Reputation, Learning and Project Choice in Frictional Economies

Farzad Pourbabaee

April 2023

Abstract

I introduce a dynamic model of learning and random meetings between a long-lived agent with unknown ability and heterogeneous projects with observable qualities. The outcomes of the agent's matches with the projects determine her posterior belief about her ability (i.e., her reputation). In a self-type learning framework with endogenous outside option, I find the optimal project selection strategy of the agent, that determines what types of projects the agent with a certain level of reputation will accept. Sections of the optimal matching set become *increasing intervals*, with different cutoffs across different types of the projects. Increasing the meeting rate has asymmetric effects on the sections of the matching sets: it unambiguously expands the section for the high type projects, while on some regions, it initially expands and then shrinks the section of the low type projects.

JEL classification: D81; D83; O31
Keywords: Reputation; Learning; Optimal Stopping

Contents

1 Introduction

Much of the theoretical literature on experimentation and project choice is about learning the other party's (namely the project's) type. In this paper, however, I shift the attention to learning the self-type, and thus engaging in a *self-experimentation* setting. Specifically, the agent in my paper does not know her type, and the only way to learn it is by matching with the projects and observing their outputs.

There are natural instances where agents *learn* their type through the course of their matches with other parties. For example, firms learn about their productivity while they are matched with workers. Colleges learn about the quality of their teaching staff while students are enrolled in their programs. Venture capitalists learn about their ability and the quality of their post-investment services while investing in the startups.[1] Common in all these cases is the cost of maintaining the match and the *tangible* created surplus (such as the high-quality output of production in the first case, students' accomplishment in the second case, and the startups' success in the last case). These tangible gains can be isomorphically captured by the choice of the *production function*, that takes in the types of partners and returns the output.

However, when the agent holds incomplete information about her type, there is also an *intangible* gain due to the learning, that cannot be nested in the former construct. Because, what is now used as an input to the production function is no longer the static type of the agent, but a dynamic state process that reflects the agent's belief of her own type. Specifically, in addition to the tangible gains, there are now information gains from agent's project choices, as present selections convey information about the agent's ability, that in turn can be used in future choices of projects. The basic research question that I ask and answer in this paper is: presented with heterogeneous projects, that differ in their expected payoff, what is the agent's optimal project selection policy as a function of her reputation?

In this economy, the agent is ex ante endowed with a high or low *immutable* type $\theta \in \{L, H\}$, that is hidden to herself. On the other hand, there are heterogeneous projects

---

[1]For example, Sørensen (2007) teases out the positive treatment effect of the VCs' involvement in their portfolio companies from the concerns regarding the sorting and selection. In addition, when it comes to VCs' underlying *ability* Gompers and Lerner (1999) argue that the empirical evidence favors the learning model (symmetric but incomplete information) compared to the signalling model (asymmetric information).

with observable qualities denoted by $q \in \{a, b\}$, which I often refer to them by a-projects and b-projects. The agent randomly meets the projects subject to the search frictions and decide to accept them or not. Once a project is accepted, there will be a random success event whose arrival intensity depends on the types of the agent and the project (namely on θ and q). Agent continuously updates her belief about the underlying type during the course of her matches. Therefore, I interpret the posterior belief as her reputation and denote it by π.

Whenever the agent pairs up with a project, a learning opportunity is created about her type. Since maintaining the match is costly, the agent effectively solves a stopping time problem, by which she weighs the *matching value function $v_q(\pi)$* (that is a function of her current reputation π and the type of the project q) against the reservation value $w(\pi)$ — the value of holding the current reputation while being unmatched, that is called the *reputation value function* throughout the paper. Because of the random meetings framework, these two functions are intertwined in the fixed-point. That is the reputation function is simply the expected discounted value of future surpluses that the agent extracts, and the matching value function is the solution to the free boundary problem with the exit option w. The continuation region of this free boundary problem determines the optimal matching set $\mathcal{M}$, that in turn defines the acceptable levels of reputation with which the agent selects and holds on to a particular project. Specifically, $(q, \pi) \in \mathcal{M}$ if $v_q(\pi) > w(\pi)$. In light of this specification and following the terminology of the optimal stopping literature, I use the *matching set* and *continuation region* interchangeably and both refer to the subset $\mathcal{M}$.

The central innovation of this paper is to study the optimality and shape of these matching sets (namely the continuation regions) when the agent has long-run incentives and learn her ability as she selects and matches with the projects. Specifically, I find and study the properties of the optimal tuple $\langle w^*, v^*, \mathcal{M}^* \rangle$. The main point of the departure from the experimentation literature (e.g., Keller et al. (2005)) is the endogeneity of the outside option w, that determines the types of acceptable projects in the agent's optimal policy. In addition, the subject of learning in the experimentation literature is the project's type, whereas in my paper the learning is about the self-type and projects provide the context for learning and a source of creating surplus.

1.1 Organization of Results

In Section 2, I introduce the dynamic learning and project selection model. Three main objects in the study of agent's optimal policy are the matching value function $v_q(\pi)$, the reputation or reservation value function $w(\pi)$, and the matching set $\mathcal{M}$. We will see how the agent's optimal policy can be translated to a fixed-point solution of a system that connects the above three elements.

In Section 3, I find the unique optimal tuple $\langle w^*, v^*, \mathcal{M}^* \rangle$ in the space of continuously differentiable value functions, i.e., $C^1[0,1]$. I study the properties of the value functions (such as monotonicity and convexity), and show the sections of the optimal matching set are *increasing intervals*, and hence the agent's optimal policy is to stay matched with a project so long as her reputation is larger than a certain threshold. The threshold for high type projects is lower, and hence the agent shows more tolerance for failure when matched to the high types. Specifically, letting $\mathcal{M}_q^*$ to indicate the interval of reputation levels at which the agent stays with a q-project, then at the optimum $\mathcal{M}_a^* \subseteq \mathcal{M}_b^*$. Due to the search frictions, there is a cost region that $\mathcal{M}_a^* \neq \varnothing$, and thus even the low type projects get selected.

I present the qualitative features of the matching sets and the value functions in Section 4. In particular, to uncover the unique role of learning on the shape of the optimal matching sets and value functions, I study the *no-learning* counterpart of the original model in Section 4.1. Specifically, I let the agent's true type to be equal to her reputation (i.e., a number $\pi \in [0,1]$), as opposed to a background hidden binary variable $\theta \in \{L, H\}$. This will shut down the learning channel, that is the Bayesian learning force will be absent in the associated Bellman equations.

Subsequently, in this no-learning environment, I find the unique optimal outcome $\langle \hat{w}, \hat{v}, \widehat{\mathcal{M}} \rangle$ in the space of continuous functions $C[0,1]$. The matching value function $\hat{v}$ is locally concave with kinks on the boundaries. Namely, it is no longer convex and continuously differentiable despite its counterpart v^* in the original learning model. Losing convexity in the value functions (due to the absence of learning incentives) leaves the matching sets smaller than their learning counterpart $\mathcal{M}^*$. Lastly, I show in this setting, lowering the search frictions *symmetrically* expands both sections of the matching set (i.e., $\widehat{\mathcal{M}}_a$ and $\widehat{\mathcal{M}}_b$).

Next, in Section 4.2, I present the comparative statics of the original learning model.

Specifically, based on the unique existence of the optimal tuple $\langle w^*, v^*, \mathcal{M}^* \rangle$ in the space of C^1 function, I present the comparative statics of this tuple with respect to the primitives of the economy. An important one among them is the impact of search frictions on the size of the sections of the optimal matching set, namely $\mathcal{M}_a^*$ and $\mathcal{M}_b^*$. I show that decreasing the search frictions unambiguously expands the high type section $\mathcal{M}_b^*$, but on some regions, initially expands and then shrinks the low type section $\mathcal{M}_a^*$. This *asymmetric* response to search frictions was not present in the no-learning matching sets $(\widehat{\mathcal{M}}_a, \widehat{\mathcal{M}}_b)$. Based on these results, I present the analogies and the implications of my findings to the economics of venture capital, and draw some policy relevant conclusions.

Finally, the paper concludes in Section 5.

1.2   Related Literature

The Bayesian learning force in the agent's decision problem in this paper is based on the exponential arrival of breakthroughs, and in that sense the paper is related to the experimentation literature with exponential news processes, initiated by Keller et al. (2005), and expanded in the follow-up works of Keller and Rady (2010) and Keller and Rady (2015). The exponential Bandit approach has also been applied to other strategic settings with payoff and informational externalities between players (Boyarchenko, 2021; Margaria, 2020; Das et al., 2020). The common theme in this line of research is that it is the project's type that is unknown to the decision maker(s), whereas in the current paper, the projects' types are observable and they provide a context for the decision maker to learn her type while they are being selected.

In the context of reputation building (when the information about the persistent or dynamic self-type is incomplete) and interpreting the reputation as the posterior belief, this paper is related to Holmström (1999); Board and Meyer-ter-Vehn (2013); Bonatti and Hörner (2017). However the kind of economic engagement that releases informative signals in these papers is the agent's effort, and in the current study is the agent's project selection.

The analysis of this paper has also the flavor of the literature on learning in labor markets such as the works by Jovanovic (1979); Moscarini (2005); Li and Weng (2017). Aside from differences in the context and motivation, the subject of learning in these studies is the *match specific* parameter, and not the underlying types of the agents. Therefore,

the information released over the present match has no bearing on the future matches and naturally the reputational aspects are absent.

There is also previous research on how agents hold *perfect* private information about themselves, and receive some form of information about the type of their partner before the match (Chakraborty et al., 2010; Liu et al., 2014). My setting is different than these works, mainly in the sense that the agent in this paper has incomplete information about herself, and one of the motives in her matching decisions (besides receiving the tangible surplus from the projects she accepts) is learning her type.

My findings also contribute to the literature about Bandits with correlated arms (Camargo, 2007; Rosenberg et al., 2013). This problem is known to be difficult, and thus very little has been achieved in economics literature. The self-experimentation model that is developed in this paper is formally equivalent to a two-armed Bandit setting, where the arms' payoffs are correlated. Specifically, what correlates the payoffs of the projects is the single dimensional variable that represents the agent's hidden type.

2 Model

2.1 Agent, Projects and Dynamic Timeline

In this subsection, I describe the elements of an economy populated by a single long-lived agent and a continuum of projects.

Agent. The agent is a long-lived individual with the rate of time preference $r > 0$. She holds incomplete information about her immutable type $\theta \in \{L, H\}$. The σ-field $\mathcal{I}_t$ aggregates all the information that is available in the economy at time $t \in \mathbb{R}_+$. The agent cares about her reputation, which is the posterior belief about her type. Given the filtration $\boldsymbol{I} = \{\mathcal{I}_t\}$, $\pi_t = \mathsf{P}\left(\theta = H \mid \mathcal{I}_t\right)$ refers to her time-t reputation.

Projects. The entities on the other side of this economy are treated as projects that are selected by the agent. Specifically, they have no bargaining power against the agent.[2] Each

---

[2]This assumption makes the analysis substantially simpler, yet it downplays the strategic role of "project owners" in the optimal outcome. However, given the paper's focus on the agent's side and her reputational concerns, such an abstraction seems plausible. Also from the empirical standpoint, for example in the venture capital literature, it is shown that firms can continue their projects without their original entrepreneurs as the project owner; see Ueda (2004) and the references therein such as Gorman and Sahlman (1989); Hellmann and Puri (2002).

project is endowed with a type $q \in \{a, b\}$, which is publicly observable. The (unnormalized) mass of type-q projects is φ_q for $q \in \{a, b\}$, exogenously replenished and held constant.

Meetings and project selection. The agent randomly meets the projects subject to the search frictions, with the meeting rate of $\kappa > 0$, and the matching technology is *quadratic*. That is the probability with which the agent meets a type-q project over the period $\mathrm{d}t$ is approximately equal to $\kappa\varphi_q\mathrm{d}t$. Furthermore, the matches are one-to-one, that is both parties have capacity constraint over the number of partners they can accept.

Output and reputation. Given the selection of a type-q project by the type-θ agent, the success arrives at the rate of λ_q, if the agent is of high type. Otherwise, there will be no success. This means breakthroughs are *conclusive* about the agent's ability.[3] Type-b projects are superior to type-a, in the sense that $\lambda_b > \lambda_a$. The agent has to cover the flow cost of $c > 0$ that is common across all projects — and, for a non-trivial setting, one has to assume $c < \lambda_a < \lambda_b$. In return, she receives the right to terminate the project at her will, so conceptually a stopping time problem is solved by the agent ex post to every selection of a project. The flow cost c captures both the running cost of the project and learning about the self-type θ. I assume there is a mechanism in the economy that tracks the output of each project and records the Bayes-updated posterior of the agent. This information is reflected in the filtration $\boldsymbol{I}$. The posterior dynamics for the reputation process (resulted from the Bayes law) follows

$$\mathrm{d}\pi_t = -\pi_t(1 - \pi_t)\lambda_q\mathrm{d}t, \tag{2.1}$$

prior to the success. Since the breakthroughs are conclusive, upon the success, π_t immediately jumps up to one.

Figure 1 summarizes the dynamic timeline for the agent, who starts the cycle with reputation π, and after some exponentially distributed time meets a project randomly drawn from the population of available ones. A decision to accept or reject the contacting project is made by the agent. Upon rejection, she returns to the initial node, and

---

[3]The analysis with inconclusive breakthroughs, in which the success comes around even with a low type agent leads to several intractable steps, hence is omitted. Due to its tractability, conclusive breakthroughs are used in a number of recent studies (Bonatti and Hörner, 2017; Boyarchenko, 2021; Margaria, 2020; Das et al., 2020). Also, the choice of exponential processes to model the breakthroughs is more natural when news arrive in discrete and randomly separated instants, than the Wiener process treatment of experimentation (e.g., see Bolton and Harris (1999) and Pourbabaee (2020)).

conditioned on acceptance an investment problem with the flow cost of c is solved.

Finally, I interpret the success as an event in which the breakthrough happens before the agent stops the project, thereby rationally updating her belief upwards. And the failure refers to the outcome, where the project is terminated before the success arrives, thus the agent returns to the unmatched status with a lower reputation. Importantly, after a success or failure event, the match is dissolved and both the agent and the project become available.

The self-experimentation model presented here is formally equivalent to a special two-armed Bandit setting, in which the payoffs to the two arms are correlated. Specifically, the agent's type θ is the variable that correlates the payoffs of the two arms. Therefore, the upcoming analysis can also be read through the lens of experimentation literature with correlated arms.
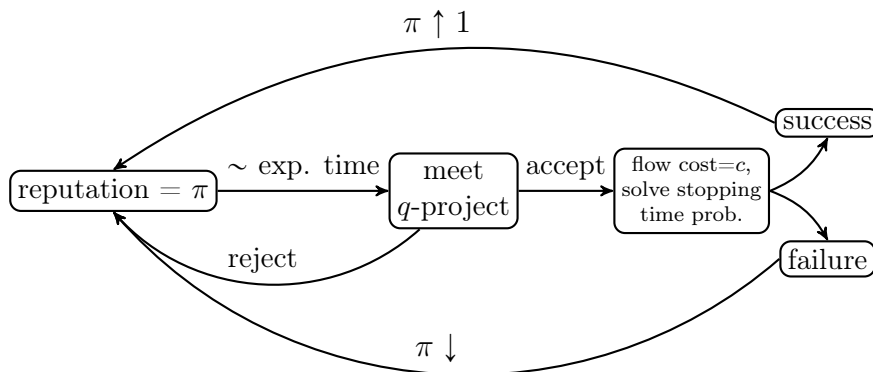
Figure 1: Decision timeline for the agent

2.2 Value Functions and Matching Sets

In this section, I show that the optimal strategy of the agent can be summarized in the choice of the matching sets. Additionally, I present the *necessary* Bellman equations that every optimal C^1 value functions, associated with the optimal matching sets, must satisfy.

Let $w(\pi)$ be the optimal value of holding reputation π, when the agent is *unmatched.* This function shall be treated as the agent's outside option and is weighed against the optimal matching value function upon the meetings.[4] The matching value function when

---

[4]Henceforth, I often drop the word "optimal", as it is clear from the context.

the agent with reputation π selects and stays with a type-q project is $v_q(\pi)$, that is the expected value of discounted future payoffs generated by this project.

The optimality of the match between the agent of reputation π and a type-q project requires that $v_q(\pi) > w(\pi)$, in that case I say $(q, \pi) \in \mathcal{M} \subseteq \{a, b\} \times [0, 1]$, where $\mathcal{M}$ is called the matching set (or interchangeably the continuation region). Also, understood from the context, $\mathcal{M}(\pi)$ (respectively, $\mathcal{M}_q$) refers to the π (respectively, q) *section* of this two dimensional set. In addition, often in the paper I use the indicator function $\chi_q(\pi)$ to denote whether the agent with reputation π accepts a type-q project, that is whether $(q, \pi) \in \mathcal{M}$ or not.

Recall that φ denotes the mass of available projects in the economy (that are treated exogenously as the primitives of the model). The agent also meets type-q projects at the rate of $\kappa \varphi_q$. If a q-project is acceptable, it leads to a surplus of $v_q(\pi) - w(\pi)$ for the agent. Hence, the following Bellman equation follows:

$$rw(\pi) = \kappa \sum_{q \in \mathcal{M}(\pi)} \varphi_q \big(v_q(\pi) - w(\pi)\big). \tag{2.2}$$

Next, I formally define the optimal matching value function, $v_q(\pi)$, and present the necessary Bellman equation that it satisfies.

Imagine a match between the agent with an initial reputation π and a type-q project. Let σ_q represent the random exponential time of success with the unit payoff and the arrival intensity of λ_q if $\theta = H$. Therefore, the matching value function $v_q(\cdot)$ is an endogenous outcome of a free boundary problem with the outside option w. In that, the agent selects an optimal stopping time τ, upon which she stops backing the project, taking into account the project's success payoff and her reputation value w:

$$v_q(\pi) = \sup_\tau \left\{ \mathsf{E}\left[e^{-r\sigma_q} - c\int_0^{\sigma_q} e^{-rs}\mathrm{d}s + e^{-r\sigma_q}w(\pi_{\sigma_q}); \sigma_q \leqslant \tau\right] \right.$$
$$\left. + \mathsf{E}\left[-c\int_0^\tau e^{-rs}\mathrm{d}s + e^{-r\tau}w(\pi_\tau); \sigma_q > \tau\right] \right\}. \tag{2.3}$$

Specifically, τ is adapted to the filtration $\boldsymbol{I}$. Namely, $\{\tau \leqslant t\}$ is $\{\pi_s : s < t\}$-measurable for every $t \in \mathbb{R}_+$.

Formally, in the above stopping time problem, if the success happens before the agent stops backing the project (namely when $\sigma_q \leqslant \tau$), the agent collects the discounted unit

payoff, has paid the flow cost until time σ_q, and successfully leaves the project with the updated reputation function $w(\pi_{\sigma_q})$, where π_{σ_q} is the updated posterior belief reflecting the successful exit, so $\pi_{\sigma_q} = 1$.

On the other hand, if the agent stops the project before the success realization (namely when $\tau < \sigma_q$), then she has just paid the flow cost up until time τ, and leaves with the updated reputation function $w(\pi_\tau)$, reflecting the fact that the success has not happened until time τ. Therefore, the exit option at the stopping time τ is the agent's reservation value of holding reputation π_τ, i.e., $w(\pi_\tau)$.

Because of the dynamic programming principle, any C^1 value function of the above stopping time problem must satisfy the following HJB equation:

$$rv_q(\pi) = \max\left\{rw(\pi), -c + \lambda_q\pi\big(1 + w(1) - v_q(\pi)\big) - \lambda_q\pi(1 - \pi)v_q'(\pi)\right\}. \qquad (2.4)$$

The above HJB is presented in the variational form, that is the first expression in the *rhs* is the value of stopping — denying the project and holding on to the outside option w — and the second expression represents the Bellman equation over the continuation region $\mathcal{M}_q$, on which $v_q(\pi) > w(\pi)$. The first term in the Bellman equation is the flow cost of the project borne by the agent, the second term is the expected flow of the created surplus, and the last term captures the marginal reputation loss due to the lack of success.

Induced by the above stopping time problem, the matching set $\mathcal{M}$ can thus be interpreted as the continuation region for the free boundary problem (2.4), namely

$$\mathcal{M} = \{(q, \pi) \in \{a, b\} \times [0, 1] : v_q(\pi) > w(\pi)\}, \qquad (2.5)$$

and on the *stopping region* $\mathcal{M}^c$ (namely the complement of $\mathcal{M}$), the matching value function equals the agent's reputation function, i.e., $v_q(\pi) = w(\pi)$.

The overarching goal of this paper is to study the optimal outcome, which is the solution to the following fixed-point problem: the tuple $\langle w, v, \mathcal{M}\rangle$ constitutes an optimal outcome, if (i) given v and $\mathcal{M}$, the reputation value function w satisfies (2.2) and (ii) given w, the matching value function v and the matching set $\mathcal{M}$ together solve the free boundary system (2.3) and (2.5). I seek to find the C^1 optimal value functions.

The two-way feedback between the reputation function w and the matching variables $\langle v, \mathcal{M}\rangle$ are plotted in figure 2. The link connecting w to the $\langle v, \mathcal{M}\rangle$ block is upheld by

the stopping time problem (2.3). The opposite link from the matching variables block to w is supported by the Bellman equation for the reputation function in (2.2). The optimal outcome is formally the fixed-point to the endogenous loops of this figure.

Figure 2: Endogenous feedbacks

I should emphasize that in the optimal stopping time literature the exit option is usually exogenously set, and thus finding the optimal strategy only requires solving the free boundary problem. The main stretch in our setting is that the exit option itself is endogenously determined by the value function associated with the stopping time problem, and this elaborates the solution method.

3 Optimum as the Fixed-Point

In previous section, the optimal outcome was expressed as the fixed-point to the system of necessary conditions (2.2), (2.3) and (2.5).

Below in Section 3.1, I appeal to the fact that any C^1 solution to the stopping time problem (2.3) satisfies the Bellman equation (2.4). Additionally, it satisfies two other conditions known as the *majorant* and *superharmonic* properties. Hence, I initiate the search for the optimal tuple in the larger space of C^1 functions that satisfy the aforementioned two properties, as well as the system of conditions (2.2), (2.4) and (2.5).

Next, in Section 3.2, I show the predicted outcome (determined by the above necessary conditions and denoted by $\langle w^*, v^*, \mathcal{M}^* \rangle$) is unique. Then, I show this unique tuple is indeed the optimal outcome, that is replacing (2.3) with (2.4) was innocuous, and the pair $(v^*, \mathcal{M}^*)$ solves the stopping time problem in (2.3) given w^*.

3.1 Necessary Conditions

In this section, I first show the monotonicity of the matching value functions in q. That is to prove for any solution v to the stopping time problem (2.3), one has $v_b(\pi) \geqslant v_a(\pi)$

for all π. Second, I highlight another two necessary conditions, called the *majorant* and *superharmonic* properties, that the optimal value functions must satisfy.[5]

Proposition 3.1 (Monotonicity)**.** *Optimal matching value functions must satisfy* $v_b(\pi) \geqslant v_a(\pi)$ *for all* π.

Proof sketch. Since $\lambda_b > \lambda_a$, for every initial π the success arrives faster with a b-project. This means when deciding to match with a b-project, the agent can mimic the matching strategy of an a-type, and guarantees herself a payoff of at least $v_a(\pi)$. The proof is actually more subtle and is presented in the appendix.

Corollary 3.2. At the optimum $\mathcal{M}_a \subseteq \mathcal{M}_b$ and $w(\pi) = 0$ if and only if $v_b(\pi) = 0$.

Proof. The justification for $\mathcal{M}_a \subseteq \mathcal{M}_b$ is that $\pi \in \mathcal{M}_a$ implies $v_a(\pi) > w(\pi)$. Since $v_b(\pi) \geqslant v_a(\pi)$, then $v_b(\pi) > w(\pi)$, and hence $\pi \in \mathcal{M}_b$. In regard to the second claim, note that by equation (2.2) at the optimum w is a linear combination of v_a and v_b (with possibly zero weights). Since $v_b \geqslant v_a$ and both are non-negative, the second claim follows. $\qquad\square$

Corollary 3.3. At the optimum, $1 \in \mathcal{M}_a$ if and only if

$$\lambda_a - c > \frac{\kappa\varphi_b(\lambda_b - c)}{r + \kappa\varphi_b + \lambda_b}. \tag{3.1}$$

We say that the economy is in the *low cost regime* if the above inequality holds, and otherwise is in the *high cost regime*. In particular, it says at $\pi = 1$ — where the learning channel is absent — selecting an a-project is optimal if its payoff exceeds the opportunity cost, that is induced by waiting for a superior b-project.

One important analogy of this derivation with labor markets is that increasing the search frictions, e.g., by reducing κ here, lowers the opportunity cost of hiring low skilled individuals, and thus increases their chances of selection by reputation seeking employers.

Remark 3.4. It is noteworthy to mention that $w(1)$ itself is an endogenous object, that takes different forms depending on the cost regime. Its values in the high and low cost regimes are respectively expressed in equations (A.3) and (A.4) of the appendix in the proof of Corollary 3.3. To prevent additional notation, in the following, I use $w(1)$ to refer to both of these expressions, as the cost regime is clear from the context.

---

[5]These two conditions are standard in the literature of optimal stopping and can be found in Chapter 2 of Peskir and Shiryaev (2006).

Turning to the second group of necessary conditions, the dynamics of the reputation process can be compactly represented by

$$d\pi_t = (1 - \pi_{t-}) (d\iota_t - \lambda_q \pi_{t-} dt) ,$$

where ι is the success indicator process, that is $\iota_t := 1_{\{t \geqslant \sigma\}}$. The infinitesimal generator associated with this stochastic process is $\mathcal{L}_q : C^1[0,1] \to C[0,1]$, where for a generic $u \in C^1[0,1]$:[6]

$$[\mathcal{L}_q u] (\pi) = \lambda_q \pi \big(1 + w(1) - u(\pi)\big) - \lambda_q \pi (1 - \pi) u'(\pi).$$

For every candidate fixed-point tuple $\langle w, v, \mathcal{M} \rangle$ in the space of C^1 functions, that satisfy the system (2.2), (2.3) and (2.5) the following two conditions must hold for the optimal v and w at all $\pi \in [0,1]$ and $q \in \{a,b\}$:

(i) *Majorant property*: $v_q(\pi) \geqslant w(\pi)$.

(ii) *Superharmonic property*: $[\mathcal{L}_q v_q](\pi) - r v_q(\pi) - c \leqslant 0$.

The first condition simply means that in every match the agent has the option to terminate the project, thus enjoying her reputation value w. The second condition means *on expectation* a typical agent *loses* if she decides to keep the match on the stopping region.

Usually in the "one dimensional" experimentation settings, meaning where the continuation region is one dimensional, the agent follows threshold strategy and thus the continuation region is naturally a connected subset. However, in the current setting, where the matching set is "two-dimensional", consisting of two sections $\mathcal{M}_a$ and $\mathcal{M}_b$, one may expect a situation in which one of these subsets contains two disjoint intervals, and hence not be connected. In the next three results, using Proposition 3.1 and the above two conditions, I will rule out this possibility and show that both sections of the matching set are connected intervals.

Lemma 3.5 (Lowest boundary point). *Let $\beta := \inf \mathcal{M}_b$, that is the lowest boundary point of the high type section of the optimal matching set. Then,*

$$\beta = \frac{c}{\lambda_b \big(1 + w(1)\big)}. \tag{3.2}$$

---

[6]Space of continuously differentiable functions on $(0,1)$ with continuous extension to the boundary $\{0,1\}$.

Proof. Since the value functions are continuous, $\mathcal{M}_b$ is an open subset, and hence $\beta \notin \mathcal{M}_b$. Corollary 3.2 implies that $w(\pi) = 0$ for all $\pi \leqslant \beta$. This means the matching value function v_b must smoothly meet the zero function at β, i.e., $v_b(\beta) = v_b'(\beta) = 0$. Inserting this into the Bellman equation (2.4) yields (3.2). $\qquad\square$

Next lemma shows that at the optimum, $\mathcal{M}_b$ is an *increasing interval*. That is if $\pi \in \mathcal{M}_b$, then $\pi' \in \mathcal{M}_b$ for all $\pi' > \pi$. To show this claim, suppose to the contrary that $\exists \pi' > \pi$ such that $\pi' \notin \mathcal{M}_b$. Then, Corollary 3.2 implies tha $v_b(\pi') = 0$, whereas $v_b(\pi) > 0$ because $\pi \in \mathcal{M}_b$. This combination will be ruled out in the next lemma, thus proving that $\mathcal{M}_b$ is an increasing interval.

Lemma 3.6 (Single crossing). *Let v_b be the optimal matching value function in C^1. If $v_b(\pi) > 0$, then $v_b(\pi') > 0$ for all $\pi' > \pi$.*

Proof. Assume there exists $\pi' > \pi$ such that $v_b(\pi') = 0$. Since $\pi \in \mathcal{M}_b$, then clearly $\beta < \pi < \pi'$. Also, since $v_b \geqslant 0$, then π' is a global minimum and thus $v_b'(\pi') = 0$. Because of superharmonic property at π', one has

$$0 \geqslant \mathcal{L}_b v_b(\pi') - r v_b(\pi') = \lambda_b \pi' \big(1 + w(1)\big) - c\,,$$

where the equality holds because $v_b(\pi') = v_b'(\pi') = 0$. This in turn implies that $\pi' \leqslant c / \lambda_b \big(1 + w(1)\big) = \beta$, that is a contradiction. $\qquad\square$

As argued previously, the above lemma shows that at the optimum $\mathcal{M}_b$ is the increasing interval $(\beta, 1]$, and hence is unique. That in turn leaves only one candidate for the optimal v_b, that satisfies the Bellman equation (2.4) on the continuation region $(\beta, 1]$. A particular solution for this Bellman differential equation is

$$-\frac{c}{r} + \frac{\lambda_b}{r + \lambda_b} \left(1 + w(1) + \frac{c}{r}\right) \pi\,,$$

and the homogenous solution is $(1 - \pi)^{1 + r/\lambda_b} \pi^{-r/\lambda_b}$. Since $v_b(\beta) = v_b'(\beta) = 0$, the only candidate for the optimal v_b is

$$v_b^*(\pi) = -\frac{c}{r} + \frac{\lambda_b}{r + \lambda_b} \left(1 + w(1) + \frac{c}{r}\right) \pi$$

$$+ \left(\frac{c}{r} - \frac{\lambda_b}{r + \lambda_b} \left(1 + w(1) + \frac{c}{r}\right) \beta\right) \left(\frac{1 - \pi}{1 - \beta}\right)^{1 + r/\lambda_b} \left(\frac{\pi}{\beta}\right)^{-r/\lambda_b}. \tag{3.3}$$

15

Exploiting the above characterization as the only viable candidate for the optimal v_b (in any C^1 fixed-point outcome) and the fact that $\mathcal{M}_a \subseteq \mathcal{M}_b$, I prove in the following proposition that $\mathcal{M}_a$ is also an increasing interval — especially, it means $\mathcal{M}_a$ cannot have disjoint subsets. Its proof includes multiple steps, thus relegated to the appendix.

Proposition 3.7 (Optimal $\mathcal{M}_a$). *In the low cost regime, optimal $\mathcal{M}_a$ is an increasing interval, i.e., $\mathcal{M}_a = (\alpha, 1]$ for some $\alpha \geqslant \beta$. And in the high cost regime $\mathcal{M}_a = \varnothing$.*

This proposition implies that in the high cost regime $v_a^* = \frac{\kappa \varphi_b}{r + \kappa \varphi_b} v_b^*$, and $\mathcal{M}_a^* = \varnothing$. In the low cost regime, however, $\mathcal{M}_a^* = (\alpha, 1]$ and

$$v_a^*(\pi) = -\frac{c}{r} + \frac{\lambda_a}{r + \lambda_a} \left(1 + w(1) + \frac{c}{r} \right) \pi + \gamma \left(\frac{1 - \pi}{1 - \alpha} \right)^{1 + r/\lambda_a} \left(\frac{\pi}{\alpha} \right)^{-r/\lambda_a} , \qquad (3.4)$$

in that the coefficient γ and the lower boundary point α are determined by the following boundary conditions:

$$v_a^*(\alpha) = \frac{\kappa \varphi_b}{r + \kappa \varphi_b} v_b^*(\alpha) \text{ and } v_a^{*\prime}(\alpha) = \frac{\kappa \varphi_b}{r + \kappa \varphi_b} v_b^{*\prime}(\alpha). \qquad (3.5)$$

3.2 Uniqueness and Martingale Verification

The characterizations in the previous section essentially offered a unique tuple as the only viable candidate satisfying (2.2), (2.4), (2.5) as well as the majorant and superharmonic properties. In the first theorem below, I summarize the properties of this tuple.

Theorem 3.8 (Uniqueness). *The following profile expresses the unique C^1 value functions and the matching sets, that satisfy the fixed-point conditions (2.2), (2.4), (2.5) as well as the majorant and superharmonic properties:*

(i) In each cost regime $\mathcal{M}_b^* = (\beta, 1]$, where β is determined by Lemma 3.5. Additionally, v_b^* follows (3.3).

(ii) In the high cost regime $\mathcal{M}_a^* = \varnothing$ and $v_a^* = \frac{\kappa \varphi_b}{r + \kappa \varphi_b} v_b^*$. In the low cost regime $\mathcal{M}_a^* = (\alpha, 1]$, where α is determined by (3.5), and v_a^* follows (3.4).

All the claims in this theorem, except a complete verification of the superharmonic property (especially outside of the continuation region), were justified in the previous section. Therefore, it only remains to establish the superharmonic property in the appendix.

16

The next step is to show the unique tuple expressed in the previous theorem is indeed the best response of the agent. Formally, one needs to prove that given w^*, the pair $(v^*, \mathcal{M}^*)$ describe the optimal value function and the optimal continuation region for the stopping time problem of (2.3). In the next theorem, I will apply a Martingale verification procedure to show this step.

Theorem 3.9 (Unique optimum). $\langle w^*, v^*, \mathcal{M}^* \rangle$ *is the unique optimal tuple in the space of C^1 value functions, satisfying conditions* (2.2), (2.3) *and* (2.5).

Figure 3 plots the optimal value functions in the low cost regime. In particular, it demonstrates the convexity of the value functions, and shows that at the the optimum $\alpha > \beta$. Relatedly, Figure 4 plots the optimal matching sets $\mathcal{M}^*$ in each cost regime. As explained before, at the optimum $\mathcal{M}_a^* \subseteq \mathcal{M}_b^*$, and in the high cost regime $\mathcal{M}_a^* = \varnothing$. In light of $\mathcal{M}_a^* \subseteq \mathcal{M}_b^*$, the model offers the testable prediction that the agents who exit the economy and do not engage in further activities made their last few engagements in the high-growth projects (i.e., b-types).[7]

The agent follows cutoff strategy with respect to each type of the projects, and in particular, she shows more tolerance for failure when matched to the high type projects. The threshold strategy (equivalently, that the matching sets are increasing intervals) advances the idea that agents with higher reputation have higher tolerance for failure. In other words, the distance to the endogenous separation point (α or β) is larger for a more reputable agent than a less reputable one. This observation is in line with the *learning theory* in economics of venture capital. Specifically, Gompers and Lerner (1999) argue that VCs learn about their *post-investment ability* while they are funding startups, and the more reputable ones have higher tolerance for failure, namely they spend longer time funding their portfolio companies.[8]

Specifically, it was shown in Lemma 3.5 that the endogenous termination point β is inversely related to $w(1)$, where $w(1)$ is the value of holding the maximum reputation,

---

[7]This prediction is in line with the observation that historically the politicians who lose big races for high office elections, become backbencher for a while. I acknowledge the anonymous referee for suggesting this anecdotal evidence.

[8]Also related to the economics of venture capital, my model suggests a method to endogenize the *tolerance for failure* (Tian and Wang, 2014; Manso, 2011) by relating it to the agent's reputation. Specifically, in Tian and Wang (2014) VCs learn about the quality of the startup over the course of the match, whereas reflecting in my model the startup's quality is observable and the learning is about the VC's ability. Consequently, the approach here suggests one way to endogenize the *tolerance parameter* in Tian and Wang (2014).
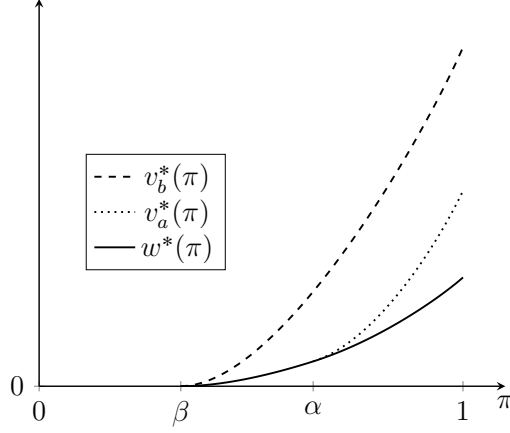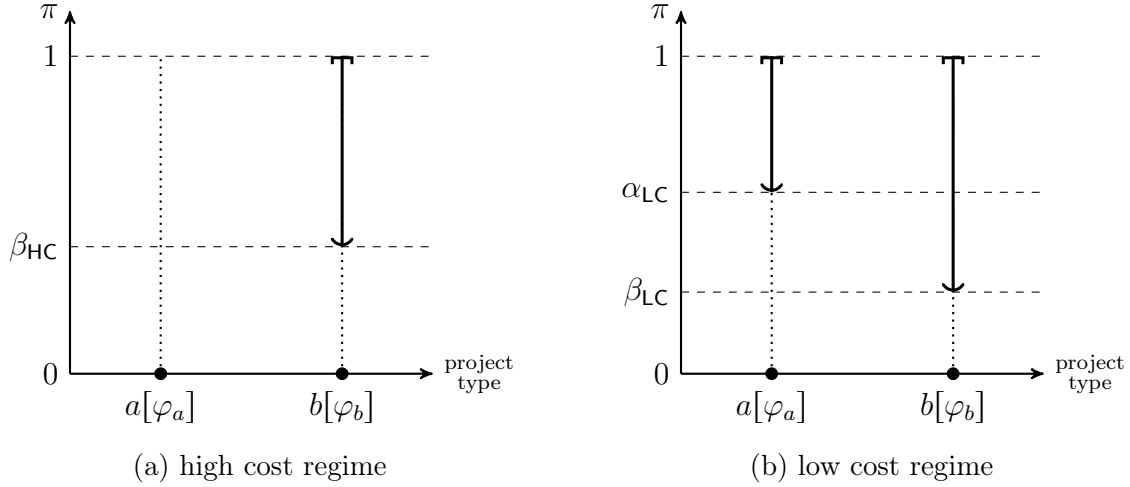
Figure 3: Value functions in the low cost regime

(a) high cost regime

(b) low cost regime

Figure 4: Optimal matching sets

namely at $\pi = 1$, in each cost regime. In the high cost regime $w(1)$ only depends on the b-parameters, because $\mathcal{M}_a^* = \varnothing$, whereas in the low cost regime it takes the a-related parameters into account as well (see Remark 3.4). Additionally, equation (3.2) shows that it is indeed through the reputation channel (i.e., $w(1)$) that learning incentives manifest themselves in the agent's selection policy. Specifically, any of the exogenous parameters of the economy can affect $w(1)$, that in turn alters the size of the continuation set.

In the next section, I compare the current learning model with its no-learning version, and perform comparative statics with respect to the underlying primitives, in particular the meeting rate (or inversely, the search frictions).

4 Qualitative Comparisons

In Section 4.1 below, I show how the shape of the matching sets and value functions change in an economy, where agent has complete information about her type, but is otherwise the same as before. This exercise helps us to uncover the unique role of learning and incomplete information in the agent's optimal selection policy.

Next, in Section 4.2, I perform the comparative statics of the optimal matching sets (in the original learning model) with respect to the primitives of the economy.

4.1 No-Learning Version

In contrast with our original model, where the agent's underlying type was the hidden binary variable $\theta \in \{L, H\}$, and π reflected the posterior belief, here I assume the actual underlying type is $\pi \in [0, 1]$ and it stays *constant* over time. Specifically, when a type-π agent selects a type-q project, the success arrives with the rate of $\lambda_q \pi$. The underlying reason behind studying this benchmark case is to understand how the learning process impacts the optimal matching sets.

The major changes happen in the Bellman equation for the matching value function. First, the Bayesian learning component that includes the π-derivative of $v_q(\pi)$ is no longer present. Second, the exit option at the time of success is $1 + w(\pi)$ instead of $1 + w(1)$. This is owed to the fact that the agent's type is persistent and she leaves the match with the same reputation that she entered. Formally, the no-learning Bellman equation for the matching value function is:

$$rv_q(\pi) = \max \left\{ rw(\pi), -c + \lambda_q \pi \big(1 + w(\pi) - v_q(\pi)\big) \right\} . \tag{4.1}$$

The expressions behind the reputation function w and $\mathcal{M}$ remain the same as in (2.2) and (2.5), respectively.

Proposition 4.1 (Unique optimum, absent of learning)**.** *There exists a unique optimal tuple* $\langle \hat{w}, \hat{v}, \widehat{M} \rangle$*, in the space of continuous value functions, that satisfy the optimality conditions* (2.2)*,* (4.1) *and* (2.5)*. Furthermore,*

(i) *In both cost regimes the matching sets are increasing intervals and* $\widehat{\mathcal{M}}_a \subseteq \widehat{\mathcal{M}}_b$*.*

(ii) *In the high cost regime* $\widehat{\mathcal{M}}_a = \varnothing$*.*

The lower boundary of optimal matching sets (in the proof of Proposition 4.1) are denoted by $\hat{\alpha} = \inf \widehat{\mathcal{M}}_a$, and $\hat{\beta} = \inf \widehat{\mathcal{M}}_b$. It is shown in the appendix that $\hat{\beta} = c/\lambda_b$ and $\hat{\alpha}$ follows equation (A.16).

Comparing $\hat{v}$ with v^*:   The optimal matching value functions in the current no-learning environment (and in the low cost regime) are plotted in Figure 5. There are two important differences with Figure 3: local concavity and kinks on the boundary of matching sets. In contrast, the value functions in the learning environment were convex and smooth. Both of these properties were due to the Bayesian learning, that are absent here.
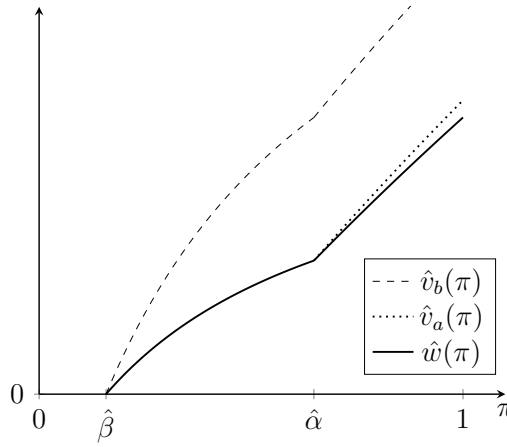

Figure 5: Value functions in the low cost regime (no-learning version)

Comparing $\hat{\beta}$ with β: It is shown in the proof of Proposition 4.1 that $\hat{\beta} = c/\lambda_b$. Comparing this with equation (3.2) for β, one notices an important difference: learning incentives affect β through the impact of $w(1)$ in its denominator. Specifically, increasing κ or φ_b, or decreasing r each strengthens the reputational motives and raises $w(1)$, thus making the agent more patient (by lowering β). In the absence of learning, all of these effects are muted in $\hat{\beta}$. Hence, in both cost regimes, the separation point β is smaller than its no-learning counterpart $\hat{\beta} = c/\lambda_b$. Therefore, the prospects of learning about the self-type and possibly reaching a higher reputation expand the matching sets and add more patience to the agent's continuation region. I refer to this force by *reputational effect* in the next section.

20

Response of $\hat{\alpha}$ to κ: Differentiating the expression for $\hat{\alpha}$ in equation (A.16) (of the appendix) with respect to κ gives that $\mathrm{d}\hat{\alpha}/\mathrm{d}\kappa > 0$, thus decreasing the search frictions shrinks $\widehat{\mathcal{M}}_a$. Hypothetically, in a frictionless world (where $\kappa \to \infty$) the agent never selects the a-projects, because its opportunity cost is effectively infinite (as she can immediately select a superior b-project). In reality however, search frictions create an endogenous wedge, by lowering the aforementioned opportunity cost, and partially tilt the incentives toward the inferior a-projects. I refer to this force by *opportunity cost effect* in the next section. This has the same frictional spirit, by which the low-skilled individuals are selected by the employers in the labor market. Hence, it is exactly in this sense that increasing the search frictions (by lowering the meeting rate κ) expands $\widehat{\mathcal{M}}_a$. Through an example in the next section, I show that this monotone response is overturned in the original learning model of Section 3 — due to the opposing force created by the reputational effect.

4.2 Comparative Statics

The results of this section pertains to the original model with learning. Observe that in both cost regimes $\mathcal{M}_a^* \subseteq \mathcal{M}_b^*$, and thus I take β as a proxy for the size of the union of matching sets, i.e., $\mathcal{M}_a^* \cup \mathcal{M}_b^*$. It is important to know the comparative statics of β (in (3.2)) with respect to the primitives of the economy.

Performing simple differentiation of (3.2), one can easily verify that $\frac{\mathrm{d}\beta}{\mathrm{d}c} > 0$, $\frac{\mathrm{d}\beta}{\mathrm{d}r} > 0$, and $\frac{\mathrm{d}\beta}{\mathrm{d}\kappa} < 0$. Namely, lower levels of flow cost, time discount rate, and search frictions (equivalently higher meeting rate) are all associated with larger $\mathcal{M}_b^*$. Specifically, raising the meeting rate κ increases the value of holding the maximum reputation $w(1)$ — because the agent meets the projects more frequently — and thus expands the optimal $\mathcal{M}_b^*$. So, as far as it relates to the impact of κ on β, it is only the reputational effect that plays a role. The opportunity cost effect has no impact on $\mathcal{M}_b^*$, as there is no better alternative than b-projects.

In the following example, I study how the optimal α (i.e., the lower boundary of $\mathcal{M}_a^*$) reacts *non-monotonically* to the search frictions. This is in contrast with the monotone response of $\hat{\alpha}$ to κ in the no-learning version (studied above).

Example 4.2 (Non-monotone response of α to κ.)**.** In this example, I show — in the low cost regime where $\mathcal{M}_a^* = (\alpha, 1] \neq \varnothing$ — there exists a range of parameters, in which the optimal α reacts non-monotonically to κ. This is in contrast with the response of its

no-learning counterpart $\hat{\alpha}$ to κ, that was shown to be unambiguously increasing due to the opportunity cost effect.

First, I explain how one can mathematically pin the down the fixed-point α, and then I argue (based on the properties of the fixed-point mapping) why the response is not monotone. Observe that in the low cost regime, α is the point at which v_a^* smoothly meets the reservation value w^*. Due to the specification in Theorem 3.8, $\alpha \in \mathcal{M}_a^{*c} \cap \mathcal{M}_b^*$, and one has

$$v_a^*(\pi) = w^*(\pi) = \frac{\kappa \varphi_b}{r + \kappa \varphi_b} v_b^*(\pi), \quad \forall \pi \leqslant \alpha.$$

Therefore, the boundary conditions in (3.5) apply. Using these conditions and the Bellman equations for v_a^* and v_b^*, one arrives at the following relation, whose fixed-point determines the optimal α:

$$\alpha = \frac{r \lambda_b c + \kappa \varphi_b (\lambda_b - \lambda_a)(c + r v_b^*(\alpha))}{r \lambda_b \lambda_a (1 + w(1))}. \tag{4.2}$$

Next, I plugged in the closed-form expression for v_b^* from (3.3) into the above relation. By varying κ, I found the fixed-point α (as a function of κ) in a numerical example whose output is plotted in Figure 6. As it appears the response is U-shaped: for small values of κ, the optimal α is decreasing, while for larger κ, it becomes increasing. One should contrast this outcome with the no-learning counterpart, in which $\frac{d\hat{\alpha}}{d\kappa} > 0$, and with the lower boundary of $\mathcal{M}_b^*$, where $\frac{d\beta}{d\kappa} < 0$.

The algebraic reason behind the U-shaped response of α to κ is that both the numerator and the denominator of the fixed-point map (4.2) are increasing in κ, therefore, the overall response is ambiguous. Intuitively however, the ratio in (4.2) highlights two opposing forces that underlie the non-monotone behavior: reputational effect and opportunity cost effect. First and similar to the case for $\mathcal{M}_b^*$, raising κ increases the value of holding the the maximum reputation $w(1)$, and this encourages the agent to stay longer with the project. This reputational effect (that originates from the learning incentives) manifests itself in the denominator of (4.2), and sets an expanding force on $\mathcal{M}_a^*$. Second and similar to the case for $\widehat{\mathcal{M}}_a$, higher κ raises the opportunity cost of choosing an a-project, thus shrinking the optimal $\mathcal{M}_a^*$. This effect is playing out in the numerator of (4.2). As it appears from Figure 6, the reputational effect dominates for small levels of the meeting rate, while as κ increases, it is the opportunity cost effect that prevails and causes $\mathcal{M}_a^*$ to shrink.
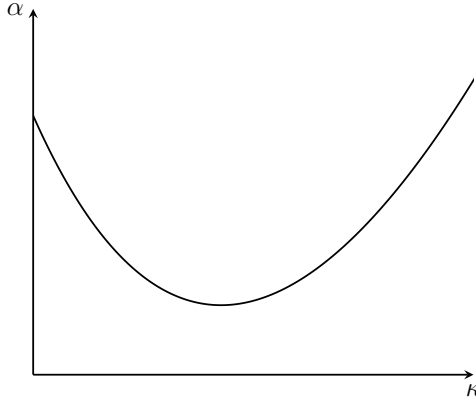
Figure 6: Response of the optimal $\mathcal{M}_a^*$ to κ

Remark 4.3. Even though the above observation on the non-monotone response was stated as an example, but by continuity it holds for an open region. A formal argument establishing this non-monotonicity is out of reach. First, because the fixed-point mapping in (4.2) is not monotone in the parameter κ, the monotone comparative statics apparatus cannot be applied to show the behavior in Figure 6. Second, implicitly differentiating both sides of (4.2) with respect to κ, and showing that $\frac{d\alpha}{d\kappa}$ is negative for small κ and positive for large κ is also intractable.

The important policy lesson behind this final comparative statics of $\mathcal{M}_a^*$ with respect to κ is that increasing the meeting rate between the two sides of the economy in the hope of achieving higher surplus is not always socially optimal. Specifically, imagine an economy where there are spillovers from successful low type projects (here the a-types) to the creation of high type opportunities (here the b-projects) — as is common in the innovation literature, where small low growth accomplishments create high growth opportunities. There are empirical evidences (Lerner et al., 2005) suggesting that small innovative firms are particularly weak in protecting their intellectual property and thus their investors do not internalize the spillover gains in their decisions.

In such circumstances, policies aimed at reducing the search frictions are initially helpful (as they increase the incentives to invest in small projects by amplifying the reputational incentives and thus expanding the region for $\mathcal{M}_a^*$), but eventually backfire and shrink the investment region for the low growth projects (due to the domination of the opportunity

cost effect as the search frictions decrease further).[9]

5   Concluding Remarks

I study the optimal project selection policy of an agent with unknown ability. The agent randomly meets the projects drawn from a heterogeneous pool, that differ in their quality. In a match between the agent and a project a breakthrough arrives at the exponential rate depending on the type of the agent and the quality of the project. Since maintaining the projects are costly, the agent effectively solves a stopping time problem, in which she weighs the expected benefit of learning about her type as well as accomplishing breakthroughs against the endogenous reservation function (that is called the reputation value function in the paper).

The matching sets indicate what type of projects an agent with certain level of reputation is willing to accept or continue the match with. In the space of continuously differentiable functions, I show there exists a unique optimum. Sections of the optimal matching set are increasing intervals, thus the agent follows cutoff strategies at the optimum. The thresholds depend on the type of the projects and are endogenously determined. They encode a number of messages. For example, lower levels of flow cost and time discount rate are associated with larger optimal matching sets. Additionally, it is shown raising the meeting rate (or lowering the search frictions) has asymmetric effects across the two types of the projects: it unambiguously expands the high type section of the matching set, while on some regions it initially expands and then shrinks the low type section.

Compared to the no-learning benchmark (where there is no incomplete information about the agent's type), the optimal continuation sets are larger, therefore the agent shows more patience before stopping the projects. This is owed to the convexity of the value functions in reputation, that itself is resulted from the learning incentives in the agent's dynamic problem.

---

[9]The curious reader can read about the failure of the public attempts to boost the Biotech sector in Malaysia by reducing the search frictions via the creation of a centralized hub for investing in early stage projects; see Lerner (2002).

A Proofs

A.1 Proof of Proposition 3.1

Suppose the agent with reputation π is approached by an a-project. Her optimal strategy is to match with the project so long as her reputation is above some level $\pi_0 \leqslant \pi$, i.e., the threshold rule. The case of $\pi_0 = \pi$ simply means the agent rejects the project. Let π_t^q represent the deterministic solution to equation (2.1), when matched to a q-project. Define t_q as the deterministic time at which this solution crosses the threshold π_0, namely:

$$t_q := \inf\{t \geqslant 0 : \pi_t^q = \pi_0\} .$$

Because of Bayes law, we have

$$\frac{\pi_0}{1 - \pi_0} = \frac{\pi}{1 - \pi} e^{-\lambda_q t_q} .$$

Therefore, $\lambda_a t_a = \lambda_b t_b$, that in turn means $\mathsf{P}(\sigma_a > t_a) = \mathsf{P}(\sigma_b > t_b)$, in that σ_q was defined as the exponential time of the breakthrough in a q-match. By the optimality of π_0 as a cutoff strategy for an a-match, one has:

$$
\begin{aligned}
v_a(\pi) = \mathsf{E}&\left[e^{-r\sigma_a} - c \int_0^{\sigma_a} e^{-rs} \mathrm{d}s + e^{-r\sigma_a} w(\pi_{\sigma_a}); \sigma_a \leqslant t_a \right] \\
+\mathsf{E}&\left[-c \int_0^{t_a} e^{-rs} \mathrm{d}s + e^{-rt_a} w(\pi_{t_a}^a); \sigma_a > t_a \right] .
\end{aligned}
$$

Since $\pi_{\sigma_a} = 1$, $\pi_{t_a}^a = \pi_0$ and t_a is a deterministic time, then

$$
\begin{aligned}
v_a(\pi) = \mathsf{E}&\left[e^{-r\sigma_a} - c \int_0^{\sigma_a} e^{-rs} \mathrm{d}s + e^{-r\sigma_a} w(1); \sigma_a \leqslant t_a \right] \\
&+ \left(-c \int_0^{t_a} e^{-rs} \mathrm{d}s + e^{-rt_a} w(\pi_0) \right) \mathsf{P}\left(\sigma_a > t_a \right) .
\end{aligned}
\tag{A.1}
$$

Recall that $v_b(\pi)$ is the *optimal* matching value function when agent is approached by a b-project, therefore, choosing (the deterministic) t_b as a stopping time when backing a

b-project leads to a weakly smaller payoff. That is by equation (2.3) it holds that

$$v_b(\pi) \geqslant \mathsf{E}\left[e^{-r\sigma_b} - c\int_0^{\sigma_b} e^{-rs}\mathrm{d}s + e^{-r\sigma_b}w(\pi_{\sigma_b}); \sigma_b \leqslant t_b\right]$$

$$+ \mathsf{E}\left[-c\int_0^{t_b} e^{-rs}\mathrm{d}s + e^{-rt_b}w(\pi_{t_b}^b); \sigma_b > t_b\right].$$

By similar reasoning, one obtains that

$$v_b(\pi) \geqslant \mathsf{E}\left[e^{-r\sigma_b} - c\int_0^{\sigma_b} e^{-rs}\mathrm{d}s + e^{-r\sigma_b}w(1); \sigma_b \leqslant t_b\right]$$

$$+ \left(-c\int_0^{t_b} e^{-rs}\mathrm{d}s + e^{-rt_b}w(\pi_0)\right)\mathsf{P}\left(\sigma_b > t_b\right). \tag{A.2}$$

Next, I compare the *rhs* of equations (A.1) and (A.2). First, observe that

$$\mathsf{E}\left[e^{-r\sigma_b} - c\int_0^{\sigma_b} e^{-rs}\mathrm{d}s + e^{-r\sigma_b}w(1); \sigma_b \leqslant t_b\right]$$

$$= \mathsf{P}(\sigma_b \leqslant t_b)\,\mathsf{E}\left[e^{-r\sigma_b} - c\int_0^{\sigma_b} e^{-rs}\mathrm{d}s + e^{-r\sigma_b}w(1)\,\middle|\,\sigma_b \leqslant t_b\right].$$

One can easily verify that since $\lambda_b > \lambda_a$ and $\lambda_a t_a = \lambda_b t_b$, the conditional distribution $(\sigma_a \mid \sigma_a \leqslant t_a)$ first order stochastically dominates the conditional distribution $(\sigma_b \mid \sigma_b \leqslant t_b)$. The expression inside the conditional expectation above is a decreasing function in σ, therefore,

$$\mathsf{E}\left[e^{-r\sigma_b} - c\int_0^{\sigma_b} e^{-rs}\mathrm{d}s + e^{-r\sigma_b}w(1)\,\middle|\,\sigma_b \leqslant t_b\right] \geqslant$$

$$\mathsf{E}\left[e^{-r\sigma_a} - c\int_0^{\sigma_a} e^{-rs}\mathrm{d}s + e^{-r\sigma_a}w(1)\,\middle|\,\sigma_a \leqslant t_a\right].$$

Since $\mathsf{P}(\sigma_b \leqslant t_b) = \mathsf{P}(\sigma_a \leqslant t_a)$, the first terms on the *rhs* of (A.1) and (A.2) compare as:

$$\mathsf{E}\left[e^{-r\sigma_b} - c\int_0^{\sigma_b} e^{-rs}\mathrm{d}s + e^{-r\sigma_b}w(1); \sigma_b \leqslant t_b\right] \geqslant$$

$$\mathsf{E}\left[e^{-r\sigma_a} - c\int_0^{\sigma_a} e^{-rs}\mathrm{d}s + e^{-r\sigma_a}w(1); \sigma_a \leqslant t_a\right].$$

Regarding the second terms, observe that $\mathsf{P}(\sigma_b > t_b) = \mathsf{P}(\sigma_a > t_a)$ and $t_b < t_a$, hence

$$
\left(-c \int_0^{t_b} e^{-rs} \mathrm{d}s + e^{-rt_b} w(\pi_0) \right) \mathsf{P}\left(\sigma_b > t_b \right) \geq
$$
$$
\left(-c \int_0^{t_a} e^{-rs} \mathrm{d}s + e^{-rt_a} w(\pi_0) \right) \mathsf{P}\left(\sigma_a > t_a \right) .
$$

The previous two inequalities jointly imply that $v_b(\pi) \geq v_a(\pi)$, thus proving Proposition 3.1.

$\square$

A.2 Proof of Corollary 3.3

At $\pi = 1$, there is no learning and hence the Bellman equation in (2.4) reduces to

$$
v_q(1) = \max \left\{ w(1), \frac{\lambda_q - c}{r + \lambda_q} + \frac{\lambda_q w(1)}{r + \lambda_q} \right\} .
$$

This implies $1 \in \mathcal{M}_q$ if and only if $\lambda_q - c > rw(1)$. Since $\lambda_b > c$ and $\mathcal{M}_a \subseteq \mathcal{M}_b$, then $1 \in \mathcal{M}_b$ always. Let w_b be the reservation value function in an outcome where $1 \notin \mathcal{M}_a$, then according to the other leg of the fixed-point system, i.e., equation (2.2), it must be that

$$
rw_b(1) = \frac{\kappa \varphi_b \left(\lambda_b - c \right)}{r + \kappa \varphi_b + \lambda_b} . \tag{A.3}
$$

Hence, $1 \notin \mathcal{M}_a$ implies that $\lambda_a - c \leq rw_b(1)$.

Conversely, assume $1 \in \mathcal{M}_a$ and let $w_{ab}(1)$ be the reservation value function in this outcome, where $1 \in \mathcal{M}_a$. Specifically, one obtains

$$
rw_{ab}(1) = \frac{\kappa \varphi_b \left(\lambda_b - c \right) \left(r + \lambda_a \right) + \kappa \varphi_a \left(\lambda_a - c \right) \left(r + \lambda_b \right)}{\left(r + \lambda_a \right) \left(r + \lambda_b \right) + \kappa \varphi_b \left(r + \lambda_a \right) + \kappa \varphi_a \left(r + \lambda_b \right)} . \tag{A.4}
$$

Then, $1 \in \mathcal{M}_a$ means $\lambda_a - c > rw_{ab}(1)$. Also, because of optimality in equation (2.2), one has $w_{ab}(1) > w_b(1)$, hence it must be that $\lambda_a - c > rw_b(1)$.

$\square$

A.3 Proof of Proposition 3.7

To prove this proposition, I will first show that $\mathcal{M}_a$ is always an interval, meaning that it is always connected. By Corollary 3.3, $1 \in \mathcal{M}_a$ in the low cost regime. So the following lemma already establishes that $\mathcal{M}_a$ must be an increasing interval in the low cost regime.

Lemma A.1. *In both cost regimes (low and high) the optimal $\mathcal{M}_a$ is an interval.*

Proof. Let us define $\mathcal{D}_a v_a := \mathcal{L}_a v_a - r v_a - c$. Since v_a belongs to C^1, then $\mathcal{D}_a v_a$ is continuous. In addition, superharmonicity implies that $\mathcal{D}_a v_a(\pi) \leqslant 0$ for all π, and particularly, $\mathcal{D}_a v_a(\pi) = 0$ on $\mathcal{M}_a$ by the Bellman equation. Suppose $\pi \in \mathcal{M}_a^c \cap \mathcal{M}_b$, then

$$v_a(\pi) = w(\pi) = \frac{\kappa \varphi_b}{r + \kappa \varphi_b} \, v_b^*(\pi) \, .$$

Therefore, the Bellman equation for v_b^* implies that

$$\mathcal{D}_a v_a(\pi) = \frac{-\kappa \varphi_b}{r + \kappa \varphi_b} (\lambda_b - \lambda_a) \frac{r v_b^*(\pi) + c}{\lambda_b} + \frac{r \lambda_a \pi \big(1 + w(1)\big) - cr}{r + \kappa \varphi_b} \, .$$

The unique characterization for v_b^* in (3.3) is twice differentiable. Since $\lambda_b > c$, it is easy to verify that $v_b^{*''} \geqslant 0$, and especially $v_b^{*''} > 0$ on $\mathcal{M}_b$. Hence, for $\pi \in \mathcal{M}_a^c \cap \mathcal{M}_b$ the above expression implies

$$\frac{\mathrm{d}^2}{\mathrm{d}\pi^2} \mathcal{D}_a v_a(\pi) = \frac{-\kappa \varphi_b (\lambda_b - \lambda_a)}{(r + \kappa \varphi_b) \lambda_b} \, v_b^{*''}(\pi) < 0 \, .$$

Therefore, $\mathcal{D}_a v_a$ is strictly concave on every connected subset of $\mathcal{M}_a^c \cap \mathcal{M}_b$. Now assume by contradiction that $\mathcal{M}_a$ is not connected. Thus, it shall contain two disjoint maximal open intervals, say (π_1, π_2) and (π_3, π_4), where $\pi_2 < \pi_3$. Since $\mathcal{M}_b$ is an increasing interval containing $\mathcal{M}_a$ — respectively, by Lemma 3.6 and Corollary 3.2 — it must be that $[\pi_2, \pi_3] \subset \mathcal{M}_a^c \cap \mathcal{M}_b$. The previous analysis means that $\mathcal{D}_a v_a = 0$ on $(\pi_1, \pi_2) \cup (\pi_3, \pi_4)$, and $\mathcal{D}_a v_a$ is strictly concave in between, i.e., on $[\pi_2, \pi_3]$. Thus, continuity of $\mathcal{D}_a v_a$ implies that it is positive on $[\pi_2, \pi_3]$, violating the superharmonicity, and thus proving the lemma. $\qquad \square$

The following two lemmas are aimed at proving $\mathcal{M}_a = \varnothing$ in the high cost regime. In the first one, I show a characterization for the optimal matching set that only hinges on the optimal matching value functions v. Borrowing that in the second lemma, I show $\mathcal{M}_a$ cannot have a lower boundary point in $\mathcal{M}_b$. Thus, in light of $\mathcal{M}_a \subseteq \mathcal{M}_b$, one can conclude that $\mathcal{M}_a = \varnothing$. Lastly, in both of these lemmas v_b is equal to the optimal v_b^* — found uniquely in (3.3) — but the $*$ superscript is dropped for the sake of brevity.

Lemma A.2. *At the optimum, $\pi \in \mathcal{M}_a \cap \mathcal{M}_b$ if and only if*

$$\frac{\kappa \varphi_a}{r + \kappa \varphi_a} < \frac{v_b(\pi)}{v_a(\pi)} < \frac{r + \kappa \varphi_b}{\kappa \varphi_b} \, . \tag{A.5}$$

In addition, $\pi \in \mathcal{M}_a^c \cap \mathcal{M}_b$ if and only if the second inequality above binds.

Proof. An equivalent representation for equation (2.2) is

$$w(\pi) = \frac{\kappa\big(\varphi_a v_a(\pi)\chi_a(\pi) + \varphi_b v_b(\pi)\chi_b(\pi)\big)}{r + \kappa\big(\varphi_a \chi_a(\pi) + \varphi_b \chi_b(\pi)\big)}. \tag{A.6}$$

One can check that if the inequality chain (A.5) holds, then with $\chi_a(\pi) = \chi_b(\pi) = 1$ in the above representation, both of the conditions $v_a(\pi) > w(\pi)$ and $v_b(\pi) > w(\pi)$ are satisfied, and hence the *if* part is established. For the *only if* direction, assume $\pi \in \mathcal{M}_a \cap \mathcal{M}_b$, then it must be that $\chi_a(\pi) = \chi_b(\pi) = 1$. Replacing this in (A.6) and simplifying $v_b(\pi) > w(\pi)$ results in the first inequality in (A.5). Similarly, simplifying $v_a(\pi) > w(\pi)$ leads to the second inequality in (A.5). The proof of the last claim in the lemma follows the same logic. $\qquad\square$

Lemma A.3. *Suppose $\mathcal{M}_a$ and $\mathcal{M}_b$ are the optimal matching sets in the high cost regime. Then, $\mathcal{M}_a$ cannot have a lower boundary point in $\mathcal{M}_b$.*

Proof. Assume by contradiction that $x := \inf \mathcal{M}_a$ belongs to $\mathcal{M}_b$. Then, continuous differentiability implies that

$$v_a(x) = w(x) = \frac{\kappa\varphi_b}{r + \kappa\varphi_b} v_b(x) \text{ and } v_a'(x) = w'(x) = \frac{\kappa\varphi_b}{r + \kappa\varphi_b} v_b'(x). \tag{A.7}$$

Now define $\Omega_q(x) := -c + \lambda_q x\big(1 + w(1)\big)$ and $\Gamma_q(x) := r + \lambda_q x$. Then, continuous differentiability and the Bellman equations on the continuation regions $\mathcal{M}_a$ and $\mathcal{M}_b$ lead to:

$$\frac{v_b'(x)}{v_b'(x)} = \frac{\lambda_a}{\lambda_b} \frac{\Omega_b(x) - \Gamma_b(x)v_b(x)}{\Omega_a(x) - \Gamma_a(x)v_a(x)}.$$

Simplifying the previous two equations gives

$$\left(\frac{\lambda_b}{\lambda_a} - 1\right) r v_b(x) = -c\left(\frac{r + \kappa\varphi_b}{\kappa\varphi_b}\frac{\lambda_b}{\lambda_a} - 1\right) + \frac{r\lambda_b x\big(1 + w(1)\big)}{\kappa\varphi_b}. \tag{A.8}$$

By Lemma A.2, x is a maximizer of $v_b(\cdot)/v_a(\cdot)$. Also, observe that v_a solves essentially the same Bellman equation (upto the change of constants) as v_b^*. Thus, the form of its particular and homogenous solutions are the same as v_b^*, and hence it becomes twice differentiable on $\mathcal{M}_a$. Since x is a maximizer of v_b/v_a, and this ratio strictly decreases to the right of

x, then it must be that $\lim_{\varepsilon \downarrow 0} \left(\frac{v_b(x+\varepsilon)}{v_a(x+\varepsilon)} \right)'' \leqslant 0$. Let us denote $v_q''(x) := \lim_{\varepsilon \downarrow 0} v_q''(x+\varepsilon)$ for $q \in \{a, b\}$. Then, the previous second order condition together with (A.7) imply:

$$\frac{v_b''(x)}{v_b(x)} \leqslant \frac{v_a''(x)}{v_a(x)} \Rightarrow v_b''(x) \leqslant \frac{r + \kappa\varphi_b}{\kappa\varphi_b} v_a''(x). \tag{A.9}$$

One can find an expression for the second order derivatives by differentiating the Bellman equations on the continuation region:

$$rv_q'(x) = \lambda_q \big(1 + w(1) - v_q(x)\big) - \lambda_q x v_q'(x) - \lambda_q(1 - 2x)v_q'(x) - \lambda_q x(1 - x)v_q''(x).$$

Replacing v_q' from the original Bellman equation into the above relation yields the following expression for v_q'':

$$\lambda_q x(1 - x)v_q''(x) = -\frac{r\big(1 + w(1)\big)}{1 - x} + \frac{r(r + \lambda_q)}{\lambda_q x(1 - x)} v_q(x) + \frac{c\big(r + \lambda_q(1 - x)\big)}{\lambda_q x(1 - x)}.$$

Plugging the second order derivatives from above into (A.9) and applying some rearrangements yield the following *equivalent* inequality:

$$rv_b(x) \left(\frac{\lambda_b}{\lambda_a} - 1 \right) \left(1 + \frac{r}{\lambda_a} + \frac{r}{\lambda_b} \right) \geqslant$$
$$\big(rx\big(1 + w(1)\big) - c(1 - x)\big) \left(\frac{r + \kappa\varphi_b}{\kappa\varphi_b} \frac{\lambda_b}{\lambda_a} - 1 \right) - \frac{cr}{\lambda_b} \left(\frac{r + \kappa\varphi_b}{\kappa\varphi_b} \frac{\lambda_b^2}{\lambda_a^2} - 1 \right).$$

Substituting (A.8) in the above expression, and applying several regroupings amount to:

$$x \left[\big(1 + w(1)\big)\big(\lambda_a (r + \lambda_b) - \kappa\varphi_b (\lambda_b - \lambda_a)\big) - c\big(\lambda_b + r^{-1}\kappa\varphi_b (\lambda_b - \lambda_a)\big) \right] \geqslant cr.$$

In the high cost regime $w(1)$ follows (A.3), which after substitution into the above inequality leads to an equivalent condition to (A.9), that is only in terms of the primitives of the model:

$$\frac{cr^2}{r + \kappa\varphi_b} \left(1 + \frac{\kappa\varphi_b}{r + \lambda_b} \right) + cx\lambda_b \left(1 + \frac{r}{r + \lambda_b} \frac{\kappa\varphi_b}{r + \kappa\varphi_b} \right)$$
$$\leqslant x\big(\lambda_a (r + \lambda_b) - \kappa\varphi_b (\lambda_b - \lambda_a)\big). \tag{A.10}$$

Next, I will show that the *lhs* above is always greater than the *rhs*, thus the contradiction is resulted and there is no $x = \inf \mathcal{M}_a \in \mathcal{M}_b$. Obviously at $x = 0$ the *lhs* is greater than

the *rhs*. At $x = 1$, the *rhs* is increasing in λ_a, so can be upper bounded when λ_a assumes its maximum level in the high cost regime, i.e., $c + \frac{\kappa\varphi_b(\lambda_b - c)}{r + \kappa\varphi_b + \lambda_b}$. Therefore the *rhs* of (A.10) at $x = 1$ is upper bounded by

$$\lambda_a\left(r + \lambda_b\right) - \kappa\varphi_b\left(\lambda_b - \lambda_a\right) \leqslant c\left(r + \lambda_b\right).$$

However, the *lhs* of (A.10) equals $c(r + \lambda_b)$ at $x = 1$. So (A.10) can never be satisfied. Therefore, the contradiction is resulted, and thus in the high cost regime $\mathcal{M}_a$ cannot have a lower boundary point in $\mathcal{M}_b$. $\qquad\square$

By Corollary 3.2 at the optimum $\mathcal{M}_a \subseteq \mathcal{M}_b$. Thus, the previous lemma implies that $\mathcal{M}_a = \varnothing$ in the high cost regime, and thereby concluding the proof of Proposition 3.7.

A.4   Proof of Theorem 3.8

The fact that the suggested tuple satisfies the necessary conditions (2.2), (2.4) and (2.5) as well as the majorizing condition is established in Section 3.1. It thus only remains to show that this tuple also satisfies the superharmonic condition.

Superharmonicity of v_b^*.   Obviously the superharmonic condition holds with equality on $(\beta, 1]$, because v_b^* solves the Bellman equation on this region. However, it needs to be checked on $[0, \beta]$ as well. Observe that for $\pi \in [0, \beta]$, one has $v_b^*(\pi) = 0$, thus

$$\mathcal{L}_b v_b(\pi) - r v_b(\pi) - c = \lambda_b \pi\left(1 + w(1)\right) - c \leqslant \lambda_b \beta\left(1 + w(1)\right) - c = 0\,,$$

where the last equality holds by Lemma 3.5.

Superharmonicity of v_a^*.   In the low cost regime and on the interval $(\alpha, 1]$, v_a^* clearly satisfies the superharmonic property, because it actually solves the Bellman equation. The proofs of the superharmoincity of v_a^* in the low cost regime on $[0, \alpha]$, and in the high cost regime on $[0, 1]$ follow the same logic, thus here I only present the latter. In the high cost regime and on $[0, \beta]$ one has $v_a^* = 0$, and thus

$$\mathcal{L}_a v_a^*(\pi) - r v_a^*(\pi) - c = \lambda_a \pi\left(1 + w(1)\right) - c \leqslant \lambda_b \beta\left(1 + w(1)\right) - c = 0\,.$$

The analysis of the superharmonicity of v_a^* on $(\beta, 1]$ however needs a little more work. On this region, $v_a^* = \frac{\kappa\varphi_b}{r+\kappa\varphi_b} v_b^*$, thus

$$\mathcal{L}_a v_a^*(\pi) - r v_a^*(\pi) - c = \mathcal{L}_a \left(\frac{\kappa\varphi_b}{r+\kappa\varphi_b} v_b^*\right)(\pi) - \frac{r\kappa\varphi_b}{r+\kappa\varphi_b} v_b^*(\pi) - c$$

$$= \frac{\kappa\varphi_b}{r+\kappa\varphi_b}\left(\mathcal{L}_a v_b^*(\pi) - r v_b^*(\pi) - c\right) + \frac{r\lambda_a \pi}{r+\kappa\varphi_b}\left(1 + w(1)\right) - \frac{rc}{r+\kappa\varphi_b}.$$

Adding and subtracting $\mathcal{L}_b v_b^*$ from the first parentheses above, and observing the Bellman equation for v_b^* result in

$$\mathcal{L}_a v_a^*(\pi) - r v_a^*(\pi) - c = -\frac{\kappa\varphi_b}{r+\kappa\varphi_b}(\mathcal{L}_b - \mathcal{L}_a)v_b^*(\pi) + \frac{r\lambda_a \pi}{r+\kappa\varphi_b}\left(1 + w(1)\right) - \frac{rc}{r+\kappa\varphi_b}$$

$$= -\frac{\kappa\varphi_b}{r+\kappa\varphi_b}(\lambda_b - \lambda_a)\pi\left(1 + w(1) - v_b^*(\pi) - (1-\pi)v_b^{*\prime}(\pi)\right)$$

$$+ \frac{r\lambda_a \pi}{r+\kappa\varphi_b}\left(1 + w(1)\right) - \frac{rc}{r+\kappa\varphi_b}.$$

One can easily verify that v_b^* is convex, so, $v_b^*(\pi) + (1-\pi)v_b^{*\prime}(\pi) \leqslant v_b^*(1)$. That in turn implies an upper bound on the above expression:

$$\mathcal{L}_a v_a^*(\pi) - r v_a^*(\pi) - c \leqslant -\frac{\kappa\varphi_b}{r+\kappa\varphi_b}\frac{r(\lambda_b - \lambda_a)\pi}{r+\lambda_b}\left(1 + w(1) + \frac{c}{r}\right) + \frac{r\lambda_a \pi\left(1 + w(1)\right) - cr}{r+\kappa\varphi_b}$$

$$\leqslant \left(-\frac{\kappa\varphi_b}{r+\kappa\varphi_b}\frac{r(\lambda_b - \lambda_a)}{r+\lambda_b}\left(1 + w(1) + \frac{c}{r}\right) + \frac{r\lambda_a\left(1 + w(1)\right) - rc}{r+\kappa\varphi_b}\right)^+.$$

In the second inequality above, I used the fact that the *rhs* of the first inequality is affine in π and negative at $\pi = 0$. Let us denote the argument of $(\cdot)^+$ by $\mathfrak{Z}$. It is increasing in λ_a, hence can be bounded above when λ_a is replaced with $c + rw(1)$ (i.e., its maximum value in the high cost regime):

$$\mathfrak{Z} \leqslant -\frac{\kappa\varphi_b}{r+\kappa\varphi_b}\frac{r\left(\lambda_b - c - rw(1)\right)}{r+\lambda_b}\left(1 + w(1) + \frac{c}{r}\right) + \frac{r\left(c + rw(1)\right)\left(1 + w(1)\right) - rc}{r+\kappa\varphi_b}$$

$$= -\frac{\kappa\varphi_b}{r+\kappa\varphi_b}\frac{(\lambda_b - c)(r+\lambda_b)(r+\kappa\varphi_b + c)}{r(\kappa\varphi_b + r + \lambda_b)^2} + \frac{\kappa\varphi_b}{r+\kappa\varphi_b}\frac{(\lambda_b - c)(r+\lambda_b)(r+\kappa\varphi_b + c)}{r(\kappa\varphi_b + r + \lambda_b)^2} = 0,$$

where in the second line $w(1)$ is replaced from equation (A.3). This concludes the proof of superharmonicity of v_a^* with respect to $\mathcal{L}_a$ on $(\beta, 1]$, thereby on the entire $[0, 1]$. $\qquad\square$

A.5 Proof of Theorem 3.9

Let w^* be the reputation function that is induced by v^* and $\mathcal{M}^*$ following equation (2.2). To verify the optimality, I need to show that given w^*, the matching set $\mathcal{M}^*$ is the optimal continuation region, and v^* is the optimal value function for the stopping time problem of (2.3). I apply a Martingale verification argument to establish the optimality. Define $\mathbf{v}_q(\iota, \pi) := v_q^*(\pi)1_{\{\iota=0\}} + (\iota + w^*(\pi))1_{\{\iota=1\}}$, where ι is the success indicator process. Since $\mathbf{v}$ is a bounded function, for each $q \in \{a, b\}$, one can find a bounded (and hence uniformly integrable) Martingale process M^q such that:

$$e^{-rt}\mathbf{v}_q(\iota_t, \pi_t) = \mathbf{v}_q(\iota, \pi) + \int_0^t e^{-rs}\big(\mathbb{L}_q\mathbf{v}_q(\cdot, \cdot) - r\mathbf{v}_q(\cdot, \cdot)\big)(\iota_{s-}, \pi_{s-})\,\mathrm{d}s + M_t^q, \qquad \text{(A.11)}$$

where $\mathbb{L}_q\mathbf{v}_q(\iota, \pi) := \mathcal{L}_q v_q(\pi)1_{\{\iota=0\}}$. By the majorizing condition, for every stopping time τ, one has $\mathbf{v}_q(\iota_\tau, \pi_\tau) \geqslant \iota_\tau + w^*(\pi_\tau)$, therefore

$$e^{-r\tau}\big(\iota_\tau + w^*(\pi_\tau)\big) \leqslant \mathbf{v}_q(\iota, \pi) + \int_0^\tau e^{-rs}\big(\mathbb{L}_q\mathbf{v}_q(\cdot, \cdot) - r\mathbf{v}_q(\cdot, \cdot)\big)(\iota_{s-}, \pi_{s-})\,\mathrm{d}s + M_\tau^q$$

$$\leqslant \mathbf{v}_q(\iota, \pi) + \int_0^\tau ce^{-rs}\,\mathrm{d}s + M_\tau^q,$$

wherein the second inequality I used the superharmonic property proven in Theorem 3.8. Doob's optional stopping theorem implies that $\mathsf{E}M_\tau^q = 0$, hence for every stopping time τ, one has

$$\mathbf{v}_q(\iota, \pi) \geqslant \mathsf{E}_{\pi,q,\iota}\left[e^{-r\tau}\left(\iota_\tau + w^*(\pi_\tau)\right) - c\int_0^\tau e^{-rs}\,\mathrm{d}s\right].$$

That in turn implies

$$v_q^*(\pi) \geqslant \sup_\tau \mathsf{E}_{\pi,q,\iota=0}\left[e^{-r\tau}\left(\iota_\tau + w^*(\pi_\tau)\right) - c\int_0^\tau e^{-rs}\,\mathrm{d}s\right]. \qquad \text{(A.12)}$$

To show the achievability in the above inequality, define

$$\tau_q := \inf\left\{t \geqslant 0 : \pi_t \notin \mathcal{M}_q^* \text{ or } \iota_t = 1\right\},$$

where $\mathcal{M}^*$ is characterized in Theorem 3.8. Applying equation (A.11) yields

$$
\begin{aligned}
e^{-r\tau_q}\left(\iota_{\tau_q} + w^*(\pi_{\tau_q})\right) &= e^{-r\tau_q}\mathbf{v}_q(\iota_{\tau_q}, \pi_{\tau_q}) \\
&= \mathbf{v}_q(\iota, \pi) + \int_0^{\tau_q} e^{-rs}\left(\mathbb{L}_q\mathbf{v}_q(\cdot,\cdot) - r\mathbf{v}_q(\cdot,\cdot)\right)(\iota_{s-}, \pi_{s-})\,\mathrm{d}s + M_{\tau_q}^q \\
&= \mathbf{v}_q(\iota, \pi) + \int_0^{\tau_q} ce^{-rs}\,\mathrm{d}s + M_{\tau_q}^q \,,
\end{aligned}
$$

where the second equality holds because the Bellman equation applies on $\mathcal{M}_q^*$. Taking expectations of both sides in the above equality, one obtains

$$
\mathbf{v}_q(\iota, \pi) = \mathsf{E}_{\pi, q, \iota}\left[e^{-r\tau_q}\left(\iota_{\tau_q} + w^*(\pi_{\tau_q})\right) - c\int_0^{\tau_q} e^{-rs}\,\mathrm{d}s\right].
$$

Therefore,

$$
v_q^*(\pi) = \mathsf{E}_{\pi, q, \iota=0}\left[e^{-r\tau_q}\left(\iota_{\tau_q} + w^*(\pi_{\tau_q})\right) - c\int_0^{\tau_q} e^{-rs}\,\mathrm{d}s\right],
$$

which together with (A.12) concludes the proof. $\qquad\square$

A.6   Proof of Proposition 4.1

All the arguments below are stated without using ^ on top of the variables. After I have established the unique existence of the optimum, one can bring back the ^ superscript.

Observe that the variational Bellman equation for $v_q(\pi)$ in (4.1) can be equivalently expressed as:

$$
v_q(\pi) = \max\left\{w(\pi), \frac{\lambda_q\pi - c}{r + \lambda_q\pi} + \frac{\lambda_q\pi w(\pi)}{r + \lambda_q\pi}\right\}. \tag{A.13}
$$

This representation implies that $v_q(\pi) > w(\pi)$ if and only if the second maximand is larger than the first, that happens if and only if $rw(\pi) < \lambda_q\pi - c$. Hence,

$$
\pi \in \mathcal{M}_q \Leftrightarrow v_q(\pi) > w(\pi) \Leftrightarrow \lambda_q\pi - c > rw(\pi). \tag{A.14}
$$

This already implies that in any optimal outcome, $\mathcal{M}_a \subseteq \mathcal{M}_b$, thus proving the last claim in part (i). Additionally, it shows that

$$
w(\pi) = 0 \text{ if and only if } v_b(\pi) = 0. \tag{A.15}
$$

Let $\hat{\beta} = \inf \mathcal{M}_b$, then $v_b(\hat{\beta}) = 0$, and hence $\hat{\beta} = c/\lambda_b$. Then, (A.14) and (A.15) together imply that $v_b(\pi) > 0$ for all $\pi > \hat{\beta}$, thereby proving the unique existence of $\widehat{\mathcal{M}}_b = (\hat{\beta}, 1]$, where $\hat{\beta} = c/\lambda_b$.

To show the uniqueness of $\mathcal{M}_a$, denote the reputation function in the outcome where $\pi \in \mathcal{M}_a \cap \mathcal{M}_b$ by $w_{ab}(\pi)$, and in the outcome where $\pi \in \mathcal{M}_a^c \cap \mathcal{M}_b$ by $w_b(\pi)$. One can find closed-form expressions for both of these functions using (2.2) and (A.13). Then, some long yet straightforward algebraic computations show that there exists a threshold $\hat{\alpha} > 0$, such that the following statements become equivalent:

$$ w_{ab}(\pi) > w_b(\pi) \Leftrightarrow \lambda_a \pi - c > r w_{ab}(\pi) \Leftrightarrow \pi > \hat{\alpha} \,. $$

In particular, the expression for $\hat{\alpha}$ is

$$ \hat{\alpha} = \frac{1}{2\lambda_a \lambda_b} \big(c\lambda_b - r\lambda_a + \kappa\varphi_b(\lambda_b - \lambda_a) \big) \times $$
$$ \sqrt{r^2\lambda_a^2 + 2r\lambda_a \big(c\lambda_b - \kappa\varphi_b(\lambda_b - \lambda_a) \big) + \big(c\lambda_b + \kappa\varphi_b(\lambda_b - \lambda_a) \big)^2} \tag{A.16} $$

Some further computations verify that $\hat{\alpha} \leqslant 1$ if and only if the economy is in the low cost regime. Therefore, one obtains the unique existence of $\widehat{\mathcal{M}}_a = \varnothing$ in the high cost regime, and $\widehat{\mathcal{M}}_a = (\hat{\alpha}, 1]$ in the low cost regime. This concludes the proof of both parts in Proposition 4.1. $\square$

References

[1] Simon Board and Moritz Meyer-ter-Vehn (2013). "Reputation for Quality," *Econometrica*, 81(6): 2381–2462.

[2] Patrick Bolton and Christopher Harris (1999). "Strategic Experimentation," *Econometrica*, 67(2): 349–374.

[3] Alessandro Bonatti and Johannes Hörner (2017). "Career Concerns with Exponential Learning," *Theoretical Economics*, 12(1): 425–475.

[4] Svetlana Boyarchenko (2021). "Inefficiency of Sponsored Research," *Journal of Mathematical Economics*, 95: 102497.

[5] Braz Camargo (2007). "Good news and bad News in Two-armed Bandits," *Journal of Economic Theory*, 135(1): 558–566.

[6] Archishman Chakraborty, Alessandro Citanna, and Michael Ostrovsky (2010). "Two-sided Matching with Interdependent Values," *Journal of Economic Theory*, 145(1): 85–105.

[7] Kaustav Das, Nicolas Klein, and Katharina Schmid (2020). "Strategic Experimentation with Asymmetric Players," *Economic Theory*, 69: 1147–1175.

[8] Paul Gompers and Josh Lerner (1999). "An Analysis of Compensation in the US Venture Capital Partnership," *Journal of Financial Economics*, 51(1): 3–44.

[9] Michael Gorman and William A. Sahlman (1989). "What do Venture Capitalists do?" *Journal of Business Venturing*, 4(4): 231–248.

[10] Thomas Hellmann and Manju Puri (2002). "Venture Capital and the Professionalization of Start-up Firms: Empirical Evidence," *The Journal of Finance*, 57(1): 169–197.

[11] Bengt Holmström (1999). "Managerial Incentive Problems: A Dynamic Perspective," *The Review of Economic Studies*, 66(1): 169–182.

[12] Boyan Jovanovic (1979). "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 87(5, Part 1): 972–990.

[13] Godfrey Keller and Sven Rady (2010). "Strategic Experimentation with Poisson Bandits," *Theoretical Economics*, 5(2): 275–311.

[14] ——— (2015). "Breakdowns," *Theoretical Economics*, 10(1): 175–202.

[15] Godfrey Keller, Sven Rady, and Martin Cripps (2005). "Strategic Experimentation with Exponential Bandits," *Econometrica*, 73(1): 39–68.

[16] Josh Lerner (2002). "When Bureaucrats Meet Entrepreneurs: the Design of Effective Public Venture Capital Programmes," *The Economic Journal*, 112(477): F73–F84.

[17] Josh Lerner, David Moore, and Stuart Shepherd (2005). "A Study of New Zealand's Venture Capital Market and Implications for Public Policy," *Report to the Ministry of Research Science & Technology, LECG Limited, New Zealand.*

[18] Fei Li and Xi Weng (2017). "Efficient Learning and Job Turnover in the Labor Market," *International Economic Review*, 58(3): 727–750.

[19] Qingmin Liu, George J Mailath, Andrew Postlewaite, and Larry Samuelson (2014). "Stable Matching with Incomplete Information," *Econometrica*, 82(2): 541–587.

[20] Gustavo Manso (2011). "Motivating Innovation," *The Journal of Finance*, 66(5): 1823–1860.

[21] Chiara Margaria (2020). "Learning and Payoff Externalities in an Investment Game," *Games and Economic Behavior*, 119: 234–250.

[22] Giuseppe Moscarini (2005). "Job Matching and the Wage Distribution," *Econometrica*, 73(2): 481–516.

[23] Goran Peskir and Albert Shiryaev (2006). *Optimal Stopping and Free-boundary Problems*, Springer.

[24] Farzad Pourbabaee (2020). "Robust Experimentation in the Continuous Time Bandit Problem," *Economic Theory*: 1–31.

[25] Dinah Rosenberg, Antoine Salomon, and Nicolas Vieille (2013). "On Games of Strategic Experimentation," *Games and Economic Behavior*, 82: 31–51.

[26] Morten Sørensen (2007). "How Smart is Smart Money? A Two-Sided Matching Model of Venture Capital," *The Journal of Finance*, 62(6): 2725–2762.

[27] Xuan Tian and Tracy Yue Wang (2014). "Tolerance for Failure and Corporate Innovation," *The Review of Financial Studies*, 27(1): 211–255.

[28] Masako Ueda (2004). "Banks versus Venture Capital: Project Evaluation, Screening, and Expropriation," *The Journal of Finance*, 59(2): 601–621.