ISER Seminar Series
4-6 pm, Mon, August 28, 2017

Learning while Experimenting

Ettore Damiano
University of Toronto

Li, Hao
University of British Columbia

Wing Suen
University of Hong Kong

February 4, 2017

Abstract.

An agent in an exponential bandit problem becomes more pessimistic and eventually quits as risky experimentation fails to deliver success. This agent can benefit from directly learning about the state. "Positive information acquisition" seeks news that would confirm the state that favors experimentation. It is optimally used as a last ditch effort before abandoning the risky arm. "Negative information acquisition" seeks news that would demonstrate that experimentation is futile. It is optimally used as an insurance strategy to avoid wasteful experimentation when the agent is still optimistic about the risky arm. If the arrival rate of negative information is sufficiently high, the agent eventually adopts a "perpetual" strategy of experimentation and information acquisition in such a way that his belief becomes stationary, until either he achieves a success from the risky arm or quits upon the arrival of negative news.

JEL classification. D83, L15

Keywords. Benefit of information, usefulness of information, smooth pasting, correlated arms.

1. Introduction

In multi-armed bandit models (Robins 1952), an agent choosing from alternatives with stochastic payoffs (risky arms) faces a trade-off between maximizing the expected payoff based on what he currently knows about the alternatives, and learning about the stochastic processes that generate these payoffs by experimenting with different alternatives. The dynamic trade-off between exploitation and experimentation captured in bandit models has found many applications in economics, ranging from project selection in industrial research and development (Weizman, 1979; Roberts and Weizman, 1981), to job search with firm-specific or industry-specific productivities (Jovanovich, 1979), to monopoly producers learning about market demands (Rothschild, 1984). More recently, bandit problems have been extended to multi-agent settings, including general approaches to strategic experimentation (Bolton and Harris, 1999; Keller, Rady and Cripps, 2005), and applications such as R&D races (Choi, 1991),[1] wage setting (Fellis and Harris, 1996), and price competition (Bergemann and Valimaki, 1996). In all these models, the agent learns about an individual risky alternative by trying it; that is, learning occurs only *through* experimenting.

In this paper, we study a simple bandit model where, while experimenting, the agent can also engage in a costly dynamic process of information acquisition.[2] Learning about risky arms *while* experimenting is modeled here as a pure information activity, because rewards can only arrive through experimentation. Further, unlike experimentation, the information acquisition process can be suspended and resumed as the agent sees fit. For example, in the R&D application of bandit problems, while a pharmaceutical firm has a team of laboratory chemists engaged in a process to develop a new drug, it may hire industrial academic scientists to research about the biochemical foundation behind the potential new drug. A scientific theory

[1] See also Reinganum (1981, 1982), Harris and Vickers (1985, 1987), and Malueg and Tsutsui (1997).

[2] In recent independent work, Che and Mierendorff (2016) study the problem of an agent choosing between two possible actions with uncertain payoffs, who can delay the decision to acquire more information about the state, by allocating a fixed budget of "attention" to available information structures. Information acquisition is the only source of learning in their model. In our model, the agent must choose whether to continue with the risky alternative or abandon it. Experimenting with the risky arm is in itself informative, and this learning is augmented by the agent acquiring additional information at a cost. While the models are distinct, some of the insights that emerge from their analysis are similar to ours.

will not in itself bring about the new drug, but it can inform the laboratory chemists in their trial and error process. Similarly, while a job-seeker goes through applications and interviews with different firms in an unfamiliar industry, he may be able to acquire information about the industry and his particular fit with it through word-of-mouth in his social network. The same is true for modeling consumer demand for an experience good: besides trying out the good themselves, buyers may be able to obtain useful information about the good by asking their friends about their experiences.

Although a pure information activity, learning while experimenting changes the agent's belief about the prospects of risky arms and affects the agent's experimentation. The value of information in our model is thus endogenous, and the comparison of different information structures—the main objective of this paper—depends on the current state of the experimentation process (i.e., how optimistic or pessimistic the agent is about the prospects of risky alternatives). To capture learning while experimenting, we use as the benchmark a single-agent, single-risky arm version of the exponential bandits model of Keller, Rady and Cripps (2005). The risky arm is either good and produces a success at an exponential rate, or bad with no possibility of success. Learning through experimentation takes a simple form: starting from a prior belief that the risky arm is good, as the agent continues to experiment without achieving a success, he becomes increasingly pessimistic, eventually abandoning the risky arm when the beliefs drops below a critical threshold.

We consider two opposite information structures in learning while experimenting, both modeled as an exponential process with an uncertain arrival rate. In the case of *positive information*, the agent pays to search for conclusive evidence that a success can indeed be obtained from experimentation. Of course, no such evidence can be found if the risky arm is actually bad. Moreover, even after the evidence arrives, the agent can achieve a success only by continuing with the risky arm, except now the arrival is stochastic but no longer uncertain. In the pharmaceutical firm R&D application, a scientific theory establishing the sound foundation or feasibility of the new drug may be an example of such information structure. In the job search application, positive information may obtain by surveying only one's friends who remain employed in the industry; and in the experimental consumption application, the current users of the new product. In the case of *negative information*, the agent pays to search for conclusive evidence that no success can be obtained from

experimentation. Such evidence arrives at a positive rate when the risky arm is bad, but it will never arrive in the opposite state. For the pharmaceutical firm example, negative information acquisition may take the form of paying scientists to look for counterexamples that would demonstrate fatal flaws in the scientific theory behind the new drug (such as a toxicological analysis), or exploring the possibility of a rival or superior drug. In the job search and experimental consumption examples, negative information may obtain when the agent selectively samples those who have left the industry and those who have tried and given up the new product, respectively.

The introduction of information acquisition in experimentation enriches the analysis of optimal experimentation. Without information acquisition, the question is simply when the agent should quit the risky arm; with it, we ask what factors determine whether and when information acquisition is useful to the agent. Is it best for the agent to acquire information about the risky arm when he is still optimistic about achieving a success or when he is already pessimistic? The answer turns out to depend on whether information is positive or negative. This is because the type of information affects the nature of the interaction between learning through experimenting and learning while experimenting.

Positive information reinforces the direct learning through experimenting, because failure to uncover positive evidence speeds up the downgrading of the agent's belief that a success from the risky arm can be achieved. When the agent is optimistic about the risky arm, there is relatively little use in having the positive news, and since information is costly, it is not optimal for the agent to acquire it. In contrast, positive information is more valuable to the decision of whether to quit the risky arm when the agent is pessimistic, as it can potentially avoid quitting before achieving success. We show that the benefit of positive information is the highest when the agent is just about to quit, and is lower when the agent is more optimistic. In a sense, positive information may be interpreted as a last ditch effort in trying the risky arm before irrevocably quitting it. Moreover, the agent optimally quits experimenting at a lower belief when positive information acquisition is used than when it is not used.

Unlike positive information, negative information counters learning through experimenting. Failure to uncover negative evidence slows down the downgrading of the agent's belief. When the agent is already so pessimistic that he is about to abandon the risky arm, there is little use in having the negative news. In contrast,

negative information is more beneficial when the agent is relatively optimistic, as an insurance strategy to avoid potentially waiting too long to quit. When the arrival rate of negative information is higher than the success arrival rate of the good risky arm, failure to uncover negative evidence drives up the agent's belief. In this case, if the cost of information acquisition is sufficiently low, the agent will eventually adopt a "perpetual" strategy of experimentation and information acquisition in such a way that his belief becomes stationary, until either he achieves a success from the risky arm or quits upon the arrival of negative news.

The usefulness of information, measured by the highest information acquisition cost that still justifies its use by an agent in optimal experimentation, is naturally greater if news arrives faster. Comparing positive and negative information, we find that the usefulness also depends on the efficiency of the risky arm, that is, on the arrival rate of success relative to the cost. When the risky arm is relatively inefficient, the agent tends to abandon the risky arm while his belief is still high. In this case, positive information is more useful, because the upside potential from a last ditch effort is high. In contrast, when the risky arm is relatively efficient, negative information is more useful, because it can potentially insure against a greater risk of mistakingly believing that a success is possible from the risky arm when it is actually not.

In Section 4 we extend our analysis in two directions. First, we allow the agent to acquire positive and negative information independently or jointly. We show that joint information acquisition never supersedes positive information but it may replace negative information. Second, we allow the firm to suspend the risky arm while engaging in information acquisition. We show that such "pure" information acquisition does not improve the usefulness of information, but pure positive information expands the range of beliefs for the agent to make a last ditch effort to resurrect the risky arm, while pure negative information makes the insurance strategy valuable at more pessimistic beliefs.

2. Model

Consider the following continuous-time bandit model. There is a single arm that yields uncertain returns to an agent. For simplicity we assume that the agent does not discount. There are two states of the world. In state $\mathcal{G}$, the risky arm yields a "success" at a random time according to the exponential distribution with parameter

$\lambda > 0$ so long as the agent experiments with it. In state $\mathcal{B}$, the arrival rate of success is 0. The belief that the state is $\mathcal{G}$ is denoted by γ; and the initial belief is strictly between 0 and 1. Experimenting with the risky arm, or choosing D, has a flow cost $c > 0$. Quitting, or choosing Q, is an irreversible decision and yields a terminal payoff of 0.[3] We make the assumption that experimenting with the risky arm is worthwhile if the state is known to be $\mathcal{G}$:

$$c < \lambda\pi. \tag{1}$$

Of course, quitting is optimal when the state is known to be $\mathcal{B}$.

While experimenting with the risky arm, the agent can also try to learn the state of the world. Since there are two possible states, we consider two types of information structures available to the agent in information acquisition. In "positive" information acquisition, or choosing P, at intensity $\delta \in (0, 1]$ for an interval of time $\mathrm{d}t$, in state $\mathcal{G}$ conclusive news about the true state arrives with probability $\delta\alpha\mathrm{d}t$ where $\alpha > 0$, and no news ever arrives if the state is $\mathcal{B}$. In "negative" information acquisition, or choosing N, at intensity $\delta \in (0, 1]$, in state $\mathcal{B}$ the agent learns the true state at rate $\delta\alpha$, and no news arrives in state $\mathcal{G}$. We assume that both positive and negative information acquisition have a flow cost of $\delta k > 0$.

In the main analysis, we assume that the agent cannot acquire information about the state if he does not experiment with the risky arm. This means that the agent cannot begin with information acquisition before commencing experimentation. Since quitting is irreversible, the agent cannot suspend experimentation to conduct information acquisition either. In contrast, learning about the state is a purely informational activity. The agent can suspend or resume information acquisition as he wishes. There is no direct payoff from information acquisition, although the optimal policy is trivial after the agent receives conclusive news about the state. In positive information acquisition, after learning that the state is $\mathcal{G}$, the agent will experiment with the risky arm until a success, which by assumption (1) is optimal for the agent. In negative information acquisition, after learning that the state is $\mathcal{B}$, the agent will optimally quit.

If the agent can only choose between experimenting and quitting, our model is

[3]In standard models with independent arms, it does not matter whether quitting, or taking the safe option, is reversible or not. This is not the case in the present model because we have correlated arms.

a simplified version of the exponential bandit problem introduced by Keller, Rady and Cripps (2005). Exponential bandit problems have become a major workhorse in dynamic game-theoretic models of learning since Keller, Rady and Cripps (2005), because the potentially intractable problem of computing the Gittins index boils down to determining the optimal stopping time. Our model of information acquisition in experimentation may be thought of as a multi-arm bandit problem with correlated arms. The agent can be thought of as choosing between two risky arms, corresponding to experimentation only and the combination of both experimentation and information acquisition respectively, and a safe option of quitting. The new risky arm is correlated with the original one, unlike in the standard multi-arm bandit problem solved by Gittins (1979), because the potential payoffs from them are determined by the same underlying state.[4] So far as we know, there is no generally applicable index rule that replaces the Gittins index. As in a few papers in the economics literature that deal with multi-armed bandit with correlated arms (Camargo, 2007; Klein and Rady, 2011), we solve the optimization problem starting from first principles.

After some preliminary analysis of the case of experimentation only, we will first add positive information acquisition and negative information acquisition separately as a third choice in addition to experimenting and quitting. We then extend the analysis in two ways. In Section 4.1 we consider the case where the agent can choose one or both types of information acquisition besides experimenting, and in Section 4.2 we allow the agent to engage in information acquisition while suspending experimentation with the risky arm, considering positive information and negative information separately. Throughout the paper, we use $V(\gamma)$ to denote the value function for the agent's optimal policy, where γ is the agent's belief. We use the same notation in the main model as in the model extensions, as the context will make the meaning of the notation clear.

3. Analysis

Without the possibility of information acquisition, our model is just the single-agent version of Keller, Rady and Cripps (2005). To make the present paper self-contained, we briefly derive the optimal policy below.

[4]The correlation among risky arms remains if information acquisition is independent from experimentation in the "pure" information acquisition model discussed in Section 4.2 as well.

The belief γ that the state is $\mathcal{G}$ goes down as the agent chooses D and no success occurs in some time interval. By Bayes' rule, the updated belief is

$$\gamma + \mathrm{d}\gamma = \frac{\gamma(1 - \lambda \mathrm{d}t)}{\gamma(1 - \lambda \mathrm{d}t) + 1 - \gamma}.$$

Keeping only terms of order $\mathrm{d}t$, we have

$$\mathrm{d}\gamma = -\gamma(1 - \gamma)\lambda \mathrm{d}t.$$

In the region of beliefs where D is optimal, the value function $V(\gamma)$ satisfies the Bellman equation:

$$V(\gamma) = -c\mathrm{d}t + \gamma\lambda\pi\mathrm{d}t + (1 - \gamma\lambda\mathrm{d}t)V(\gamma + \mathrm{d}\gamma).$$

Using the expression for $\mathrm{d}\gamma$, $V(\gamma)$ follows the differential equation:

$$\gamma(1 - \gamma)\lambda V_D'(\gamma) = -c + \gamma\lambda\pi - \gamma\lambda V(\gamma), \tag{2}$$

where, to indicate the optimal policy in this case, we write $V_D'(\gamma)$ instead of $V'(\gamma)$ in the left-hand-side above. Alternatively, the above can be read as defining the slope, $V_D'(\gamma)$, of the function $V(\gamma)$ whenever the policy choice is D at γ, whether or not it is optimal at that belief.

Since quitting is irreversible, with a constant payoff equal to 0, the optimal policy is given by a cutoff γ_{QD} such that the agent chooses Q for $\gamma \leq \gamma_{QD}$ and D otherwise. The value of γ_{QD} can be found by value-matching ($V(\gamma_{QD}) = 0$) and smooth-pasting ($V'(\gamma_{QD}) = 0$). This yields

$$\gamma_{QD} = \frac{c}{\lambda\pi},$$

which is strictly between 0 and 1.

The value function can be found by solving the differential equation (2) with boundary condition $V(\gamma_{QD}) = 0$. For future reference, we denote it as $V_D(\gamma)$ and give the explicit formula:

$$V_D(\gamma) = \pi(\gamma - \gamma_{QD}) - \frac{c}{\lambda}(1 - \gamma)\left(\log\frac{\gamma}{1 - \gamma} - \log\frac{\gamma_{QD}}{1 - \gamma_{QD}}\right) \tag{3}$$

for $\gamma > \gamma_{QD}$, and 0 otherwise. The value function $V_D(\gamma)$ given in (3) is homogeneous of degree 0 in λ and c. We sometimes refer to the output-cost ratio λ/c as a measure of the *efficiency of experimentation*. For any $\gamma > \gamma_{QD}$, $V_D(\gamma)$ strictly increases in the value of success π and in the efficiency of experimentation λ/c.

3.1. Positive information acquisition

Before formally characterizing the optimal policy when positive information acquisition is available to the agent, we provide a heuristic argument to illustrate the benefit of positive information to a agent that adopts the optimal experimentation policy given above. We define the *benefit of positive information*, $B_P(\gamma)$, as the time rate of change in the expected payoff to a agent that gains access to free positive information for an instant of time $\mathrm{d}t$ before going back to his optimal experimentation policy. Suppose the belief is above γ_{QD}. The expected payoff without the free positive information is just $V_D(\gamma)$. Conditional on the state being $\mathcal{G}$, the state is revealed to the agent with probability $\alpha \mathrm{d}t$, after which the agent optimally chooses D until success, with expected payoff $\pi - c/\lambda$. In the absence of either a success from the risky arm or positive news, the new updated belief is $\gamma + \mathrm{d}\gamma$, with

$$\mathrm{d}\gamma = -\gamma(1-\gamma)(\lambda + \alpha)\mathrm{d}t.$$

For $\gamma > \gamma_{QD}$, the benefit of positive information is defined by:

$$\begin{aligned} B_P(\gamma)\mathrm{d}t = [&-c\mathrm{d}t + \gamma\lambda\pi\mathrm{d}t + \gamma\alpha\mathrm{d}t(\pi - c/\lambda) + (1 - \gamma\lambda\mathrm{d}t - \gamma\alpha\mathrm{d}t)V_D(\gamma + \mathrm{d}\gamma)] \\ &- V_D(\gamma), \end{aligned}$$

where the term is square brackets is the expected payoff with free positive information for a brief interval of time. Using (2) for a first-order approximation of $V_D(\gamma + \mathrm{d}\gamma)$, we obtain

$$B_P(\gamma) = \alpha\left[(1-\gamma)c/\lambda\right]. \tag{4}$$

If the agent is already engaged in experimenting, the benefit of positive information is decreasing in γ. In particular, $B_P(\gamma)$ is maximized when the belief is γ_{QD} and the agent is just ready to quit.[5] Positive information is used in this case as a last ditch effort before abandoning the risky arm. When the agent is more optimistic, the benefit is lower because of a smaller change in the agent's belief when news from positive information acquisition arrives (from γ to 1). Finally, the benefit of positive information is higher when experimentation is less efficient, i.e., when λ/c

[5] In our model quitting is an irreversible decision so the benefit of positive information is nil in the region of beliefs below γ_{QD}. When quitting is not an irreversible decision, for $\gamma < \gamma_{QD}$ the benefit of positive information is given by $\gamma\alpha(\pi - c/\lambda)$, which is increasing in γ. Thus the benefit of positive information is still maximized at γ_{QD}.

is lower. This is because when experimentation is less efficient, the agent quits at a less pessimistic belief, in which case positive information can potentially generate a greater gain by resurrecting the risky arm.

With experimentation only, the agent's belief that the state is $\mathcal{G}$ goes down when he chooses D but there is no success. Positive information acquisition speeds up belief updating, as no positive news is bad news about the state. For a small interval of time $\mathrm{d}t$ and a fixed information acquisition intensity $\delta \in (0, 1]$, the value function $V(\gamma)$ satisfies

$$V(\gamma) = -(c + \delta k)\mathrm{d}t + \gamma(\lambda\pi\mathrm{d}t + \delta\alpha(\pi - c/\lambda)\mathrm{d}t) + (1 - \gamma\lambda\mathrm{d}t - \gamma\delta\alpha\mathrm{d}t)V(\gamma + \mathrm{d}\gamma),$$

with $\mathrm{d}\gamma = -\gamma(1-\gamma)(\lambda + \delta\alpha)\mathrm{d}t$. From this, we obtain the differential equation:

$$\gamma(1-\gamma)(\lambda + \delta\alpha)V'_{P,\delta}(\gamma) = -(c + \delta k) + \gamma(\lambda\pi + \delta\alpha(\pi - c/\lambda)) - \gamma(\lambda + \delta\alpha)V(\gamma), \quad (5)$$

where we have used $V'_{P,\delta}(\gamma)$ instead of $V'(\gamma)$ to indicate the policy chosen in this region. Note that the above assumes a fixed intensity of information acquisition.

By the assumption of irreversible quitting, the value from Q is 0 regardless of the belief about the state. We first argue that, due to the linearity inherent in the continuous-time model, it is without loss of generality to restrict the agent to choosing positive information acquisition at full intensity, i.e., $\delta = 1$. The proof is standard and is relegated to the appendix.

Lemma 1. *There is no interval of beliefs for which P at intensity $\delta < 1$ is optimal.*

For the remainder of this section, we write V_P instead of $V_{P,1}$. The following result imposes further restrictions on $V'(\gamma)$ by considering one-shot deviations.

Lemma 2. *At a given belief γ: (i) if D is optimal, then $V'_D(\gamma) \geq V'_P(\gamma)$; (ii) if P is optimal, then $V'_P(\gamma) \geq V'_D(\gamma)$; and (iii) if Q is optimal, then $0 \geq \max\{V'_D(\gamma), V'_P(\gamma)\}$.*

Proof. For part (i), suppose D is optimal in the neighborhood of γ. Then, it must not be better to switch to P for an interval of time $\mathrm{d}t$ and then continue with D. Thus,

$$\begin{aligned} V(\gamma) \geq &-(c + k)\mathrm{d}t + \gamma(\lambda\pi\mathrm{d}t + \alpha(\pi - c/\lambda)\mathrm{d}t) \\ &+ (1 - \gamma\lambda\mathrm{d}t - \gamma\alpha\mathrm{d}t)V(\gamma + \mathrm{d}\gamma), \end{aligned}$$

where $V(\gamma + \mathrm{d}\gamma)$ is the continuation payoff, given by

$$V(\gamma + \mathrm{d}\gamma) = V(\gamma) - \gamma(1-\gamma)(\lambda + \alpha)V_D'(\gamma)\mathrm{d}t.$$

in first-order approximation. The above implies

$$\begin{aligned} 0 &\geq -(c+k) + \gamma(\lambda\pi + \alpha(\pi - c/\lambda)) - \gamma(\lambda+\alpha)V(\gamma) - \gamma(1-\gamma)(\lambda+\alpha)V_D'(\gamma) \\ &= \gamma(1-\gamma)(\lambda+\alpha)(V_P'(\gamma) - V_D'(\gamma)), \end{aligned}$$

where the equality follows from the definition of $V_P'(\gamma)$ (equation (5) with $\delta = 1$). Thus, $V_D'(\gamma) \geq V_P'(\gamma)$. Part (ii) can be established similarly. For part (iii), if Q is optimal in the neighborhood of some γ, then it must not be optimal to switch to either D or P for an interval of time $\mathrm{d}t$ and then quit. Note that the continuation payoff $V(\gamma + \mathrm{d}\gamma)$ is zero by assumption. Thus,

$$0 = V(\gamma) \geq -c\mathrm{d}t + \gamma\lambda\pi\mathrm{d}t = \gamma(1-\gamma)V_D'(\gamma),$$

where the equality follows from the definition of $V_D'(\gamma)$ because $V(\gamma) = 0$. Thus, $V_D'(\gamma) \leq 0$. Similar, $V_P'(\gamma) \leq 0$. ∎

Lemma 2 is stated in terms of necessary conditions for the optimality of the two non-quitting options of D and P. Nonetheless, so long as quitting is not optimal at γ, an immediate corollary of Lemma 2 is that we also have the sufficient conditions if the comparisons are strict: D is optimal if $V_D'(\gamma) > V_P'(\gamma)$, and P is optimal if $V_P'(\gamma) > V_D'(\gamma)$.

Lemma 2 implies that the standard smooth-pasting condition holds. From parts (i) and (ii), we conclude that $V_D'(\gamma) = V_P'(\gamma)$ at any belief γ where the optimal policy changes between D to P. When the optimal policy switches from D to Q as the belief crosses some threshold γ, from part (iii) we have $V_D'(\gamma) \leq 0$, but if $V_D'(\gamma) < 0$ then it would not be optimal to choose D just above γ.[6] Thus, at the switching point we also have the smooth pasting condition $V_D'(\gamma) = 0$. Similarly, at any switching point γ where the optimal policy changes from P at just above γ to Q just below it, we must have $V_P'(\gamma) = 0$.

[6] At the threshold γ, the agent is indifferent between D and Q, so $V(\gamma) = 0$. If $V_D'(\gamma) < 0$, then since the belief moves down when the agent chooses D, the value at any belief just above γ is negative, which implies that the agent should have chosen Q instead.

From (2) and (5) we have

$$\gamma(1-\gamma)(\lambda+\alpha)(V_P'(\gamma)-V_D'(\gamma))=\alpha(1-\gamma)c/\lambda-k. \tag{6}$$

The right-hand side of the above equation is the benefit of positive information less its cost to the agent. Define

$$k_P \equiv \max_{\gamma} B_P(\gamma) = \alpha\frac{c}{\lambda}\left(1-\frac{c}{\lambda\pi}\right).$$

It is obvious that $V_P'(\gamma) < V_D'(\gamma)$ for all γ if $k > k_P$. Furthermore, let

$$\gamma_{PD} = 1-\frac{k/\alpha}{c/\lambda}$$

be the belief that satisfies the smooth-pasting condition joining the P region and the D region, and let

$$\gamma_{QP} = \frac{c+k}{\lambda\pi+\alpha(\pi-c/\lambda)}$$

be the belief that satisfies the smooth-pasting condition joining the Q region and the P region. Recall that $\gamma_{QD} = c/(\lambda\pi)$ satisfies the smooth-pasting condition joining the Q region and the D region. As k rises, γ_{PD} decreases, γ_{QP} increases, and γ_{QD} remains unchanged.

We are ready to present the characterization of the unique optimal policy in the positive information acquisition model.

Proposition 1. *In the positive information acquisition model:*

(i) If $k \geq k_P$, then the optimal policy is Q when $\gamma \leq \gamma_{QD}$, and D when $\gamma > \gamma_{QD}$.

(ii) If $k < k_P$, then there exists a non-degenerate positive information acquisition region $(\gamma_{QP}, \gamma_{PD}]$ containing γ_{QD}, such that the optimal policy is Q when $\gamma \leq \gamma_{QP}$, P when $\gamma \in (\gamma_{QP}, \gamma_{PD}]$, and D when $\gamma > \gamma_{PD}$.

Proof. By (6), the sign of $V_P'(\gamma) - V_D'(\gamma)$ can change only once as γ increases, and only from positive to negative. By Lemma 2, there is at most one interval of beliefs for which P is optimal. Clearly, D is optimal for γ sufficiently close to 1 and Q is optimal for γ sufficiently close to zero. Thus, the unique candidate for optimal policies that call for P for some beliefs is given by: Q when $\gamma \leq \gamma_{QP}$, P when $\gamma \in (\gamma_{QP}, \gamma_{PD})$, and D for $\gamma \geq \gamma_{PD}$.

It is straightforward to verify that $\gamma_{QP} > \gamma_{QD} > \gamma_{PD}$ if and only if $k > k_P$, and $\gamma_{QP} < \gamma_{QD} < \gamma_{PD}$ if and only if $k < k_P$. For case (i), suppose that P is optimal for some beliefs. Since $k \geq k_P$ implies $\gamma_{PD} \leq \gamma_{QP}$, the unique candidate optimal policy involving P is not feasible. Thus, P is not part of the optimal policy. The optimal policy in this case is then the same as when the agent can only choose between D and Q.

For case (ii), suppose that P is not part of the optimal policy. Then, D is optimal at γ just above γ_{QD}. But since $k < k_P$, we have $\gamma_{QD} < \gamma_{PD}$. This implies that D is optimal at $\gamma < \gamma_{PD}$, contradicting Lemma 2(i). Thus, P is optimal for some beliefs. The unique candidate policy involving P is thus the optimal policy. ■

Figure 1 illustrates the optimal policy described in Proposition 1. In case (i), information acquisition cost is so high that the agent's optimal policy is the same as when positive information acquisition is unavailable. We interpret k_P as a measure of the *usefulness of positive information*, as it is the highest information acquisition cost for which P is chosen in the optimal policy for some belief. The usefulness of positive information is increasing in its news arrival rate α and in the reward π from successful experimentation. Further, it is non-monotone in the efficiency of experimentation. When λ/c is large, experimentation is very efficient and there is little need for further costly information acquisition. Experimentation alone is enough to enable the agent to decide optimally whether or not to quit the risky arm. Therefore raising the efficiency of experimentation lowers the usefulness of positive information. If λ/c is too small experimentation is very inefficient, but experimenting is necessary to obtain the reward. Information acquisition instead does not directly help obtain the reward and would cut further into an already thin expected surplus. In this case, raising the efficiency of experimentation will increase the usefulness of positive information.

When positive information acquisition is used in the agent's optimal policy, it raises the value function and enables the agent to quit at a more pessimistic belief about the state.[7] That is, γ_{QP} is strictly below γ_{QD}. Recall that γ_{QD} is the belief that maximizes the benefit of positive information. Proposition 1 implies that the positive

[7]The value function $V(\gamma)$ can be found by first solving the differential equation (5) with boundary condition $V(\gamma_{QP}) = 0$, and then use the solution value at γ_{PD} as the boundary condition to solve the differential equation (2).

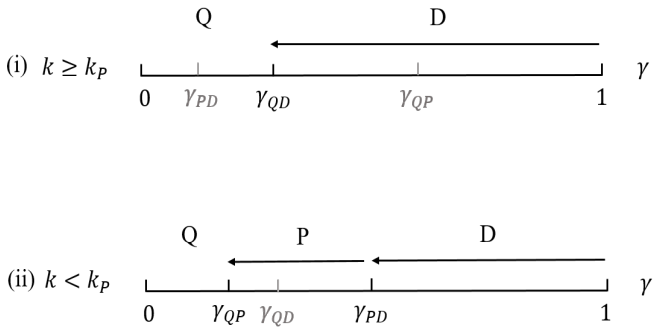

Figure 1. *Optimal policy in the positive information acquisition model. The direction of the arrows indicate how the belief evolves when the risky arm brings no success and information acquisition brings no news.*

information acquisition region must contain γ_{QD}. As we have seen from the the discussion of the benefit of positive information, positive information acquisition is used as a last ditch effort before abandoning the risky arm permanently. Consistent with the fact that $B_P(\gamma)$ decreases in γ whenever the agent experiments with the risky arm, the agent optimally refrains from positive information acquisition if it is sufficiently optimistic about the state (i.e., if $\gamma > \gamma_{PD}$).

3.2. Negative information acquisition

As in positive information acquisition, we first derive an expression for the *benefit of negative information*, $B_N(\gamma)$, defined as the time rate of change in the expected payoff to the agent who gains access to free negative information for an instant of time dt before going back to his optimal experimentation policy. Suppose the belief is above γ_{QD}. Conditional on $\mathcal{B}$, with probability αdt the agent learns the state, in which case it will optimally quit, with a payoff of 0. In the absence of either a success from the risky arm or negative news, the updated belief is $\gamma + \mathrm{d}\gamma$, with $\mathrm{d}\gamma = \gamma(1-\gamma)(\lambda - \alpha)$. For $\gamma > \gamma_{QD}$, the benefit of negative information is defined by:

$$B_N(\gamma)\mathrm{d}t = [-c\mathrm{d}t + \gamma\lambda\pi\mathrm{d}t + (1 - \gamma\lambda\mathrm{d}t - (1-\gamma)\alpha\mathrm{d}t)V_D(\gamma + \mathrm{d}\gamma)] - V_D(\gamma),$$

which gives

$$B_N(\gamma) = \alpha\left[\gamma\pi - c/\lambda - V_D(\gamma)\right]. \tag{7}$$

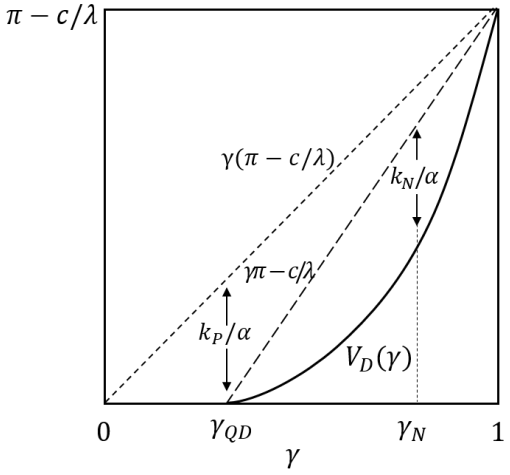

Figure 2. *The benefit of negative information is the news arrival rate α times the distance between the dashed line and $V_D(\gamma)$. This benefit is maximized at γ_N. The benefit of positive information is α times the distance between the dashed line and the dotted line. This benefit is maximized at γ_{QD}.*

Because $V_D(1) = \pi - c/\lambda$ and $V_D(\gamma_{QD}) = 0$, the term $\gamma\pi - c/\lambda$ in equation (7) is shown by the chord between $V_D(1)$ and $V_D(\gamma_{QD})$ in Figure 2. The benefit of negative information $B_N(\gamma)$ is α times the distance between this chord and $V_D(\gamma)$. Also note that equations (4) and (7) together imply that

$$B_P(\gamma) + B_N(\gamma) = \alpha \left[\gamma V_D(1) + (1-\gamma)V_D(0) - V_D(\gamma)\right].$$

Thus, the total benefit of positive and negative information is proportional to the value of an experiment that reveals the true state. See Figure 2.

Since V_D is convex, the benefit of negative information is concave for $\gamma > \gamma_{QD}$. This means that negative information is particularly valuable when the agent has intermediate beliefs about the state. Unlike positive information, the benefit of negative information is zero when the agent is about to quit, because learning that the state is $\mathcal{B}$ does not change the agent's decision. As the agent becomes more optimistic, the benefit initially increases, but it eventually decreases because negative information acquisition is unlikely to generate any news when the state is likely to be $\mathcal{G}$. Thus, the benefit of negative information stems from saving the wasteful effort in using the risky arm when the agent is still optimistic about it but the state is actually $\mathcal{B}$.

Negative information acquisition slows down the belief downgrading resulting from having no success from the risky arm. For a small interval of time $\mathrm{d}t$ and a fixed information acquisition intensity $\delta \in (0,1]$, in the absence of a success from the risky arm or negative news, the differential equation for belief updating is

$$\mathrm{d}\gamma = -\gamma(1-\gamma)(\lambda - \delta\alpha)\mathrm{d}t.$$

Thus, the belief can even go up if $\delta\alpha$ is greater than than λ. When $\alpha \geq \lambda$, we define $\delta^* \equiv \lambda/\alpha$ to be the intensity of negative information acquisition at which the agent's belief does not change.

Suppose that N at intensity $\delta \in (0,1]$ is optimal. The value function $V(\gamma)$ satisfies the Bellman equation

$$V(\gamma) = -(c+\delta k)\mathrm{d}t + \gamma\lambda\pi\mathrm{d}t + (1-(\gamma\lambda + (1-\gamma)\delta\alpha)\mathrm{d}t)V(\gamma + \mathrm{d}\gamma),$$

where we have used the fact it is optimal to quit immediately if negative news reveals the state to be $\mathcal{B}$. Using the differential equation for belief updating, we obtain a differential equation for the value function:

$$\gamma(1-\gamma)(\lambda - \delta\alpha)V'_{N,\delta}(\gamma) = -(c+\delta k) + \gamma\lambda\pi - (\gamma\lambda + (1-\gamma)\delta\alpha)V(\gamma). \tag{8}$$

In the special case of the N region where the level is equal to δ^*, the value function satisfies

$$V^*(\gamma) = -(c+\delta^* k)\mathrm{d}t + \gamma\lambda\pi\mathrm{d}t + (1-(\gamma\lambda + (1-\gamma)\delta^*\alpha)\mathrm{d}t)\, V^*(\gamma),$$

which yields:

$$V^*(\gamma) = \gamma\pi - c/\lambda - k/\alpha.$$

Lemma 3. *Suppose that $\alpha \neq \lambda$. There is no interval of beliefs for which N at level $\delta < 1$ is optimal.*

The proof of Lemma 3 is in the appendix. For the remainder of this section, we write V_N instead of $V_{N,1}$ in the belief region where N is optimal.

Lemma 4. *At a given belief γ: (i) if N is optimal, then $V'_N(\gamma) \geq V'_D(\gamma)$; (ii) if D is optimal, then $V'_D(\gamma) \geq V'_N(\gamma)$ when $\alpha < \lambda$ and $V'_D(\gamma) \leq V'_N(\gamma)$ when $\alpha > \lambda$; and (iii) if Q is optimal, then $V'_D(\gamma) \leq 0$, and $V'_N(\gamma) \leq 0$ when $\alpha < \lambda$ and $V'_N(\gamma) \geq 0$ when $\alpha > \lambda$.*

Proof. Part (i) can be established in the same way as in Lemma 2. For part (ii), suppose that D is optimal in the neighborhood of γ. Then for all $\delta \in (0, 1]$,

$$\begin{aligned} V(\gamma) \geq & -(c + \delta k)\mathrm{d}t + \gamma\lambda\pi\mathrm{d}t \\ & + (1 - \gamma\lambda\mathrm{d}t - (1-\gamma)\delta\alpha\mathrm{d}t)\left(V(\gamma) - \gamma(1-\gamma)(\lambda - \delta\alpha)V_D'(\gamma)\mathrm{d}t\right). \end{aligned}$$

In particular, the above holds for $\delta = 1$, and thus

$$\begin{aligned} 0 &\geq -(c+k) + \gamma\lambda - (\gamma\lambda + (1-\gamma)\alpha)V(\gamma) - \gamma(1-\gamma)(\lambda - \alpha)V_D'(\gamma) \\ &= \gamma(1-\gamma)(\lambda - \alpha)\left(V_N'(\gamma) - V_D'(\gamma)\right). \end{aligned}$$

It follows that $V_D'(\gamma) \geq V_N'(\gamma)$ if $\alpha < \lambda$ and the reverse holds if $\alpha > \lambda$. Part (iii) is similar to (ii). ∎

Lemma 4 implies the standard smooth-pasting condition when $\alpha < \lambda$. But smooth-pasting may not hold when $\alpha > \lambda$, because $V_N'(\gamma) \geq V_D'(\gamma)$ regardless of whether N or D is optimal in the neighborhood of γ (and also because it is possible that $V_N'(\gamma) > 0$ at the switching point between Q and N). The first case is simpler because the belief cannot go up even when N is optimal; the second case leaves that possibility open. As a result, to characterize the optimal policy in the negative information model, we need to consider two cases: $\alpha < \lambda$ and $\alpha \geq \lambda$.

3.2.1. Slow negative information acquisition: $\alpha < \lambda$

From (2) and (8) we have

$$\gamma(1-\gamma)(\lambda - \alpha)(V_N'(\gamma) - V_D'(\gamma)) = \alpha\left(\gamma\pi - V(\gamma) - c/\lambda\right) - k. \tag{9}$$

The right-hand side of the above equation is the benefit of negative information minus its cost. To state the characterization of the optimal policy, define:

$$k_N \equiv \max_{\gamma} B_N(\gamma),$$

and let $\gamma_N = \arg\max_\gamma B_N(\gamma)$. Obviously, $V_N'(\gamma) < V_D'(\gamma)$ for all γ if $k > k_N$. Figure 3 illustrate the optimal policy in the slow negative information case.

Proposition 2. *In the negative information acquisition model with $\alpha < \lambda$:*

(i) If $k \geq k_N$, the optimal policy is Q when $\gamma \leq \gamma_{QD}$, and D when $\gamma > \gamma_{QD}$.

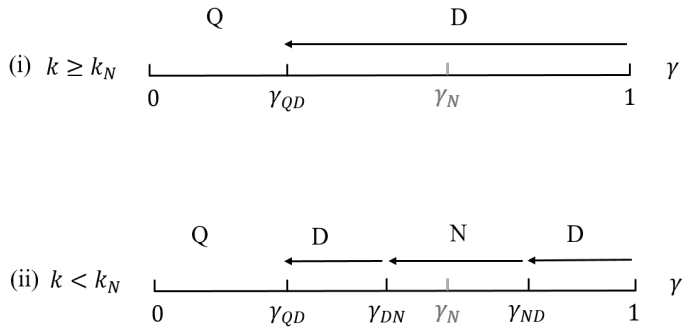

Figure 3. *Optimal policy under slow negative information acquisition.*

(ii) If $k < k_N$, there exists a non-degenerate negative information acquisition region $(\gamma_{DN}, \gamma_{ND}]$ containing γ_N, such that the optimal policy is Q when $\gamma \leq \gamma_{QD}$, D when $\gamma \in (\gamma_{QD}, \gamma_{DN}]$, N when $\gamma \in (\gamma_{DN}, \gamma_{ND}]$, and D when $\gamma > \gamma_{ND}$.

Proof. As γ approaches 1, $V(\gamma)$ goes to $\pi - c/\lambda$. From (9), the sign of $V_N'(\gamma) - V_D'(\gamma)$ is negative, and D is optimal when γ is sufficiently close to 1. Further, at $\gamma = \gamma_{QD}$, the sign of $V_N'(\gamma) - V_D'(\gamma)$ is also negative as $V(\gamma_{QD}) = 0$. Finally, since $V_D(\gamma)$ is convex, it follows from part (ii) of Lemma 4 that $V_N'(\gamma) - V_D'(\gamma)$ is concave whenever it is negative. Thus, the equation $V_N'(\gamma) - V_D'(\gamma) = 0$ either has no solution and hence $V_N'(\gamma) < V_D'(\gamma)$ for all γ, or has two solutions with $V_N'(\gamma) \geq V_D'(\gamma)$ if and only if γ is between the two solutions. Putting all these together, we find that in the unique candidate for the optimal policy in which N is used for some beliefs, the agent chooses Q for $\gamma \leq \gamma_{QD}$, N for some interval and D otherwise.

For (i), suppose that $k \geq k_N$ and N is optimal in a neighborhood of some belief γ. Then, $\gamma > \gamma_{QD}$ and $V(\gamma) \geq V_D(\gamma)$, where V_D again is the value function given by equation (3) when there is no information acquisition. But then $V_N'(\gamma) - V_D'(\gamma)$ has the same sign as

$$\alpha\left(\gamma\pi - V(\gamma) - c/\lambda\right) - k \leq B_N(\gamma) - k \leq 0,$$

with strict inequality except when $k = k_N$ and $\gamma = \gamma_N$, which contradicts part (i) of Lemma 4. Thus, N is not part of the optimal policy. The statement of (i) then follows immediately.

For (ii), suppose $k < k_N$ and N is never optimal. Then, we have $V(\gamma) = V_D(\gamma)$ for all $\gamma > \gamma_{QD}$. But this implies $V_N'(\gamma_N) - V_D'(\gamma_N) > 0$ given that $k < k_N$, which contradicts Lemma 4(ii). Thus, N is optimal for some belief above γ_{QD}. The optimal policy is given by the unique candidate where N is chosen in some interval $(\gamma_{DN}, \gamma_{ND}]$, where γ_{DN} and γ_{ND} are the two solutions to $V_N'(\gamma) - V_D'(\gamma) = 0$ above γ_{QD}.

To show that the negative information acquisition region must contain γ_N, note that γ_{DN} is the smaller root of the equation $B_N(\gamma) - k = 0$. Since γ_N maximizes the concave function $B_N(\gamma) - k$, we have $\gamma_{DN} < \gamma_N$. Next, γ_{ND} is the larger root to the equation $\alpha(\gamma\pi - V(\gamma) - c/\lambda) - k = 0$, and $V(\gamma)$ is convex. Therefore, $V'(\gamma_{ND}) > \pi$. Because γ_{ND} is the boundary between the N region and the D region, smooth-pasting implies that $V_D'(\gamma_{ND}; \gamma_{ND}) = V'(\gamma_{ND}) > \pi$, where the function $V_D(\cdot; \gamma_{ND})$ is the solution to the differential equation (2) with boundary condition that its value at γ_{ND} is $V(\gamma_{ND})$. Recall that $V_D(\cdot)$ given in equation (3) is the solution to the same differential equation (2) but satisfying a different boundary condition that its value at γ_{ND} is equal to $V_D(\gamma_{ND})$. Because N is preferred to D for $\gamma \in (\gamma_{DN}, \gamma_{ND}]$, we have $V_D(\gamma_{ND}) \leq V(\gamma_{ND})$, which by the differential equation (2) implies that $V_D'(\gamma_{ND}) \geq V_D'(\gamma_{ND}; \gamma_{ND}) > \pi$. Since $V_D'(\gamma_N) = \pi$, the strict convexity of $V_D(\gamma)$ implies that $\gamma_{ND} > \gamma_N$. ∎

As in the positive information model, slow negative information acquisition is not used if its cost k is sufficiently high or the evidence discovery rate α is sufficiently low. The optimal policy in case (i) of Proposition 2 is the same as when the agent chooses between Q and D only. We can interpret k_N as a measure of the *usefulness of negative information* in terms of its cost. Having summarized the usefullness of information in a single parameter, for a fixed arrival rate, α, we can compare the usefulness of negative vs. positive information. We have the following proposition.

Proposition 3. *For positive and negative information acquisition with the same α, there is a critical value of the efficiency of the risky arm such that negative information is more useful than positive information if and only if λ/c is larger than that critical value.*

Proof. Using the explicit formula for $V_D(\gamma)$ given in (3), the first-order condition for maximizing $B_N(\gamma)$ can be written as

$$\log \frac{\gamma_N}{1-\gamma_N} - \log \frac{\gamma_{QD}}{1-\gamma_{QD}} = \frac{1}{\gamma_N}.$$

Since $\gamma_{QD} = c/(\lambda\pi)$, γ_N strictly decreases from 1 to 0 as λ/c increases from π to infinity. Moreover, substituting the above first-order condition into the explicit formula for $B_N(\gamma_N)$, we obtain

$$k_N = \alpha \frac{c}{\lambda} \frac{1-\gamma_N}{\gamma_N}.$$

Comparing this to the expression for k_P, we have that $k_N > k_P$ if and only if $(1-\gamma_N)/\gamma_N > 1-\gamma_{QD}$. Use the first-order condition again to eliminate γ_{QD}, this condition is equivalent to

$$\gamma_N - \mathrm{e}^{1/\gamma_N}(2\gamma_N - 1) > 0.$$

The left-hand-side is decreasing in γ_N, and is positive for γ_N close to 0 and negative for γ_N close to 1. Hence there is a critical value for γ_N such that the expression is positive for γ_N lower than that critical value. The proposition then follows because the relation between γ_N and λ/c is monotone. ■

We have shown in Section 3.1 that positive information is used as a last ditch effort before abandoning the risky arm. In contrast, there is little value in having conclusive negative news for the state being $\mathcal{B}$ when the agent is so pessimistic that it is about to quit. In case (ii) of Proposition 2, the agent's optimal policy is D when the belief γ is just above γ_{QD}, even if negative information is very cheap. Instead, negative information is more valuable when the agent is relatively optimistic, as it can potentially avoid waiting too long to quit. Thus, slow negative information acquisition can be interpreted an insurance strategy against the risk of mistaking the state to be $\mathcal{G}$ when it is actually $\mathcal{B}$. An insurance strategy is more useful when the risk is relatively high, while a last ditch effort is more useful when the upside potential is relatively high, which explains why slow negative information is more useful than positive information when the risky arm is relatively efficient (i.e., when λ/c is high).

Whenever N is chosen, the negative information acquisition region must contain γ_N, the belief that maximizes the benefit of negative information. This belief is higher than γ_{QD}, the belief that maximizes the benefit of positive information. The comparison between negative information acquisition as an insurance strategy and positive information acquisition as a last ditch effort suggests that negative information acquisition tends to be chosen when beliefs are more optimistic. To formalize

this comparison, we say that an interval of beliefs is *higher than* another interval if both the upper bound and the lower bound of the first interval are higher than those of the second. We say that an interval of beliefs *expands* if the upper bound increases and the lower bound decreases.

Proposition 4. *For positive and negative information with the same $k < \min\{k_P, k_N\}$ and the same $\alpha < \lambda$, the information acquisition region with negative information is higher than with positive information. Moreover, the region of information acquisition expands with both negative and positive information as k falls.*

Proof. For the lower bound, Proposition 2 establishes that $\gamma_{DN} > \gamma_{QD}$. Moreover, it can be verified that $\gamma_{QD} > \gamma_{QP}$ when $k < k_P$. Thus, $\gamma_{DN} > \gamma_{QP}$.

For the upper bound, the proof of Proposition 2 shows that $V'(\gamma_{ND}) > \pi$ and $V(\gamma_{ND}) = \gamma_{ND}\pi - c/\lambda - k/\alpha$. Substitute these values into the differential equation (2) yields

$$\gamma_{ND} > \frac{c/\lambda}{c/\lambda + k/\alpha}.$$

The right-hand-side of the above is greater than γ_{PD}. Thus, $\gamma_{ND} > \gamma_{PD}$.

Finally, it is straightforward to verify from the explicit formulas for γ_{QP} and γ_{PD} that the positive information acquisition region expands as k falls. For the negative region, the lower bound γ_{DN} is the smaller root to the equation $B_N(\gamma) - k = 0$. Since $B_N(\gamma)$ is concave and does not depend on k, the smaller root γ_{DN} decreases as k falls. To show that the upper bound γ_{ND} increases as k falls, let $V_N(\gamma)$ be the solution to the differential equation (8). The general solution is

$$V_N(\gamma) = \gamma\pi - \gamma\frac{c+k}{\lambda} - (1-\gamma)\frac{c+k}{\alpha} + \gamma^{\frac{-\alpha}{\lambda-\alpha}}(1-\gamma)^{\frac{\lambda}{\lambda-\alpha}}A,$$

for some constant A. Imposing the boundary condition $V_N(\gamma_{DN}) = \gamma_{DN}\pi - c/\lambda - k/\alpha$, we obtain:

$$A = \frac{\lambda-\alpha}{\lambda\alpha}\left((1-\gamma_{DN})c - \gamma_{DN}k\right)\gamma_{DN}^{\frac{\alpha}{\lambda-\alpha}}(1-\gamma_{DN})^{\frac{-\lambda}{\lambda-\alpha}}.$$

The upper bound of the negative information acquisition region, γ_{ND}, is the larger

root that satisfies the equation, $V_N(\gamma_{ND}) = \gamma_{ND}\pi - c/\lambda - k/\alpha$. Therefore,

$$\begin{aligned}
\left(V_N'(\gamma_{ND}) - \pi\right)\frac{\partial \gamma_{ND}}{\partial k} &= -\frac{1}{\alpha} + \gamma_{ND}\frac{1}{\lambda} + (1-\gamma_{ND})\frac{1}{\alpha} - \gamma_{ND}^{\frac{-\alpha}{\lambda-\alpha}}(1-\gamma_{ND})^{\frac{\lambda}{\lambda-\alpha}}\frac{\partial A}{\partial k} \\
&= -\gamma_{ND}\frac{\lambda-\alpha}{\lambda\alpha} - \left(\frac{\gamma_{ND}}{\gamma_{DN}}\right)^{\frac{-\alpha}{\lambda-\alpha}}\left(\frac{1-\gamma_{ND}}{1-\gamma_{DN}}\right)^{\frac{\lambda}{\lambda-\alpha}}\frac{\lambda-\alpha}{\lambda\alpha} \\
&\quad \times \left(-\gamma_{DN} + \left(\frac{\alpha}{\lambda-\alpha}\frac{c}{\gamma_{DN}} - \frac{\lambda}{\lambda-\alpha}\frac{k}{1-\gamma_{DN}}\right)\frac{\partial \gamma_{DN}}{\partial k}\right) \\
&= \gamma_{ND}\frac{\lambda-\alpha}{\lambda\alpha}\left(\left(\frac{(1-\gamma_{ND})/\gamma_{ND}}{(1-\gamma_{DN})/\gamma_{DN}}\right)^{\frac{\lambda}{\lambda-\alpha}} - 1\right) \\
&\quad + \left(\frac{\gamma_{ND}}{\gamma_{DN}}\right)^{\frac{-\alpha}{\lambda-\alpha}}\left(\frac{1-\gamma_{ND}}{1-\gamma_{DN}}\right)^{\frac{\lambda}{\lambda-\alpha}}\left(\frac{k}{(1-\gamma_{DN})\alpha} - \frac{c}{\gamma_{DN}\lambda}\right)\frac{\partial \gamma_{DN}}{\partial k}
\end{aligned}$$

The first-term on the right-hand-side is negative because $\gamma_{ND} > \gamma_{DN}$. The second term is also negative because

$$V_D'(\gamma_{DN})\gamma_{DN}(1-\gamma_{DN})\lambda = -c + \gamma_{DN}\lambda - \gamma_{DN}\lambda\left(\gamma_{DN}\pi - \frac{c}{\lambda} - \frac{k}{\alpha}\right),$$

which implies

$$\frac{k}{(1-\gamma_{DN})\alpha} - \frac{c}{\gamma_{DN}\lambda} = V_D'(\gamma_{DN}) - \pi < 0.$$

Finally, $V_N'(\gamma_{ND}) > \pi$ because γ_{ND} is the larger root. Hence, γ_{ND} decreases in k. ■

Other things equal, a fall in negative information acquisition cost k raises the attractiveness of N relative to D. However, a lower k also raises the value $V(\gamma_{ND})$ at the upper threshold of the negative information acquisition region. This tends to reduce the attractiveness of N relative to D, as shown in equation (9). Proposition 4 shows that the first effect dominates, so that the negative region always expands as information acquisition cost falls.

The lower bound γ_{DN} of the negative information acquisition region may be greater than or smaller than the upper bound γ_{PD} of the positive region. For example, if k is just below k_N and $k_N < k_P$, γ_{DN} is very close to γ_{ND}, which by Proposition 4 strictly exceeds γ_{PD}; hence the two regions do not overlap. If k is close to 0, γ_{DN} goes to γ_{QD}, while γ_{PD} goes to 1; hence the two information acquisition regions do overlap.

3.2.2. Fast negative information acquisition: $\alpha \geq \lambda$

Clearly, D is optimal when the belief γ is sufficiently close to 1. Suppose that there is some $\gamma^* \in (\gamma_{QD}, 1)$ such that it is optimal to choose N just below γ^* and D in the region $(\gamma^*, 1]$. Because the belief goes down when it is above γ^* and the belief goes up when it is below γ^*, in the absence of any discovery or news arrival the belief remains at γ^* once it reaches there. This means that $V(\gamma^*)$ is equal to the payoff from N at intensity $\delta^* = \lambda/\alpha$ so that the belief remains stationary. We have

$$V(\gamma^*) = V^*(\gamma^*) = \gamma^*\pi - c/\lambda - k/\alpha.$$

Furthermore, since the policy of N at intensity δ^* is feasible, optimality of N at full intensity just below γ^* implies that $V(\gamma) \geq V^*(\gamma)$ for γ slightly below γ^*. Likewise, we have $V(\gamma) \geq V^*(\gamma)$ for γ slightly above γ^*. Together they imply

$$\lim_{\gamma \uparrow \gamma^*} V_N'(\gamma) \leq V^{*\prime}(\gamma) = \pi \leq \lim_{\gamma \downarrow \gamma^*} V_D'(\gamma).$$

Lemma 4, however, requires that $V_N'(\gamma) \geq V_D'(\gamma)$ for γ in the neighborhood of γ^*. Thus, we must have $V_N'(\gamma^*) = V_D'(\gamma^*) = \pi$. Since $V(\gamma^*) = V^*(\gamma^*)$ and $V_D'(\gamma^*) = \pi$, we can use the differential equation (2) to solve for γ^*:

$$\gamma^* = \frac{c/\lambda}{c/\lambda + k/\alpha}.$$

This is the highest value of γ for which N can be optimal.

To determine the lowest value of γ for which N is optimal, for the generic case of $\alpha > \lambda$, let $V_N(\gamma)$ be the solution to the differential equation (8) with boundary condition $V_N(\gamma^*) = V^*(\gamma^*)$. As shown by the following proposition, the optimal policy changes qualitatively depending on whether the function $V_N(\gamma)$ reaches zero before or after the belief γ_{QD}. In the former case the optimal policy transitions from N to Q as the belief becomes smaller, while in the latter case there is a region of D before quitting. As the value of $V_N(\gamma_{QD})$ is monotone in the cost of information acquisition, we define $k_D \in (0, k_N)$ to be the value of k such that $V_N(\gamma_{QD}) = 0$.[8] For the knife-edge case of $\alpha = \lambda$, we define $k_D = 0$.

[8] When $k = k_N$, $V_N(\gamma_{QD}) < V_D(\gamma_{QD}) = 0$ because $V_N(\gamma^*) = V_D(\gamma^*)$ and $V_N'(\gamma) > V_D'(\gamma)$ for $\gamma < \gamma^*$. When $k = 0$, the explicit solution to the differential equation gives $V_N(\gamma_{QD}) > 0$. Since $V_N(\gamma_{QD})$ is decreasing in k, k_D exists and is unique.

Proposition 5. *In the negative information acquisition model with $\alpha \geq \lambda$:*

(i) If $k \geq k_N$, the optimal policy is Q when $\gamma \leq \gamma_{QD}$ and D when $\gamma > \gamma_{QD}$.

(ii) If $k \in [k_D, k_N)$, then there exists a non-degenerate negative information acquisition region $(\gamma_{DN}, \gamma^*]$ containing γ_N such that the optimal policy is Q when $\gamma \leq \gamma_{QD}$, D when $\gamma \in [\gamma_{QD}, \gamma_{DN}]$, N when $\gamma \in (\gamma_{DN}, \gamma^*]$, and D when $\gamma > \gamma^*$.

(iii) If $k < k_D$, then there exists $\gamma_{QN} < \gamma_{QD}$ such that the optimal policy is Q when $\gamma \leq \gamma_{QN}$, N when $\gamma \in (\gamma_{QN}, \gamma^*]$, and D when $\gamma > \gamma^*$.

Proof. For now, assume that $\alpha > \lambda$. Since by construction γ^* is the highest belief for which N can be optimal, by convexity of the value function, a necessary condition for N to be optimal for some interval of beliefs is that $V^*(\gamma^*) > V_D(\gamma^*)$, where V_D solves the differential equation (2) with boundary condition $V_D(\gamma_{QD}) = 0$. But $V^*(\gamma) - V_D(\gamma) = (B_N(\gamma) - k)/\alpha$ is non-positive for $k \geq k_N$. Case (i) follows immediately.

Next, observe that when $k = k_N$, we have $\gamma^* = \gamma_N$, because γ^* satisfies the condition that $B'_N(\gamma^*) = 0$ and $B_N(\gamma^*) = k$ (the same condition that defines γ_N). Since γ^* decreases in k while γ_N does not depend on k, we have $\gamma^* > \gamma_N$ if and only if $k < k_N$.

Suppose $k < k_N$. By definition of γ^*, $V'_D(\gamma^*) = \pi$ if $V_D(\gamma^*) = V^*(\gamma^*)$. From the differential equation (2), $V'_D(\gamma^*) \leq \pi$ if $V_D(\gamma^*) \geq V^*(\gamma^*)$. Therefore, by the convexity of V_D, $V_D(\gamma^*) \geq V^*(\gamma^*)$ implies $V'_D(\gamma) \leq \pi$ for all $\gamma \leq \gamma^*$, which in turn implies $V_D(\gamma_N) > V^*(\gamma_N)$, a contradiction. We conclude that if $k < k_N$ then $V^*(\gamma^*) > V_D(\gamma^*)$. By construction, the convexity of V_N implies that $V_N(\gamma) > V^*(\gamma) > V_D(\gamma)$ for γ just below γ^*, so N is optimal. If $k \geq k_D$, we have $V_N(\gamma_{QD}) \leq 0 = V_D(\gamma_{QD})$. But since $V_N(\gamma^*) = V^*(\gamma^*) > V_D(\gamma^*)$, there exists $\gamma_{DN} \in [\gamma_{QD}, \gamma^*)$ such that $V_N(\gamma_{DN}) = V_D(\gamma_{DN})$. Furthermore, $V_N(\gamma) > V^*(\gamma)$ for all $\gamma < \gamma^*$ by the convexity of V_N. From (9) we have that $V'_N(\gamma) - V'_D(\gamma)$ crosses zero only once and only from below. Thus, γ_{DN} is uniquely defined. The interval for which N is optimal is $(\gamma_{DN}, \gamma^*]$. Case (ii) then follows.

If $k < k_D$, then $V_N(\gamma_{QD}) > 0$. There exists $\gamma_{QN} < \gamma_{QD}$ such that $V_N(\gamma_{QN}) = 0$. By the single-crossing property of $V'_N(\gamma) - V'_D(\gamma)$ established above, $V_N(\gamma) \geq V_D(\gamma)$ for all $\gamma \leq \gamma^*$, so D is not optimal. Case (iii) follows immediately.

Finally, in the knife-edge case of $\alpha = \lambda$, we have $V(\gamma) = V^*(\gamma)$ whenever N is optimal. It is easy to verify that $V^*(\gamma_{QD}) < 0$ for any $k > 0$. The characterization of

(i) and (ii) above applies without change. ■

The usefulness of fast negative information has the same expression as that of slow negative information. This is not a coincidence: in the case of $\alpha < \lambda$, we have defined γ_N as the maximizer of the value of negative information, with the first-order condition $V_D'(\gamma_N) = \pi$, and the usefulness of negative information as the cost k such that $V_D(\gamma_N) = V^*(\gamma_N)$. These are the same two conditions satisfied by γ^* at the threshold cost k between case (i) and case (ii) of Proposition 5. As a result, the usefulness of negative information is proportional to its news arrival rate, the faster the more useful.

The comparison of the regions between positive and negative information described in Proposition 4 still holds in the case of fast negative information. The negative information acquisition region when $\alpha \geq \lambda$ is either $(\gamma_{DN}, \gamma^*]$ in case (ii), or $(\gamma_{QN}, \gamma^*]$ in case (iii) of Proposition 5. A direct comparison of the upper bound of the regions establishes that

$$\gamma^* > \gamma_{PD}.$$

For the lower bound, when $k \leq \min\{k_P, k_N\}$, in case (ii) of Proposition 5, we have

$$\gamma_{DN} \geq \gamma_{QD} \geq \gamma_{QP}.$$

In case (iii), by Lemma 4, a necessary condition for N to be optimal just above γ_{QN} is $V_N'(\gamma_{QN}) \geq 0$. From the expression (8) for $V_N'(\gamma)$, using $V_N(\gamma_{QN}) \geq 0$ and $\alpha > \lambda$ we have

$$\gamma_{QN} \geq (c + k)/\lambda > \gamma_{QP}.$$

Thus, as with slow negative information acquisition, fast negative information is optimally used when the agent is relatively optimistic, compared with positive information acquisition, which is optimally used when the agent is relatively pessimistic.

The comparison between fast negative information and slow negative information is not straightforward because information acquisition becomes more efficient for the same cost as α rises above λ. To preserve the same efficiency under the two cases, we fix the output-to-cost ratio α/k when making the comparison. The following result shows that, for the same efficiency level, fast negative information acquisition is used when the agent is relatively pessimistic while slow negative information acquisition is used when it is more optimistic.

Proposition 6. *Suppose $\alpha/k = \alpha'/k'$ with $\alpha < \lambda < \alpha'$. Then, the information acquisition region under slow negative information (i.e., k and α) is higher than that under fast negative information (i.e., k' and α').*

Proof. Because the output-to-cost ratio is the same for these two cases, the value of γ^* and the $V^*(\gamma)$ function do not depend on whether negative information acquisition is fast or slow. In the proof of Proposition 4, we already show that

$$\gamma_{ND}(k,\alpha) > \frac{c/\lambda}{c/\lambda + k/\alpha} = \frac{c/\lambda}{c/\lambda + k'/\alpha'} = \gamma^*(k',\alpha').$$

For the lower bound of the information acquisition region, suppose case (ii) of Proposition 5 is relevant. Then γ_{DN} is determined by the intersection between $V_D(\gamma)$ and $V^*(\gamma)$ in the slow negative information case, but by the intersection between $V_D(\gamma)$ and $V_N(\gamma)$ in the fast negative information case. Since $V^*(\gamma) < V_N(\gamma)$ for all $\gamma < \gamma^*$, the intersection point under the former (slow) case is above the intersection point under the latter (fast) case. If instead case (iii) of Proposition 5 is relevant, then

$$\gamma_{DN}(k,\alpha) > \gamma_{QD} > \gamma_{QN}(k',\alpha'). \qquad \blacksquare$$

Fast negative information acquisition critically differs from slow negative information acquisition in that the belief is driven up instead of down when N is used and there is no success from the risky arm or news from information acquisition. As a result, the standard smooth-pasting condition does not hold at the lower bound of the fast negative information acquisition region.[9] In case (ii) of Proposition 5, the proof establishes a single-crossing property of $V_N'(\gamma) - V_D'(\gamma)$, which implies that, consistent with Lemma 4,

$$\lim_{\gamma \uparrow \gamma_{DN}} V_D'(\gamma) < \lim_{\gamma \downarrow \gamma_{DN}} V_N'(\gamma).$$

The reason that smooth-pasting fails at γ_{DN} is that D is optimal at γ just below γ_{DN} so the continuation belief goes down, while N is optimal just above γ_{DN} so the continuation belief goes up. The kink in the value function implies completely different optimal dynamics on the two sides of γ_{DN}: to the left, the agent uses only the risky arm and quits when the belief reaches γ_{QD}; to the right, the agent engages

[9] The smooth-pasting condition does hold at the upper bound γ^*, as both the left and the right derivatives are equal to π.

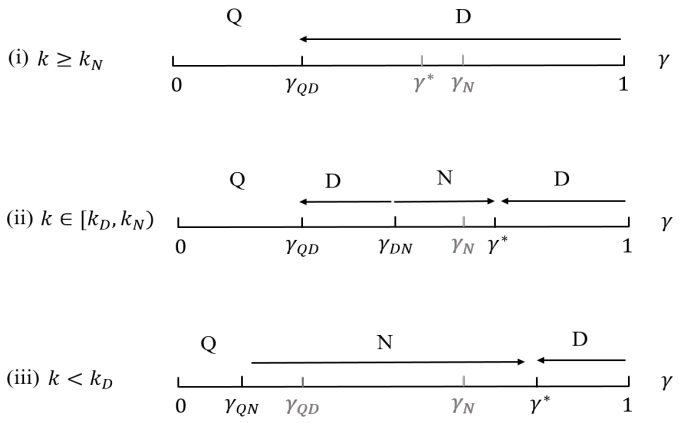

Figure 4. *Optimal policy under fast negative information. Smooth pasting fails at γ_{DN} in case (ii) and at γ_{QN} in case (iii). The belief is stationary at γ^* until either a success from the risky arm or the arrival of negative news.*

negative information acquisition in addition to using the risky arm and never quits unless he learns that the state is $\mathcal{B}$. See Figure 4. A similar failure of smooth-pasting occurs at γ_{QN} in case (iii) of Proposition 5, as

$$0 < \lim_{\gamma \downarrow \gamma_{QN}} V_N'(\gamma).$$

Even though the agent chooses N at a belief just higher than the quitting belief γ_{QN}, fast negative information is *not* a last-ditch effort before abandoning the risky arm. The agent quits immediately if the starting belief is just below γ_{QN}; above γ_{QN}, negative information acquisition is so fast that the agent optimally engages in both information acquisition and the risky arm, and never quits until the state is revealed.

Under fast negative information acquisition, the belief is driven up in the absence of success from the risky arm or arrival of negative news. When it reaches γ^*, there is "perpetual" negative information acquisition at the intensity of $\delta^* = \lambda/\alpha$.[10] This is an absorbing state which is reached from above where D is optimal but has so

[10]Instead of an information acquisition intensity of $\delta^* < 1$ at γ^*, one can imagine just below γ^* the agent chooses N at the full intensity, switches to just D at the full level immediately when the belief shoots above γ^*, switches back to N immediately when the belief falls back below γ^*, and so on.

far failed to deliver a success, or from below where N is optimal but has so far failed to reveal the negative news. The expected payoff associated with γ^* is simply $V^*(\gamma^*) = \gamma^*\pi - c/\lambda - k/\alpha$. The expression can be easily understood after noting that the expected duration of the discovery process in state $\mathcal{G}$ is the same as the expected time for news arrival in state $\mathcal{B}$ (i.e., $\lambda = \delta^*\alpha$). Thus the expected cost is to the agent is just the flow cost of discovery and information acquisition ($c + \delta^* k$) times the state independent expected duration $1/\lambda$.

4. Extensions

4.1. Joint information acquisition

Our separate analysis of positive and negative information acquisition can be extended straightforwardly to allow the agent to choose positive or negative information acquisition independently of each other. If the agent chooses both positive and negative information acquisition at the same time, we say that it engages in joint acquisition (choosing J). To simplify the analysis, we assume that the cost k and arrival rate α are identical for positive and negative information acquisition, and we consider only the case of slow information acquisition ($\alpha < \lambda$).

The expressions of $V'_D(\gamma)$, $V'_P(\gamma)$ and $V'_N(\gamma)$ remain the same, as (2), (5) and (8), respectively. When the agent chooses J for an instant $\mathrm{d}t$, the updated belief when there is no success from the risky arm, and no positive or negative news from joint information acquisition is $\gamma + \mathrm{d}\gamma$, with $\mathrm{d}\gamma = -\gamma(1-\gamma)\lambda \mathrm{d}t$. This is because, with the same news arrival rate α, the belief downgrading under positive information acquisition exactly cancels the belief upgrading under negative information acquisition. Correspondingly, $V'_J(\gamma)$ is given by

$$\gamma(1-\gamma)\lambda V'_J(\gamma) = -(c+2k) + \gamma(\lambda\pi + \alpha(\pi - c/\lambda)) - (\gamma\lambda + \alpha)V(\gamma). \tag{10}$$

It is straightforward to establish the following counterpart to Lemmas 2 and 4: if J is optimal in a neighborhood of belief γ, then $V'_J(\gamma) \geq \max\{V'_D(\gamma), V'_P(\gamma), V'_N(\gamma)\}$, and similar slope comparisons hold if the optimal policy is P, N or D. The comparison between P and D remains the same as in the positive information acquisition model, as it does not depend on the value function: $V'_P(\gamma) \geq V'_D(\gamma)$ if and only if

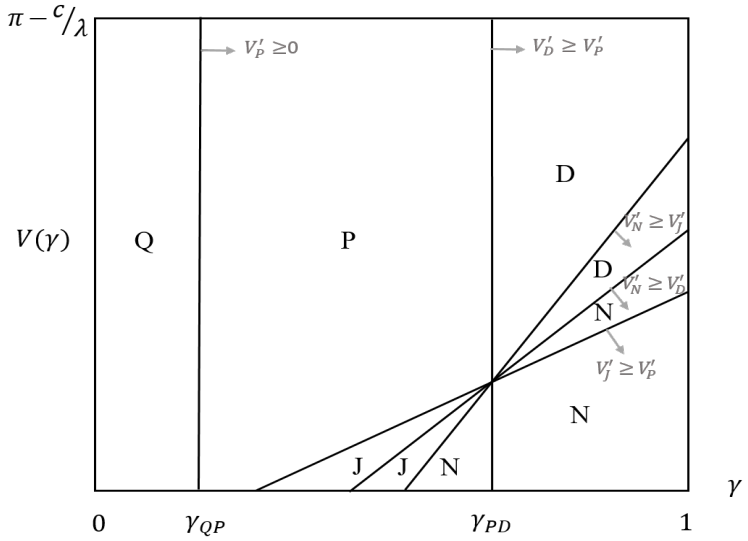

Figure 5. *Optimal policy in the (γ, V) space for the case $k < k_P$.*

$\gamma \leq \gamma_{PD}$. In addition, we have

$$
\begin{aligned}
V'_N(\gamma) \geq V'_J(\gamma) &\iff -\frac{c+k}{\alpha} + \frac{(\lambda-\alpha)}{\alpha}\frac{k}{\alpha} + \gamma\left(\pi + \frac{(\lambda-\alpha)}{\alpha}\frac{c}{\lambda}\right) \geq V(\gamma);\\
V'_N(\gamma) \geq V'_P(\gamma) &\iff -\frac{2(c+k)}{\lambda+\alpha} + \gamma\left(\pi + \frac{(\lambda-\alpha)}{(\lambda+\alpha)}\frac{c}{\lambda}\right) \geq V(\gamma);\\
V'_N(\gamma) \geq V'_D(\gamma) &\iff -\frac{c}{\lambda} - \frac{k}{\alpha} + \gamma\pi \geq V(\gamma);\\
V'_J(\gamma) \geq V'_P(\gamma) &\iff -\frac{c+k}{\lambda+\alpha} - \frac{k}{\alpha} + \gamma\left(\pi - \frac{\alpha}{(\lambda+\alpha)}\frac{c}{\lambda}\right) \geq V(\gamma);\\
V'_J(\gamma) \geq V'_D(\gamma) &\iff -\frac{2k}{\alpha} + \gamma\left(\pi - \frac{c}{\lambda}\right) \geq V(\gamma). \qquad (11)
\end{aligned}
$$

The linear functions on the left-hand sides of (11) all intersect at the same point of $\gamma = \gamma_{PD}$, and their slopes are ranked in decreasing order. These linear functions, together with the vertical line $\gamma = \gamma_{PD}$, divide the (γ, V) space into different regions, for which the optimal policy given any pair of γ and $V(\gamma)$ can be determined by comparing the slopes of the value functions. Figure 5 depicts such comparison for the case $k < k_P$.

Figure 5 shows that, for the case $k < k_P$, J can be optimally used only for $\gamma \leq$

γ_{PD}.[11] Moreover, if J is optimally chosen for some interval of beliefs, at the lowest such belief the agent optimally switches to P and then to Q (without choosing D or N) as the belief goes down. This observation motivates the characterization of the transition between J and P.

Lemma 5. *There exists $k_J > 0$ such that $V'_J(\gamma) \geq V'_P(\gamma)$ for some γ if $k \leq k_J$, and $V'_J(\gamma) < V'_P(\gamma)$ for all γ otherwise.*

The proof of Lemma 5 involves showing that $V'_J(\gamma) - V'_P(\gamma)$ is single-crossing from above in k, and is provided in the appendix. We are now in the position to characterize the optimal policy in the joint information acquisition model.

(i) Suppose that $k \geq \max\{k_P, k_N\}$. In the positive information acquisition model, P is never optimal because $k \geq k_P$; the same conclusion holds here even though the agent could also engage in J or N, because the necessary conditions for P to be optimal are independent of the value function. Also, J is never optimal. This follows because, a necessary condition for $V'_J(\gamma) \geq V'_P(\gamma)$ is that $V'_P(\gamma) > 0$, which requires $\gamma > \gamma_{QP}$; while for J to be optimal, we also need $\gamma \leq \gamma_{PD}$. These two requirements are contradictory when $k \geq k_P$. Given that neither P nor J can be optimal, we know from the negative information acquisition model that $k \geq k_N$ implies that N is never optimal. Therefore,

- If $k \geq \max\{k_P, k_N\}$, then the optimal policy is: Q when $\gamma \leq \gamma_{QD}$, and D when $\gamma > \gamma_{QD}$.

(ii) Suppose that $k_P \leq k < k_N$. As stated in Proposition 3, this requires that λ/c to be large. As in the first case, P and J are never optimal. From the negative information acquisition model, we know that $k < k_N$ implies case (ii) of Proposition 2:

- If $k_P \leq k \leq k_N$, then the optimal policy is: Q when $\gamma \leq \gamma_{QD}$, D when $\gamma \in (\gamma_{QD}, \gamma_{DN}]$, N when $\gamma \in (\gamma_{DN}, \gamma_{ND}]$, and D when $\gamma > \gamma_{ND}$.

(iii) Suppose that $k_N \leq k < k_P$. This requires that λ/c to be small. As in the first case, N is never optimal: $V'_N(\gamma) - V'_D(\gamma)$ takes the same form as in the negative information acquisition model, which is non-positive because $k \geq k_N$, and thus remains non-positive as the availability of P and J can only raise the value function V. Now, recall that $k < k_P$ implies that $\gamma_{QP} < \gamma_{QD}$. We know that Q is optimal for

[11] For the case $k \geq k_P$, the linear functions in Figure 5 intersect at a point below the horizontal axis; so J is never optimal.

$\gamma \leq \gamma_{QP}$ and P is optimal for γ just above γ_{QP}. If $k \geq k_J$, then the optimal policy is given by case (ii) of Proposition 1. If $k < k_J$, Lemma 5 and the convexity of the value function imply that there is a non-degenerate interval of beliefs $(\gamma_{PJ}, \gamma_{JP})$ such that such that $V_J'(\gamma) > V_P'(\gamma)$ if only if γ belongs to this interval. But since N is never optimal and P is not optimal for $\gamma > \gamma_{PD}$, we have $\gamma_{JP} \leq \gamma_{PD}$.[12] Therefore,

- (a) If $\max\{k_N, k_J\} \leq k < k_P$, then the optimal policy is: Q when $\gamma \leq \gamma_{QP}$, P when $\gamma \in (\gamma_{QP}, \gamma_{PD}]$, and D when $\gamma > \gamma_{PD}$.
- (b) If $k_N \leq k < \min\{k_P, k_J\}$, then the optimal policy is: Q when $\gamma \leq \gamma_{QP}$, P when $\gamma \in (\gamma_{QP}, \gamma_{PJ}]$, J when $\gamma \in (\gamma_{PJ}, \gamma_{JP}]$, P when $\gamma \in (\gamma_{JP}, \gamma_{PD}]$, and D when $\gamma > \gamma_{PD}$.

(iv) Suppose that $k < \min\{k_P, k_N\}$. If $k \geq k_J$, then J is never optimal. In this case, if N is ever optimal for an interval of beliefs, then the lowest such belief must be greater than γ_{PD}. This is because if J is never optimal, then neither is N when $\gamma < \gamma_{PD}$.[13] Then, since $k < k_P$ implies that $\gamma_{QP} < \gamma_{PD}$, the agent optimally chooses Q for $\gamma \leq \gamma_{QP}$ and P for $\gamma \in (\gamma_{QP}, \gamma_{PD}]$, switching to D at γ_{PD}, as in case (ii) of Proposition 1. The rest of the optimal policy is D for $\gamma > \gamma_{PD}$ if N is never used, or otherwise given by case (ii) of Proposition 2. The agent may not choose N in spite of $k < k_N$, because the use of P for $\gamma \in (\gamma_{QP}, \gamma_{PD}]$ has raised the value function and hence reduced the usefulness of negative information.[14] For the same reason, when N is optimally used, the negative information acquisition region $(\gamma_{DN}, \gamma_{ND}]$ is different from (a subset of) that given by the negative information acquisition model. If $k < k_J$, and if further N is never optimal, then as in case (iii) above, J is optimal for some interval $(\gamma_{PJ}, \gamma_{JP}]$ to the left of γ_{PD}. Moreover, if N is optimal for an interval of beliefs, then the interval must contain γ_{PD}. This is because the linear function associated with $V_N'(\gamma) > \max\{V_P'(\gamma), V_D'(\gamma)\}$ becomes flatter as the belief γ crosses γ_{PD} from below, which by the convexity of the value function implies that if N is never optimal for any $\gamma < \gamma_{PD}$ then it cannot be optimal for any $\gamma > \gamma_{PD}$. We have the following characterization:

[12] The agent must switch from J to P, instead of to D directly. In Figure 5, the agent switches from J to P to the left of the vertical line $\gamma = \gamma_{PD}$.

[13] In Figure 5, to the left of the vertical line $\gamma = \gamma_{PD}$, the agent can switch to N only from J, not from P.

[14] Instead of $V_D(\gamma)$, as given by (3), that we have used in deriving the usefulness of negative information k_N, the new value function solves the differential equation (2) with boundary condition at γ_{PD} determined by the value function in the positive information acquisition model.

- (a) If $k_J \leq k < \min\{k_P, k_N\}$, then the optimal policy is: Q when $\gamma \leq \gamma_{QP}$, P when $\gamma \in (\gamma_{QP}, \gamma_{PD}]$, followed by either
  - D when $\gamma > \gamma_{PD}$; or
  - D when $\gamma \in (\gamma_{PD}, \gamma_{DN}]$, N when $\gamma \in (\gamma_{DN}, \gamma_{ND}]$, and D when $\gamma > \gamma_{ND}$.
- (b) If $k < \min\{k_J, k_P, k_N\}$, then the optimal policy is: Q when $\gamma \leq \gamma_{QP}$, P when $\gamma \in (\gamma_{QP}, \gamma_{PJ}]$, followed by either
  - J when $\gamma \in (\gamma_{PJ}, \gamma_{JP}]$, P when $\gamma \in (\gamma_{JP}, \gamma_{PD}]$, and D when $\gamma > \gamma_{PD}$; or
  - J when $\gamma \in (\gamma_{PJ}, \gamma_{JN}]$, N when $\gamma \in (\gamma_{JN}, \gamma_{ND}]$, and D when $\gamma > \gamma_{ND}$.

To summarize, there are several general lessons that can be drawn from this exercise. First, the idea that positive information acquisition serves as a last ditch effort before quitting the risky arm and negative information acquisition serves as an insurance strategy against fruitlessly experimenting with the risky arm remains valid in the extended model. Whenever P is chosen, it is optimal just above the quit point. Whenever N or J is chosen, they are *not* optimal just above the quit point. Moreover, whenever both P and N are chosen, N is chosen at beliefs that are more optimistic than those for P, as in case (iv) of the model. Second, joint information acquisition never completely supersedes positive information acquisition, in the sense that whenever J is chosen, P must be optimal at some belief. Further, P is chosen if and only if $k < k_P$, the same condition as described in Proposition 1. The additional options to engage in negative information acquisition or joint information acquisition do not change this condition. Third, the possibility of engaging in positive or joint information acquisition reduces the net benefit of negative information. When N is the only option other than D, choosing N is better than choosing D if

$$\lambda\pi - c/\lambda - k/\alpha - V_D(\gamma) \geq 0$$

When the options of choosing P or J are also available, they raise the value of the agent to some $V(\gamma) \geq V_D(\gamma)$, which makes the insurance strategy of N more costly. Consistent with this observation, case (iv)(b) in the joint information acquisition model shows that it is possible that N is not chosen despite $k < k_N$.

4.2. Pure information acquisition

So far we have considered information acquisition as an activity that the agent may engage in while using the risky arm. This seems a natural assumption for many applications of the bandit problem, but it may still be theoretically interesting to

analyze a model where the agent can suspend the risky arm before embarking on information acquisition as a pure information activity. Below we consider pure positive and negative information acquisition separately.

4.2.1. Pure positive information acquisition

Consider the case where the agent can suspend the risky arm while engaging in *pure* positive information acquisition. In this model, the agent can choose the risky arm only (D), the risky arm and positive information acquisition (P), pure positive information acquisition ($P\dagger$), or quit irrevocably (Q). When the agent is choosing $P\dagger$, the belief in the absence of the arrival of positive news evolves according to $\mathrm{d}\gamma = -\gamma(1-\gamma)\alpha \mathrm{d}t$. The corresponding differential equation for the value function is given by

$$\gamma(1-\gamma)\alpha V'_{P\dagger}(\gamma) = -k + \gamma\alpha(\pi - c/\lambda) - \gamma\alpha V(\gamma). \tag{12}$$

Since the agent can suspend the risky arm instead of having to quit irreversibly, the condition for $P\dagger$ to be optimal in a neighborhood of belief γ is that $V'_{P\dagger}(\gamma) \geq \max\{V'_D(\gamma), V'_P(\gamma), 0\}$.

The comparison between P and D remains the same as in Section 3.1 (i.e., $V'_P(\gamma) \geq V'_D(\gamma)$ if and only if $\gamma \leq \gamma_{PD}$). In addition, we have

$$V'_{P\dagger}(\gamma) \geq V'_D(\gamma) \iff V'_{P\dagger}(\gamma) \geq V'_P(\gamma) \iff \gamma \leq \gamma_{PD}. \tag{13}$$

Thus, if P is preferred to D, then $P\dagger$ is preferred to P, meaning that simultaneously engaging in the risky arm and positive information acquisition is never optimal.

Suppose that $k \geq k_P$. Without pure positive information acquisition, the optimal policy is Q when $\gamma \leq \gamma_{QD}$, and D when $\gamma > \gamma_{QD}$. For any $\gamma < \gamma_{QD}$, since $V(\gamma) = 0$, from (12) we have $V'_{P\dagger}(\gamma) < 0$ so $P\dagger$ is not optimal. For any $\gamma \geq \gamma_{QD}$, from (13) we have $V'_{P\dagger}(\gamma) - V'_D(\gamma) \leq 0$, and thus again $P\dagger$ is not optimal. It follows immediately that the optimal policy is the same as Proposition 1(i). This is not surprising because the benefit of positive information does not change when the agent can suspend the risky arm. As a result, the usefulness of positive information does not improve.

Next, suppose that $k < k_P$. Without pure positive information acquisition, the optimal policy is given by Proposition 1(ii), with a quitting belief at γ_{QP}. However, using $V(\gamma_{QP}) = 0$ we can easily verify from (12) that $V'_{P\dagger}(\gamma_{QP}) > 0$, a contradiction. Thus, $P\dagger$ is optimal for γ just above $\gamma_{QP\dagger}$, which is defined by the smooth-pasting

condition and the value-matching condition of $V'_{P\dagger}(\gamma_{QP\dagger}) = 0$ and $V(\gamma_{QP\dagger}) = 0$, and given by

$$\gamma_{QP\dagger} = \frac{k}{\alpha(\pi - c/\lambda)} < \gamma_{QP}.$$

Thus, the optimal policy is given by Q when $\gamma \leq \gamma_{QP\dagger}$, $P\dagger$ when $\gamma \in (\gamma_{QP\dagger}, \gamma_{PD}]$, and D when $\gamma > \gamma_{PD}$.

The information acquisition region $(\gamma_{QP}, \gamma_{PD}]$ under positive information acquisition is higher than the region $(\gamma_{QP\dagger}, \gamma_{PD}]$ under pure positive information acquisition. For belief $\gamma \in (\gamma_{QP\dagger}, \gamma_{QP})$, the ex ante probability of finding news about state $\mathcal{G}$ is too low to justify the cost if positive information acquisition has to be accompanied by the risky arm. However, pure positive information acquisition $P\dagger$ can still be justified despite the long odds because it is relatively cheap without experimentation. Pure positive information acquisition thus expands the range of beliefs at the lower end for the agent to make a last ditch effort to resurrect the risky arm when the agent can suspend it.

4.2.2. Pure negative information acquisition

In pure negative information acquisition (choosing $N\dagger$), the agent incurs a flow cost at the rate of k and receives conclusive negative news at the rate of α, while suspending the risky arm. For simplicity, we again assume that $\alpha < \lambda$.[15] The updated belief upon no news after choosing $N\dagger$ evolves according to $\mathrm{d}\gamma = \gamma(1-\gamma)\alpha \mathrm{d}t$. The belief goes up because the absence of the arrival of negative news makes the agent more optimistic about the state. This feature implies that the analysis of the model of pure negative information acquisition below parallels the fast negative information acquisition model of Section 3.2.2 regardless of the value of α.

The differential equation for the value function when the agent optimally chooses $N\dagger$ is given by

$$\gamma(1-\gamma)\alpha V'_{N\dagger}(\gamma) = k + (1-\gamma)\alpha V(\gamma). \tag{14}$$

Following the analysis in Section 3.2.2, at the highest belief for which $N\dagger$ is optimally chosen, the belief stays stationary until either a success from the risky arm or the arrival of negative news. The payoff associated with such a stationary belief $\gamma^\dagger$ is

[15] The analysis for the case of $\alpha \geq \lambda$ proceeds in the same way as below. The only difference is that, to keep the belief stationary, instead of mixing between $N\dagger$ at full intensity and D at intensity α/λ, the agent mixes between $N\dagger$ at intensity λ/α and D at full intensity.

$V^*(\gamma^\dagger)$. As in Section 3.2.2, the value of this belief is determined by $V(\gamma^\dagger) = V^*(\gamma^\dagger)$ and $V_D'(\gamma^\dagger) = \pi$, the same conditions as those in the fast negative information acquisition model. Therefore, $\gamma^\dagger$ is equal to γ^* given in Section 3.2.2.

Solve the differential equation (14) with boundary condition at γ^* given by $V^*(\gamma^*)$. Let $V_{N\dagger}(\gamma)$ be the solution, and define $k_D^\dagger \in (0, k_N)$ such that $V_{N\dagger}(\gamma_{QD}) = 0$. Now, we can provide a characterization of the optimal policy in the pure negative information acquisition model.

Suppose that $k \geq k_N$. In the negative information acquisition model, N is never optimal. Since the highest belief $\gamma^\dagger$ for which $N\dagger$ is optimal is the same as γ^* in the fast negative information acquisition case, the proof of Proposition 5(i) applies to imply that $N\dagger$ is not optimal for any belief either. Being able to suspend the risky arm while engaging in negative information acquisition does not increase its usefulness. The optimal policy stays the same as when the agent can only choose between D and Q.

Suppose that $k < k_N$. With pure negative information acquisition a feasible choice, the agent optimally chooses $N\dagger$ for γ just below γ^*. If $k \in [k_D^\dagger, k_N)$, the lowest value of γ for which $N\dagger$ is optimally chosen is $\gamma_{DN\dagger}$, given by

$$V_{N\dagger}(\gamma_{DN\dagger}) = V_D(\gamma_{DN\dagger}).$$

The optimal policy is: Q for $\gamma \leq \gamma_{QD}$, D for $\gamma \in (\gamma_{QD}, \gamma_{DN\dagger}]$, $N\dagger$ for $\gamma \in (\gamma_{DN\dagger}, \gamma^*]$, and D for $\gamma \geq \gamma^*$. If instead $k < k_D^\dagger$, the lowest value of γ for which n is optimally chosen is $\gamma_{QN\dagger}$, given by

$$V_{N\dagger}(\gamma_{QN\dagger}) = 0.$$

The optimal policy is: Q for $\gamma \leq \gamma_{QN\dagger}$, $N\dagger$ for $\gamma \in (\gamma_{QN\dagger}, \gamma^*]$, and D for $\gamma > \gamma^*$.

For the same α, the region for which $N\dagger$ is optimal is higher than the negative information acquisition region. Comparing the upper bounds of the two regions, the proof of Proposition 4 shows that $\gamma_{ND} > \gamma^*$. For the lower bound, γ_{DN} in the negative information acquisition model is determined by the intersection between $V_D(\gamma)$ and $V^*(\gamma)$, while $\gamma_{DN\dagger}$ in the pure negative model (with $k \geq k_D^\dagger$) is determined by the intersection between $V_D(\gamma)$ and $V_{N\dagger}(\gamma)$. Since $V_{N\dagger}(\gamma) > V^*(\gamma)$ for $\gamma < \gamma^*$, we have $\gamma_{DN} > \gamma_{DN\dagger}$. And if $k < k_D^\dagger$, we have $\gamma_{DN} > \gamma_{QD} > \gamma_{QN\dagger}$. Therefore, $N\dagger$ is optimal at a more pessimistic range of beliefs than N is.

Our characterization of the optimal strategy also implies that N is never optimal

when $N\dagger$ is available. To see why this is true, recall that, under the slow negative information acquisition model of Section 3.2.1, the value function satisfies $V(\gamma) < V^*(\gamma)$ for $\gamma \in (\gamma_{DN}, \gamma_{ND})$. Instead of choosing N in this region, the agent can mix between $N\dagger$ with intensity 1 and D with intensity α/λ so that the belief becomes stationary. This alternative strategy to arrest the downward updating of belief will yield a payoff of $V^*(\gamma)$, which dominates the payoff in the slow negative information acquisition model. To be sure, this deviation is not optimal. For $\gamma \in (\gamma^*, \gamma_{ND}]$, it is better to choose D than to mix, because the agent optimally postpones negative information acquisition until the belief reaches γ^*, knowing that it can keep the belief there indefinitely (until either the risky arm produces a success or negative news arrives). For $\gamma \in (\gamma_{DN\dagger}, \gamma^*)$, it is better to choose $N\dagger$ than to mix, because the agent optimally postpones the risky arm until the belief is pushed back to a higher level γ^* that makes it more promising.

5. Discussion

There are potentially two broad directions where the model can be further extended. One is to generalize the idea of information acquisition in an experimentation model beyond positive and negative structures, maintaining the key feature that information acquisition has no direct payoffs. In the present paper, only two states are possible, one in which the risky arm is good and is expected to deliver a success, and the other in which the arm is bad and no success is ever possible. As a result, there is limited scope to address the issue of which type of information, and when, the agent should try to acquire. In a general environment with more than two possible states, we may model the dynamic trade-off between a narrow but in-depth search for information about the state versus a broad but cursory one. For example, in the R&D application of the bandit problem, a firm might be engaged in a product development process, where the potential new product can have several features that may be represented by a multi-dimensional state space. How to model the state space and build different types of information structures remains a challenge.

The second direction is to extend the single-agent model to a game between rival agents competing to be the first to achieve success from a risky arm. For example, imagine firms in the same market competing to develop a new product. There may be some uncertainty in the true state of the world, which could either be such that firms face identical and independent prospects of successfully developing a new

product, or no success is possible by any firm because there is no demand for it or because the technology is not feasible. Information acquisition can be introduced in such a competitive framework, by assuming that agents have an independent, reversible and costly option of uncovering conclusive news about the state. The challenge is that if agents' information acquisition activities are not observable and news from information acquisition is not shared, even if they start with the same information about the state, private information will emerge among them. Recent papers that have introduced private information to bandit problems, including Moscarini and Squintani (2010), Farrell and Simcoe (2009), and Guo and Roesler (2016), may suggest a way to analyze the strategic interactions that come with learning while experimenting.

Appendix

Proof of Lemma 1. Suppose it is optimal to choose P at intensity $\delta \in (0,1)$ when the belief is in a neighborhood of some γ. Then, it must not be better to switch to D for small time interval, implying:

$$\begin{aligned} V(\gamma) &= -(c+\delta k)\mathrm{d}t + \gamma(\lambda\pi\mathrm{d}t + \delta\alpha(\pi - c/\lambda)\mathrm{d}t) \\ &\quad + (1-\gamma\lambda\mathrm{d}t - \gamma\delta\alpha\mathrm{d}t)(V(\gamma) - \gamma(1-\gamma)(\lambda+\delta\alpha)V'_{P,\delta}(\gamma)\mathrm{d}t) \\ &\geq -c\mathrm{d}t + \gamma\lambda\pi\mathrm{d}t + (1-\gamma\lambda\mathrm{d}t)(V(\gamma) - \gamma(1-\gamma)\lambda V'_{P,\delta}(\gamma)\mathrm{d}t), \end{aligned}$$

where the inequality is strict if the agent strictly prefers P at intensity δ to D. From the differential equations in the D region and in the P region, we can show that the agent cannot be indifferent between D and P at any level σ in an interval of beliefs: such indifference implies that $\gamma = 1 - (k/\alpha)/(c/\lambda)$, which is a contradiction. Since $\delta > 0$, the above inequality implies

$$-k + \gamma\alpha(\pi - c/\lambda) - \gamma\alpha V(\gamma) - \gamma(1-\gamma)\alpha V'_{P,\delta}(\gamma) > 0.$$

Now, the payoff from deviating to P at a higher level $\hat{\delta} > \delta$ for an interval of time $\mathrm{d}t$ and then reverting to using the optimal strategy δ is

$$\begin{aligned} \hat{V}(\gamma) &= -(c+\hat{\delta}k)\mathrm{d}t + \gamma(\lambda + \hat{\delta}\alpha(\pi - c/\lambda))\mathrm{d}t \\ &\quad + (1-\gamma(\lambda+\hat{\delta}\alpha)\mathrm{d}t)(V(\gamma) - \gamma(1-\gamma)(\lambda+\hat{\delta}\alpha)V'_{P,\delta}(\gamma)\mathrm{d}t) \end{aligned}$$

Ignoring terms involving $(\mathrm{d}t)^2$, $\hat{V}(\gamma) - V(\gamma)$ is equal to

$$(\hat{\delta} - \delta)\left(-k + \gamma\alpha(\pi - c/\lambda) - \gamma\alpha V(\gamma) - \gamma(1-\gamma)\alpha V'_{P,\delta}(\gamma)\right)\mathrm{d}t,$$

which is strictly positive by the previous inequality, a contradiction to optimality. ∎

Proof of Lemma 3. Suppose it is optimal to choose N at some intensity $\delta \in (0,1)$ and suppose first that $\delta \neq \delta^*$. The payoff from deviating to chooses N at a higher level $\hat{\delta} > \delta$ for an interval of time $\mathrm{d}t$ and then reverting to using the optimal strategy δ is

$$\begin{aligned} \hat{V}(\gamma) &= -(c+\hat{\delta}k)\mathrm{d}t + \gamma\lambda\pi\mathrm{d}t \\ &\quad + (1-(\gamma\lambda + (1-\gamma)\hat{\delta}\alpha)\mathrm{d}t)(V(\gamma) - \gamma(1-\gamma)(\lambda - \hat{\delta}\alpha)V'_{N,\delta}(\gamma)\mathrm{d}t). \end{aligned}$$

Therefore $\hat{V}(\gamma) - V(\gamma)$ is equal to (ignoring all $o(\mathrm{d}t)$ terms):

$$(\hat{\delta} - \delta)\left(-k - (1-\gamma)\alpha V(\gamma) + \gamma(1-\gamma)\alpha V'_{N,\sigma}(\gamma)\right)\mathrm{d}t.$$

Since it is optimal to choose N at level $\delta > 0$, we have

$$\begin{aligned} V(\gamma) &= -(c+\delta k)\mathrm{d}t + \gamma\lambda\pi\mathrm{d}t \\ &\quad + (1-\gamma\lambda\mathrm{d}t - (1-\gamma)\delta\alpha\mathrm{d}t)(V(\gamma) - \gamma(1-\gamma)(\lambda-\delta\alpha)V'_{N,\delta}(\gamma)\mathrm{d}t) \\ &\geq -c\mathrm{d}t + \gamma\lambda\pi\mathrm{d}t + (1-\gamma\lambda\mathrm{d}t)(V(\gamma) - \gamma(1-\gamma)\lambda V'_{N,\delta}(\gamma)\mathrm{d}t), \end{aligned}$$

where the inequality is strict if the agent strictly prefers N at intensity δ to D. From the differential equations in the D region and in the N region, we can show that the agent cannot be indifferent between D and N at a fixed intensity in an interval of beliefs: such indifference implies that $V(\gamma) = \gamma - c/\lambda - k/\alpha$, and hence $\gamma = c\alpha/(c\alpha + k\lambda)$, which is a contradiction. Now, since $\delta > 0$, the above inequality implies $\hat{V}(\gamma) - V(\gamma) > 0$, which contradicts the optimality of N at level δ.

Next, suppose that $\delta = \delta^*$. Then,

$$V(\gamma) = V^*(\gamma) = \gamma\pi - c/\lambda - k/\alpha,$$

with $V'(\gamma) = \pi$. Therefore,

$$\hat{V}(\gamma) - V(\gamma) = \alpha(\hat{\delta} - \delta^*)\left[(1-\gamma)\frac{c}{\lambda} - \gamma\frac{k}{\alpha}\right]\mathrm{d}t.$$

If the bracketed term on the right-hand-side is positive, payoff can be increased by raising $\hat{\delta}$ above δ^*. If the term is negative, payoff can be raised by lowering $\hat{\delta}$ below δ^*. ∎

Proof of Lemma 5. Let $V_P(\gamma)$ solve the differential equation (5) with the boundary condition $V_P(\gamma_{QP}) = 0$. For $\gamma \geq \gamma_{QP}$, the explicit solution is given by

$$V_P(\gamma) = \frac{c+k}{\lambda+\alpha}\left(\frac{\gamma - \gamma_{QP}}{\gamma_{QP}} - (1-\gamma)\left(\log\frac{\gamma}{1-\gamma} - \log\frac{\gamma_{QP}}{1-\gamma_{QP}}\right)\right).$$

Define

$$\begin{aligned} f(k) &\equiv \max_{\gamma}\left\{-\frac{c+k}{\lambda+\alpha} - \frac{k}{\alpha} + \gamma\left(\pi - \frac{\alpha}{(\lambda+\alpha)}\frac{c}{\lambda}\right) - V_P(\gamma)\right\} \\ &= \max_{\gamma}\left\{-\frac{k}{\alpha} + \frac{c+k}{\lambda+\alpha}(1-\gamma)\left(\log\frac{\gamma}{1-\gamma} - \log\frac{\gamma_{QP}}{1-\gamma_{QP}}\right)\right\}. \end{aligned}$$

Note that $f(0)$ is positive, and $f(k)$ is negative for k sufficiently large. Thus, there exists $k_J > 0$ such that $f(k_J) = 0$. To prove that k_J is unique, we show that $f(k)$ is single-crossing from above. By the envelope theorem,

$$f'(k) = -\frac{1}{\alpha} + \frac{1}{\lambda+\alpha}(1-\gamma_J)\left(\log\frac{\gamma_J}{1-\gamma_J} - \log\frac{\gamma_{QP}}{1-\gamma_{QP}}\right) - \frac{c+k}{\lambda+\alpha}\frac{1-\gamma_J}{\gamma_{QP}(1-\gamma_{QP})}\frac{\partial\gamma_{QP}}{\partial k},$$

where γ_J is the solution to the maximization problem. At $k = k_J$, we have $f(k_J) = 0$, so the derivative reduces to

$$f'(k_J) = -\frac{c}{\alpha(c+k_J)} - \frac{c+k_J}{\lambda+\alpha}\frac{1-\gamma_J}{\gamma_{QP}(1-\gamma_{QP})}\frac{\partial\gamma_{QP}}{\partial k} < 0.$$

Thus, $f(k)$ crosses zero once and from above, which implies that $f(k) \geq 0$ if $k \leq k_J$ and $f(k) \leq 0$ if $k \geq k_J$. When $f(k) \leq 0$, there is no γ such that $V_J'(\gamma) > V_P'(\gamma)$. When $f(k) > 0$, we have $V_J'(\gamma_J) > V_P'(\gamma_J)$. Since the objective function associated with the maximization problem is concave, there is an interval of belief $V_J'(\gamma) \geq V_P'(\gamma)$ for all γ in that interval. ■

References

Bergemann, D., and Valimaki, J. "Learning and strategic pricing," *Econometrica* 64 (1996): 1125–1150.

Bergemann, D., and Valimaki, J. "Experimentation in markets," *Review of Economic Studies* 67 (2000): 213–234.

Bolton, P., and Harris, C. "Strategic experimentation," *Econometrica* 67 (1999): 349–374.

Camargo, B. "Good news and bad news in two-armed bandits," *Journal of Economic Theory* 133 (2007): 558–566.

Che, Y. K., and Mierendorff, K. "Optimal sequential decision with limited attention," Columbia University working paper, 2016.

Choi, J. P. "Dynamic R&D competition under 'hazard rate' uncertainty," *RAND Journal of Economics* 22 (1991): 596–610.

Farrell, J., and Simcoe, T. "Choosing the rules for formal standardization," Boston University working paper, 2009.

Gittins, J. C. "Bandit processes and dynamic allocation indices," *Journal of the Royal Statistical Society* 41 (1979): 148–177.

Guo, Y., and Roesler, A. K. "Private learning and exit decisions in collaboration," Northwestern University working paper, 2016.

Harris, C., and Vickers, J. "Perfect equilibrium in a model of a race," *Review of Economic Studies* 52 (1985): 193–209.

Harris, C., and Vickers, J. "Racing with uncertainty," *Review of Economic Studies* 54 (1987): 1–21.

Moscarini, G., and Squintani, F. "Competitive experimentation withe private information: the survivor's curse," *Journal of Economic Theory* 145 (2010): 639–660.

Klein, N., and Rady, S. "Negatively correlated bandits," *Review of Economic Studies* 78 (2011): 693–732.

Keller, G., Cripps, M., and Rady, S. "Strategic experimentation with exponential bandits," *Econometrica* 73 (2005): 39–68.

Malueg, D., and Tsutsui, S. "Dynamic competition with learning," *RAND Journal of Economics* 28 (1997): 751–772.

Reinganum, J. "Dynamic games of innovation," *Journal of Economic Theory* 25 (1981): 21–41.

Reinganum, J. "A dynamic game of R&D: Patent protection and competitive behavior," *Econometrica* 50 (1982): 671–688.

Roberts, K., and Weitzman, M. "Funding criteria for research, development and exploration of projects," *Econometrica* 49 (1981): 1261–1288.

Weitzman, M. "Optimal search for the best alternative," *Econometrica* 47 (1979): 641–654.