\{\theta < \theta^* : A + B\theta = v^*(\theta)\}$. The following curve leads to a higher profit:

$$v^{**}(\theta) := \begin{cases} v^*(\theta) \text{ if } \theta < \theta_1 \\ P(A + B\theta) \text{ elsewhere.} \end{cases}$$

Claim 3 *If $v^*(\theta^*) > \underline{c}$ then $\left(\hat{a}(\theta), \hat{b}(\theta)\right)$ is discontinuous at θ^*.*

Proof: Assume that $v^*(\theta^*) > \underline{c}$ and $\left(\hat{a}(\theta), \hat{b}(\theta)\right)$ is continuous at θ^*. Notice that we have established that: i) $\hat{v}(\theta) = v^*(\theta)$ for all $\theta < \theta^*$ (Claim 2). Hence if $\left(\hat{a}(\theta), \hat{b}(\theta)\right)$ is continuous at θ^* then $\hat{v}$ is differentiable at θ^* and $\hat{v}'(\theta^*) = v^{*\prime}(\theta^*)$. Assume that there exists $\hat{\theta} \in (\theta^*, 1)$ such that

$\hat{v}\left(\hat{\theta}\right) = v^*\left(\hat{\theta}\right)$ (the other case is straightforward). This immediately implies that $\theta^* < \theta'$. For small $\varepsilon > 0$ consider the implementable curve $\tilde{v}(\varepsilon, \cdot)$:

$$\tilde{v}(\varepsilon, \theta) := \begin{cases} v^*(\theta) \text{ if } \theta < \theta^* - \varepsilon \text{ or } \theta > \hat{\theta} \\ v^*(\theta^* - \varepsilon) + \left(\frac{v^*(\hat{\theta}) - v^*(\theta^* - \varepsilon)}{\hat{\theta} - \theta^* + \varepsilon}\right)(\theta - \theta^* + \varepsilon) \text{ otherwise.} \end{cases}$$

This new curve is linear in $\left(\theta^* - \varepsilon, \hat{\theta}\right)$. Notice that (for small ε) there exists $\delta(\varepsilon) > 0$ such that $\tilde{v}(\varepsilon, \theta^* + \delta(\varepsilon)) = v^*(\varepsilon, \theta^* + \delta(\varepsilon))$. That is, the new curve $\tilde{v}(\varepsilon, \theta)$ is above $v^*(\theta)$ for all $\theta \in (\theta^* - \varepsilon, \theta^* + \delta(\varepsilon))$.

Let $\bar{g} = \max_\theta g(\theta)$, $\underline{g} = \min_\theta g(\theta)$, $h = \max_\theta h(\theta)$ and $\underline{h} = \min_c h(c)$. Take a curve $\tilde{v}$ and some θ. From Lemma 1, if all types $(\theta, c) : c \leq \tilde{v}(\theta)$ experiment, then the principal's expected profit conditional on θ is:

$$\int_{\underline{c}}^{\tilde{v}(\theta)} \left(\theta H + (1-\theta)L - c - \frac{H(c)}{h(c)}\right) h(c)dc.$$

Notice that the function $R(c) := \left(c + \frac{H(c)}{h(c)}\right)$ is Lipschitz continuous with constant $k = 1 + \max_c \left|\left(\frac{h^2(c) - h'(c)H(c)}{h(c)^2}\right)\right|$. Therefore, if $\tilde{v}\ (\theta) - v^*(\theta) = k > 0$ then

$$\begin{aligned} & \left| \begin{array}{l} \int_{\underline{c}}^{\tilde{v}(\theta)} \left(\theta H + (1-\theta)L - c - \frac{H(c)}{h(c)}\right) h(c)dc \\ - \int_{\underline{c}}^{v^*} \left(\theta H + (1-\theta)L - c - \frac{H(c)}{h(c)}\right) h(c)dc \end{array} \right| \\ = & \left| \int_{v^*(\theta)}^{\tilde{v}(\theta)} \left(\frac{H(c)}{h(c)} - \frac{H(v^*(\theta))}{h(v^*(\theta))}\right) h(c)dc \right| \\ \leq & \ h \int_0^{v'(\theta) - v^*(\theta)} kxdx = \left(\frac{k\bar{h}\left(v'(\theta) - v^*(\theta)\right)^2}{2}\right). \end{aligned} \tag{23}$$

Our assumptions on H imply that the curve v^* is C^2. Straightforward algebra shows that there exists $\varepsilon_1 > 0$ such that $\varepsilon < \varepsilon_1$ implies $\delta(\varepsilon) < 2\varepsilon$. Hence $|\theta^* + \delta(\varepsilon) - (\theta^* - \varepsilon)| < 3\varepsilon$. Thus

$$\sup_{\theta \in [\theta^* - \varepsilon, \theta^* + \delta(\varepsilon)]} |\tilde{v}(\theta) - v^*(\theta)| \leq 3\varepsilon \sup_{\tilde{\theta}} v^*(\tilde{\theta}) \leq 3\varepsilon(H - L), \tag{24}$$

where we have used $\theta H + (1-\theta)L - v^*(\theta) - \frac{H(v^*(\theta))}{h(v^*(\theta))} = 0$. Therefore, using $|\theta^* + \delta(\varepsilon) - (\theta^* - \varepsilon)| < 3\varepsilon$, (23) and (21) we conclude that over the region $\theta \in (\theta^* - \varepsilon, \theta^* + \delta(\varepsilon))$ an upper bound to the loss due to this new curve is

$$3\varepsilon\bar{g} \times \left(\frac{k\bar{h}\left(3\varepsilon(H-L)\right)^2}{2}\right) = \left(\frac{27(H-L)^2\bar{g}k\bar{h}}{2}\right)\varepsilon^3. \tag{25}$$

Next, we calculate a lower bound to the gain obtained from the new curve $\tilde{v}$ over $\left(\theta^* + \delta(\varepsilon), \hat{\theta}\right)$. For that, notice that for all $\theta \in \left(\theta^* + \delta(\varepsilon), \hat{\theta}\right)$ we have $\hat{v}(\theta) < \tilde{v}(\theta) < v^*(\theta)$.

Notice that the slope of the curve $\tilde{v}(\varepsilon,\cdot)$ is $B(\varepsilon)=\left(\frac{v^*(\hat{\theta})-v^*(\theta^*-\varepsilon)}{\hat{\theta}-\theta^*+\varepsilon}\right)$. We have:

$$B'(0)=0 \quad B''(0)=-\left(\frac{v''^*(\theta^*)}{\hat{\theta}-\theta^*}\right).$$

Hence, using a Taylor expansion, there exists $\varepsilon_2>0$ such that $\varepsilon<\varepsilon_2$ implies that the slope of the curve $\hat{v}'$ is at least $\frac{1}{4}\left(\frac{v''^*(\theta^*)}{\hat{\theta}-\theta^*}\right)\varepsilon^2$ greater than the slope of $\tilde{v}(\varepsilon,\cdot)$ for each $\theta\in\left(\theta^*+\delta(\varepsilon),\hat{\theta}\right)$.

Next, consider some interval $[\theta_a,\theta_b]\subseteq\left(\theta^*+\left(\frac{\hat{\theta}-\theta^*}{4}\right),\theta^*+3\left(\frac{\hat{\theta}-\theta^*}{4}\right)\right)$. From the slope difference obtained above we conclude that

$$\min_{\theta\in[\theta_a,\theta_b]}\left(\tilde{v}(\varepsilon,\theta)-v^*(\theta)\right)\geq\left(\frac{v''^*(\theta^*)}{16}\right)\varepsilon^2. \tag{26}$$

There exists $\kappa>0$ and $\varepsilon_3>0$ such that $\varepsilon<\varepsilon_3$ implies

$$\min_{\theta\in[\theta_a,\theta_b]}\left(v^*(\theta)-\hat{v}(\theta)\right)>\kappa \text{ and } \min_{\theta\in[\theta_a,\theta_b]}\left(v^*(\theta)-\tilde{v}(\varepsilon,\theta)\right)>\kappa. \tag{27}$$

Therefore, there exists $\xi>0$ such that for all $\theta\in[\theta_a,\theta_b]$ and for all $c\in[v^*(\theta),\tilde{v}(\varepsilon,\theta)]$ we have:

$$\left(\theta H+(1-\theta)L-c-\frac{H(c)}{h(c)}\right)>\xi. \tag{28}$$

From (26) and (28) we conclude that a lower bound to the gain obtained from the new curve $\tilde{v}$ over $\left(\theta^*+\delta(\varepsilon),\hat{\theta}\right)$ is:

$$|\theta_b-\theta_a|\times\left(\frac{v''^*(\theta^*)}{16}\right)\varepsilon^2\times\xi\times\underline{gh} \tag{29}$$

Finally, if $\varepsilon<\min\left\{\varepsilon_1,\varepsilon_2,\varepsilon_3,\left(\frac{|\theta_b-\theta_a|\times\left(\frac{v''^*(\theta^*)}{16}\right)\times\xi\times\underline{gh}}{27(H-L)^2\bar{g}kh}\right)\right\}$ we conclude that the lower bound on the gains from the curve $\tilde{v}$ estimated by (29) is greater than the lower bound on the loss obtained from $\tilde{v}$ estimated by(25). This contradicts the assumption that $\hat{v}$ is optimal. Q.E.D.

References

[1] Bergemann, D. and U. Hege (1998): "Venture Capital Financing, Moral Hazard, and Learning", Journal of Banking and Finance, 22, 703-735.

[2] Bergemann, D. and U. Hege (2005): "The Financing of Innovation: Learning and Stopping", RAND Journal of Economics, 36, 719-752.

[3] Bonatti, A. and J. Hörner (2011): "Collaborating", American Economic Review, 101, 632–663.

[4] Canidio, A. and P. Legros (2012): "The Value of Failures: Career Concerns, Experimentation and Entrepreneurship", Unpublished.

[5] Campbell, Ederer, Spinnewijn (2012): "Delay and Deadlines: Free-riding and Information Revelation in Partnerships", Unpublished.

[6] Ederer, F. (2009): "Launching a Thousand Ships: Incentives for Parallel Innovation", Unpublished.

[7] Garfagnini, U. (2011): "Delegated Experimentation", Unpublished.

[8] Gerardi, D. and L. Maestri (2012): "A Principal-Agent Model of Sequential Testing", Theoretical Economics, 7, 425-463.

[9] Gittins, J. C. (1979): "Bandit Processes and Dynamic Allocation Indices", Journal of the Royal Statistical Society. Series B (Methodological), 41, 148-177.

[10] Halac, M., N. Kartik and Q. Liu (2012): "Optimal Contracts for Experimentation", Unpublished.

[11] Hörner, J. and L. Samuelson (2012): "Incentives for Experimenting Agents", Unpublished.

[12] Kaplan, S. and P. Stromberg (2003): "Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts", Review of Economic Studies,70, 281-315.

[13] Keller, G., S. Rady, and M. Cripps (2005): "Strategic Experimentation with Exponential Bandits", Econometrica, 73, 39-68.

[14] Klein, N. (2012): "The Importance of Being Honest", Unpublished.

[15] Manso, G. (2011): "Motivating Innovation", Journal of Finance, 66, 1823-1860.

[16] Milgrom, P. (2004): Putting Auction Theory to Work, Cambridge University Press, New York.

[17] Rochet, J. C. and L. Stole (2003): "The Economics of Multidimensional Screening", in Advances in Economics and Econometrics: Theory and Applications - Eight World Congress, series Econometric Society Monographs, n. 36, Cambridge University Press, 783-826.

[18] Rothschild, M (1974): "A two-armed bandit theory of market pricing", Journal of Economic Theory, 9, 185-202.