Keeping Your Options Open*

Jean Guillaume Forand†

February 27, 2013

Abstract

In standard models of experimentation, the costs of project development consist of (*a*) the direct cost of running trials as well as (*b*) the implicit opportunity cost of leaving alternative projects idle. Another natural type of experimentation cost, the cost of holding on to the option of developing a currently inactive project, has not been studied. In a multi-armed bandit model of experimentation in which inactive projects have explicit maintenance costs and can be irreversibly discarded, I fully characterise optimal experimentation policies and show that the decision-maker's incentive to actively manage its options has important implications for the order of project development. In the model, an experimenter searches for a success among a number of projects by choosing both those to develop now and those to maintain for (potential) future development. In the absence of maintenance costs, optimal experimentation policies have a 'stay-with-the-winner' property: the projects that are more likely to succeed are developed first. Maintenance costs provide incentives to bring the option value of less promising projects forward, and under optimal experimentation policies, 'going-with-the-loser' can be optimal: projects that are less likely to succeed are sometimes developed first.

JEL Classification: D83, D81, C61
Keywords: Experimentation, Maintenance Costs, Multi-Armed Bandits

1 Introduction

When experimentation is costly, decision-makers must choose which alternatives to actively investigate and which to leave 'on the back burner'. A firm engaged in research and development can obtain comparable product innovations through various technologies. Investing in multiple technologies simultaneously is costly, so how should the firm prioritises its allocation of funds to competing alternatives? Academic researchers typically have many ongoing projects of varying

*I would like to thank Li Hao, Martin Osborne and Colin Stewart for their supervision, comments and suggestions. I would also like to thank Ettore Damiano, John Duggan, Francisco Gonzalez, Carolyn Pitchik, Wing Suen, Thomas Wiseman, seminar participants at Guelph, Rochester and Waterloo and participants at the CETC 2011, Fall 2010 METC, NASM 2011 SEDesign 2011 and SEDynamics 2011.

†Department of Economics, University of Waterloo, Hagey Hall of Humanities, Waterloo, Ontario, Canada N2L 3G1. Email: jgforand@uwaterloo.ca. Website: http://arts.uwaterloo.ca/∼jgforand

quality, with limited time to devote to research duties. Which projects should receive more of the researcher's attention? In standard models of experimentation, putting one alternative to trial over another entails only an implicit opportunity cost: the foregone opportunity of learning about the inactive alternative. However, retaining the option to investigate a currently shelved alternative often involves explicit maintenance costs. Firms engaged in research and development routinely devote resources solely to keep open the option of developing low-priority technologies, which involves the costly upkeep of specialised equipment and paying the salaries of skilled workers or scientists that can be lost to other firms. For these firms, holding on to the rights to proprietary technology with outside market value also imposes a cost in foregone revenue. Academic researchers can bear psychological costs from lingering projects, and efforts to limit these costs may impact their allocation of time to ongoing projects.

In this paper, I present a multi-armed bandit model of experimentation in which arms (projects) that are not allocated experimentation effort impose explicit maintenance costs unless the experimenter chooses to irreversibly discard them.[1] Specifically, I adopt the continuous-time exponential bandit framework of Keller et al. (2005) and consider a model in which two independent risky projects can be of either good or bad type and only good projects eventually succeed if developed. The experimenter's belief about a project, its updated probability that the project is of the good type, decays as the project is put to trial and no success is observed. The experimenter selects discarding times for the projects and allocates trials among maintained projects to maximise its discounted expected payoffs: the benefits from obtaining a success on either project less total experimentation costs, which include the direct trial costs of active projects and the maintenance costs of inactive projects. Successes are perfect substitutes and a success on either project ends experimentation. As opposed to standard problems, at any time at which some project is maintained but inactive, the experimenter faces a choice between paying to keep that option open or discarding it (irreversibly) altogether. Discarding an inactive project liquidates its option value, which is realised in the event that currently active project is deemed unpromising. To avoid destroying this option value or paying to maintain it, the experimenter has an incentive to bring it forward by altering the order of project development.

I fully characterise optimal experimentation policies with maintenance costs and show that they entail significant departures from standard results. In the corresponding environment without maintenance costs, optimal experimentation policies are characterised by the well-known Gittins index policy.[2] In the exponential framework, this policy has the 'stay-with-the-winner' property: at any time, the project with the highest updated probability of being of the good type (the winner) is put to trial. In the presence of maintenance costs, I show that 'going-with-

[1] Bergemann and Välimäki (2008) survey the literature on multi-armed bandits with a focus on applications in economics.

[2] To each arm is assigned an index number that depends only on the ex ante characteristics and accumulated observations of that project. The optimal experimentation policy consists of always selecting a project among those with maximal indices.

the-loser' is a robust feature of optimal experimentation: the project with the lowest updated probability of being of good type (the loser) can be put to trial first. Specifically, either (a) maintenance costs are so high that the losing project is always discarded immediately for all initial beliefs about its quality, otherwise, (b), under optimal experimentation the loser will be put to trial before the winning project for a non-negligible set of initial beliefs.

Moreover, bringing the option value of a losing project forward necessarily takes the form of a simple and appealing culling rule. This project is granted a 'last chance' to succeed in a period of experimentation in which it is put to trial exclusively while the winner is maintained, after which it is permanently discarded. In other words, if the losing project is ever put to trial, the experimenter returns to the winning project only after the losing project has been discarded. This culling rule for losing projects is key for characterising optimal experimentation with maintenance costs and it takes a more general form in the case in which the experimenter has more than two risky projects. When the experimenter has three risky projects ranked by their likelihoods of success,[3] I show that if it is ever optimal to experiment with the middle-ranked project, then experimentation can proceed to the top-ranked project only when both the middle-ranked and the lowest-ranked projects have been discarded. In other words, experimenting with a middle-ranked project grants a 'last chance' to all projects of a lower or equal rank.

An outside observer that fails to take maintenance costs for inactive projects into account would conclude that the experimenter is sometimes prioritising the wrong project. Furthermore, during a culling phase the experimenter appears to not 'know when to pull the plug' and to cling to projects that have repeatedly failed to achieve results.[4] However, when inactive projects' option values are attached to a stream of maintenance costs, the experimenter 'throws good money after bad' precisely in order to convince itself that the initial investments were indeed a bad idea, thus ensuring a quicker extrication of resources from a hopeless project towards more promising ones.

1.1 Literature

The Gittins index representation of optimal policies in discounted bandit problems with independent arms is not robust to perturbations of the standard model such as correlated arms, non-geometric discounting and the simultaneous pulling of multiple arms. Much less is known about optimal experimentation policies in models in which they do not take the simple index form. Closer to this paper, Banks and Sundaram (1994) have shown that index policies are not optimal in the presence of switching costs between arms.[5] Switching costs are attributed to

[3]The argument is easily extended to more than three projects.

[4]This is reminiscent of the literature on escalation and the sunk cost fallacy. See Staw (1981), Staw and Ross (1987) and Garland (1990). McAfee et al. (2010) have expressed misgivings about interpreting such phenomena solely from a non-rational perspective.

[5]General characterisations of optimal experimentation policies with switching costs are not known. For details, see Jun (2004). An exception is Bergemann and Välimäki (2001), who exploit results of Banks and Sundaram

an inactive arm only when experimentation transitions to it and are always accompanied by an observation from that arm. Maintenance costs, on the other hand, need to be paid whenever an inactive arm is not pulled and never generate observations from that arm.[6] Nevertheless, optimal experimentation policies in bandit problems with maintenance costs fail to admit a Gittins index representation for the reason found by Banks and Sundaram (1994): the index of a given maintained arm would have to be a function of the maintenance cost, and this relationship would depend nontrivially on the characteristics of outside arms, breaking the independence required for an index characterisation.[7]

The exponential bandit framework due to Keller et al. (2005) has proved useful in applications due to its tractability. Keller et al. (2005), following Bolton and Harris (1999), study strategic experimentation and the free-riding incentives of multiple agents facing a single risky arm. Keller and Rady (2010) generalise the model to 'poisson' bandits that allow for arms of the bad type to also generate successes. Klein and Rady (2011) allow for each of two experimenters to have perfectly negatively correlated versions of the same risky arm. Strulovici (2010) applies the model in a voting framework. Bergemann and Hege (1998) introduce a discrete-time version of the model to study the moral hazard problem arising between bankers (principals) and venture capitalists (experimenters). In this vein, Bonatti and Hörner (2011), Hörner and Samuelson (2009) and Klein (2011) focus on the provision of incentives to experimenting agents.

Bonatti and Hörner (2011) derive another version of the 'stay-with-the-winner' rule when agents can experiment with multiple disjunctive projects, i.e., when, as in my model, project successes are perfect substitutes. They also uncover a 'go-with-the-loser' rule when projects are conjunctive, i.e., when successes on both projects are perfect complements. In that case, experimenting first with the losing project is optimal since a success on the winning project is worthless on its own. My results show that with maintenance costs to inactive projects, 'going-with-the-loser' is optimal even with disjunctive projects. Also, in Section 5.2, I show that 'going-with-the-loser' is still optimal if successes on various projects can be accumulated. Hence, my results are due to the incentive to economise on maintenance costs by bringing the option values of inactive projects forward, and not to the perfect substitutability of project successes.

1.2 Example

The following simple example illustrates the main lessons of the paper by clarifying why experimenting with the losing project first can be optimal when maintaining inactive projects is costly. An experimenter can devote a trial to one of two projects, A and B, in each of three periods, and can irreversibly discard any project at any time. Projects are risky in that the payoffs they

(1992b) on bandits with a countable numbers of ex ante identical arms.

[6] See the Conclusion for more discussion of the relationship between switching costs and maintenance costs in experimentation.

[7] Furthermore, it is not clear how to define an index policy in the presence of maintenance costs since experimentation policies need to specify both which arm is pulled and which arms are maintained.

deliver are unknown. A project of type *Good* delivers a one-time payoff of 1, termed a *success*, with probability $G > 0$ in any trial. A project of type *Bad* never delivers a success. Assume that experimentation ends once a single trial is successful. Direct experimentation costs are $\overline{k} > 0$ per trial, maintenance costs for a maintained but inactive project are $\underline{k} \geq 0$ per period and there is no discounting.

A project's current state is characterised by the experimenter's belief that it is of type *Good* and repeated failures make the experimenter more pessimistic about the project. Let p_J^i be the probability that project J is of type *Good* given that it has failed i trials, with $i \in \{0, 1, 2\}$. By Bayes' rule, $p_J^i = \frac{p^{i-1}(1-G)}{1-p^{i-1}G}$ for $i = \{1, 2\}$, and initial beliefs (p_A^0, p_B^0) are given. Assume that $p_B^0 \in [p_A^2, p_A^1]$. This ensures that project A is the better project ex ante and that in the absence of maintenance costs, that is if $\underline{k} = 0$, the optimal experimentation sequence is the 'go-with-the-winner' sequence AAB which allocates the first two trials to project A and the final trial to project B.

When $\underline{k} > 0$, the optimal experimentation sequence must be one of $AA|_AB$, $|_BAAA$ or $B|_BAA$, where $|_J$ represents the discarding of project J. In other words, either the experimenter sticks with the 'go-with-the-winner' rule, discards the losing project B prior to the first trial or it gives project B an early chance to succeed and discards it following a single failure. This follows since (a) if a project is discarded prior to the first trial, it must be project B, (b) if a project is discarded after a single trial, it must be project B irrespective of the project first put to trial, and among such sequences $B|_BAA$ yields the highest probability of a success, while (c) any experimentation sequence that maintains both projects in the first two trials must discard a project before the final trial, and among such sequences $AA|_AB$ yields the highest probability of a success. Let $V(s; p_A^0, p_B^0)$ be the expected payoff to experimentation sequence s given initial beliefs (p_A^0, p_B^0). Then

$$\begin{aligned}
V(AA|_AB; p_A^0, p_B^0) &= p_A^0 + (1-p_A^0)p_A^1 + (1-p_A^0)(1-p_A^1)p_B^0 \\
&\quad - \Big[(\overline{k}+\underline{k}) + (1-p_A^0)(\overline{k}+\underline{k}) + (1-p_A^0)(1-p_A^1)\overline{k}\Big], \\
V(B|_BAA; p_A^0, p_B^0) &= p_B^0 + (1-p_B^0)p_A^0 + (1-p_B^0)(1-p_A^0)p_A^1 \\
&\quad - \Big[(\overline{k}+\underline{k}) + (1-p_B^0)\overline{k} + (1-p_B^0)(1-p_A^0)\overline{k}\Big], \\
V(|_BAAA; p_A^0, p_B^0) &= p_A^0 + (1-p_A^0)p_A^1 + (1-p_A^0)(1-p_A^1)p_A^2 \\
&\quad - \Big[\overline{k} + (1-p_A^0)\overline{k} + (1-p_A^0)(1-p_A^1)\overline{k}\Big].
\end{aligned}$$

The incentive to bring the option value of project B forward entails a trade-off. The benefits of the 'go-with-the-loser' experimentation sequence $B|_BAA$ are that (a) (relative to sequence $|_BAAA$) the value of project B gets exploited and the decision to discard B is better informed while (b) (relative to sequence $AA|_AB$) saving on maintenance costs. However, to the experimentation sequence $B|_BAA$ are associated both (a) the maintenance cost (relative to $|_BAAA$)

and (b) the opportunity cost (relative to $AA|_AB$) of leaving the 'better' project A idle while experimenting with project B.

Simple calculations show that $V(AA|_AB; p_A^0, p_B^0) - V(B|_BAA; p_A^0, p_B^0)$ is decreasing in p_B^0. Hence, if $B|_BAA$ is preferred to $AA|_AB$ for some $p_B^0 \in [p_A^2, p_A^1]$, then this is also the case for all $p_B'^0 > p_B^0$. Note also that $V(|_BAAA; p_A^0, p_B^0)$ is independent of p_B^0 and that $V(|_BAAA; p_A^0, p_B^0) - \max\{V(AA|_AB; p_A^0, p_B^0), V(B|_BAA; p_A^0, p_B^0)\}$ is decreasing in p_B^0 and is strictly positive at $p_B^0 = p_A^2$. That is, when $p_B^0 = p_A^2$, all three experimentation sequences $AA|_AB$, $B|_BAA$ and $|_BAAA$ yield the same success probabilities, yet $|_BAAA$ has strictly lower costs. Hence, for fixed G, $\overline{k}$, $\underline{k}$ and p_A^0, the optimal experimentation policy can be represented by beliefs $\underline{p}$, $\overline{p}$ with $p_A^2 \leq \underline{p} \leq \overline{p} \leq p_A^1$, such that $|_BAAA$ is optimal on $[p_A^2, \underline{p}]$, $AA|_AB$ is optimal on $[\underline{p}, \overline{p}]$ and $B|_BAA$ is optimal on $[\overline{p}, p_A^1]$. In general, all three intervals can be non-empty. An example has $G = \frac{2}{5}$, $\overline{k} = \frac{6}{100}$, $\underline{k} = \frac{3}{200}$ and $p_A^0 = \frac{45}{100}$. Then it can be computed that $p_A^1 \approx 0.33$ and $p_A^2 \approx .23$, while $\underline{p} \approx 0.29$ and $\overline{p} \approx 0.32$. This is depicted in Figure 1.

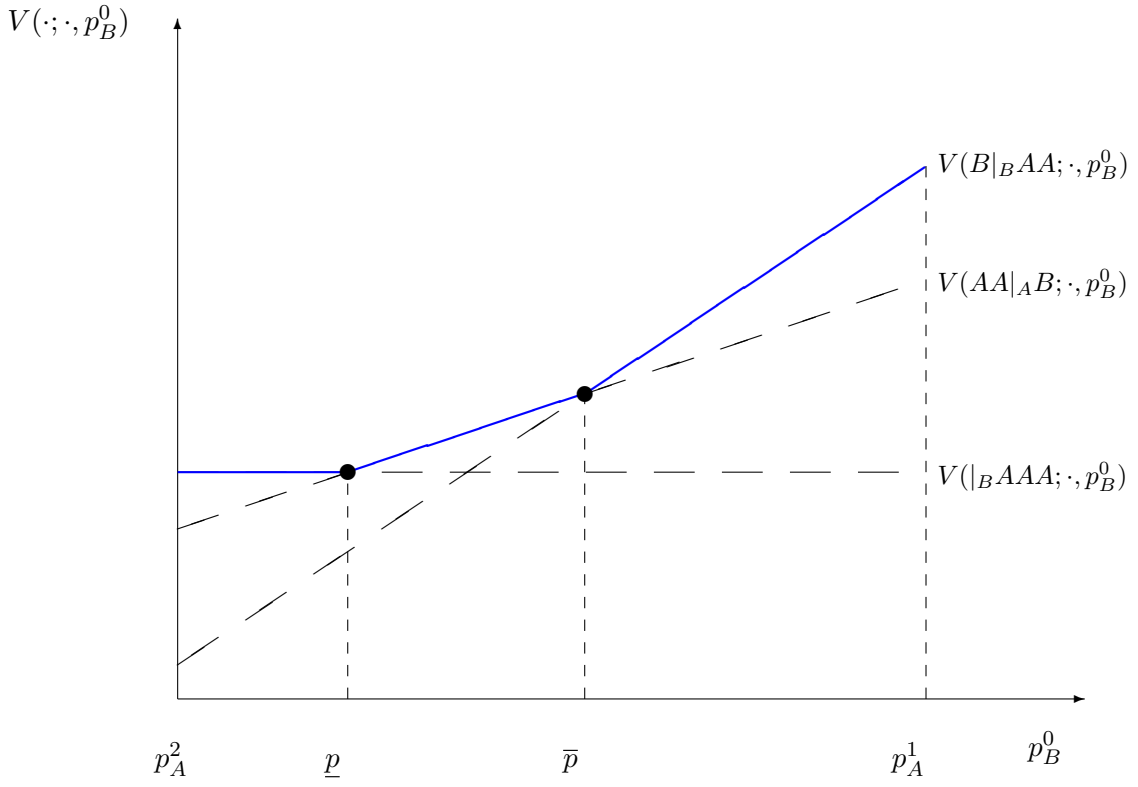


Figure 1: Example: optimal experimentation as a function of p_B^0.

2 Model

I consider a continuous time three-armed exponential bandit problem with two risky arms, A and B, and a safe arm Q.[8] Risky arms will henceforth be referred to as projects, and the safe arm as the option to quit experimentation. A trial consists of experimenting with a risky project for some time interval $[t, t+dt)$. Trials yield either successes or failures. The type of a risky project is $\theta \in \{Good, Bad\}$. A risky project of type θ that is put to trial continuously in time interval $[t, t+dt)$ succeeds with probability Gdt for some $G > 0$ if $\theta = Good$, while it fails with probability 1 if $\theta = Bad$.[9] The types of risky projects A and B are drawn independently, where (p_A^0, p_B^0) are the initial beliefs of the experimenter about the types of arms A and B, with p_J^0 the probability that project $J \in \{A, B\}$ is of type *Good*.

A risk-neutral experimenter collects the benefits of successes on risky projects and bears all experimentation costs. Its objective is to maximise its discounted expected flow payoff. A success on either risky project yields a lump-sum payment of 1 and ends the experimentation process.[10] Quitting experimentation, which I assume is irreversible, yields a flow payoff of 0.[11] Experimenting continuously with risky project J in time interval $[t, t+dt)$ entails direct experimentation cost $\overline{k}dt$. I introduce explicit costs to maintaining inactive risky projects. That is, a risky project that is maintained but not involved in a trial in time interval $[t, t+dt)$ imposes cost $\underline{k}dt$. The experimenter can irreversibly discard risky projects without cost and quitting the experimentation process also entails no costs. The experimenter discounts future payoffs at rate r.

Since experimentation ends after the first success, the only histories after which the experimenter selects a project to put to trial are intervals of time in which only failures have been observed. A *strategy* is a collection (α, t_A, t_B), where $\alpha : \mathbf{R}_+ \to [0,1] \cup \{Q\}$ is measurable and $t_A, t_B \geq 0$. The function α is an *assignment rule* and $\int_{t'}^{t'+dt} \alpha(t)dt$ specifies the fraction of time devoted to experimenting with project A in time interval $[t', t'+dt)$ if the experimenter conducts trials in that interval, while $\alpha(t) = Q$ for all $t \in [t', t'+dt)$ if the experimenter has quit experimentation at some time $t'' \leq t'$. The experimenter is allowed to share quantity dt of experimentation time between the risky projects in an interval of time $[t, t+dt)$ in which both are maintained, but note that this shared allocation leads to experimentation cost $(\overline{k} + \underline{k})dt$. The stopping times t_A and t_B specify *maintenance rules*, with project J maintained at any time $t < t_J$. Note that, by assumption, $\alpha(t) = Q$ if and only if $t \geq \max\{t_A, t_B\}$. Given initial beliefs (p_A^0, p_B^0), a strategy (α, t_A, t_B) induces uniquely defined and continuously differentiable laws of motion for beliefs

[8] See Section 5.1 for an extension to more than two risky arms.

[9] See the Conclusion for a discussion of alternative distributional assumptions for risky project types.

[10] See Section 5.2 for an extension to the case in which successes are accumulated across projects.

[11] This assumption considerably simplifies notation, while the corresponding restriction on experimentation strategies is without loss of generality for optimal experimentation.

$(p_A^t, p_B^t)_t$. These laws of motion, given by

$$\dot{p}_A^t = \begin{cases} -\alpha(t) G p_A^t (1 - p_A^t) \text{ for } t \in [0, t_A), \\ 0 \text{ for } t \geq t_A, \end{cases}$$

$$\dot{p}_B^t = \begin{cases} -(1 - \alpha(t)) G p_B^t (1 - p_B^t) \text{ for } t \in [0, t_B), \\ 0 \text{ for } t \geq t_B, \end{cases}$$

are derived by requiring that the evolution of beliefs be consistent with α and Bayes' rule, following Keller et al. (2005). For convenience, I set $p_J^t = 0$ for any $t \geq t_J$.

Let $W(\alpha, t_A, t_B; t', \tau)$ be the experimenter's payoff at time t' to strategy (α, t_A, t_B) when a success arrives at time $\tau > t'$. If a success arrives at $\tau < \min\{t_A, t_B\}$

$$W(\alpha, t_A, t_B; t, \tau) = e^{-r(\tau - t')} - \int_{t'}^{\tau} e^{-rt} (\overline{k} + \underline{k}) dt,$$

while if a success arrives at $\tau \in [\min\{t_A, t_B\}, \max\{t_A, t_B\})$

$$W(\alpha, t_A, t_B; t', \tau) = e^{-r(\tau - t')} - \int_{t'}^{\min\{t_A, t_B\}} e^{-rt} (\overline{k} + \underline{k}) dt - \int_{\min\{t_A, t_B\}}^{\tau} e^{-rt} \overline{k} dt,$$

and finally, if a success never arrives,

$$W(\alpha, t_A, t_B; t', \tau) = - \int_{t'}^{\min\{t_A, t_B\}} e^{-rt} (\overline{k} + \underline{k}) dt - \int_{\min\{t_A, t_B\}}^{\max\{t_A, t_B\}} e^{-rt} \overline{k} dt.$$

The expected payoff to strategy (α, t_A, t_B) given belief $(p_A^{t'}, p_B^{t'})$ is

$$V(\alpha, t_A, t_B; t') \equiv \mathbf{E}_\tau W(\alpha, t_A, t_B; t', \tau),$$

where the expectation is taken over the distribution of stopping times τ determined by (α, t_A, t_B) and $(p_A^t, p_B^t)_{t=t'}^{\infty}$. Finally, denote the experimenter's optimal payoff at time t' by

$$U(t') = \max_{(\alpha, t_A, t_B)} V(\alpha, t_A, t_B; t').$$

Without loss of generality, I assume that $p_A^0 \geq p_B^0$. Furthermore, it is without loss of generality to restrict attention to strategies generating belief paths $(p_A^t, p_B^t)_t$ such that $p_A^t \geq p_B^t$ for all t. Since the arrival rates of successes depend only on beliefs, which decrease continuously when a project is put to trial and no success arrives, any strategy could be replaced by a payoff-equivalent strategy in the specified class through an appropriate relabelling of projects. Henceforth, project A will always be the 'winning' project, with project B the 'losing' project.

2.1 Optimal Payoff Functions

Keller et al. (2005) provide simple expressions for the experimenter's optimal payoff functions for the two-armed exponential bandit (optimal stopping) problem. In this section, I build on these results to derive the expressions satisfied by the optimal payoff $U(t)$ at time t that will support the characterisations of Sections 3 and 4. Given optimal stopping times t_A^* and t_B^*, the optimal payoff at t must satisfy the following Bellman equation

$$\begin{aligned} u(p_A^t, p_B^t) &\equiv U(t) \\ &= \max \Bigg\{ 0,\ \mathbb{I}_{t<t_A^*} u(p_A^t, 0),\ \mathbb{I}_{t<t_B^*} u(0, p_B^t),\ \mathbb{I}_{t<\min\{t_A^*, t_B^*\}} \max_{\beta\in[0,1]} \Big\{ [\beta p_A^t G \\ &\qquad + (1-\beta) p_B^t G - (\overline{k} + \underline{k})]dt + e^{-rdt} \mathbf{E}\left[u(p_A^t + dp_A, p_B^t + dp_B) | p_A^t, p_B^t \right] \Big\} \Bigg\}. \end{aligned} \tag{1}$$

The first term in the brackets of (1) corresponds to the option quitting experimentation. The second and third terms correspond to the options of discarding projects B and A respectively, and collecting the optimal payoff to the optimal stopping problem with a single risky project and the safe project. The final term corresponds to the payoff from maintaining both projects and allocating experimentation effort optimally. Any solution u to (1) must rely on an optimal rule for the allocation of trials among projects. I approach this problem by abstracting from allocation rules and considering in isolation those intervals of time in which a single project is put to trial.

Let t lie in an open interval of time in which project A is put to trial, in which case

$$u(p_A^t, p_B^t) = [p_A^t G - (\overline{k} + \mathbb{I}_{t<t_B^*}\underline{k})]dt + e^{-rdt}\mathbf{E}\left[u(p_A^t + dp_A, p_B^t) | p_A^t\right].$$

The probability of a success in an interval of length dt is $p_A^t G dt$, and the payoff to a success is 1. The probability of failure is $1 - p_A^t G dt$. In case of failure, the payoff to the experimenter is $u(p_A^t, p_B^t) + \frac{\partial}{\partial p_A} u(p_A^t, p_B^t) dp_A$, which is equal to $u(p_A^t, p_B^t) + \frac{\partial}{\partial p_A} u(p_A^t, p_B^t) p_A^t (1 - p_A^t) G dt$. Using $1 - rdt$ as an approximation to e^{-rdt} and cancelling dominated terms yields that u satisfies the differential equation

$$u(p_A^t, p_B^t)(r + G p_A^t) = p_A^t G - (\overline{k} + \mathbb{I}_{t<t_B^*}\underline{k}) - \frac{\partial}{\partial p_A} u(p_A^t, p_B^t) p_A^t (1 - p_A^t), \tag{2}$$

which can be solved to yield

$$\begin{aligned} v(p_A^t; C_A(p_B^t)) &\equiv u(p_A^t, p_B^t) \\ &= C_A(p_B^t) \left(\frac{1 - p_A^t}{p_A^t} \right)^{\frac{r}{G}} (1 - p_A^t) + p_A^t \frac{G - (\overline{k} + \mathbb{I}_{t<t_B^*}\underline{k})}{r + G} - (1 - p_A^t) \frac{\overline{k} + \mathbb{I}_{t<t_B^*}\underline{k}}{r}, \end{aligned} \tag{3}$$

where $C_A(p_B^t)$ is a constant of integration. The setup here is slightly different than in Keller et al. (2005), but the expression (3) admits a similar interpretation. The term $p_A^t \frac{G-(\overline{k}+\mathbb{I}_{t<t_B^*}\underline{k})}{r+G} - (1-p_A^t)\frac{\overline{k}+\mathbb{I}_{t<t_B^*}\underline{k}}{r}$ is the payoff to risky project A in the absence of the ability to quit experimentation or turn to project B, while the term $C_A(p_B^t)\left(\frac{1-p_A^t}{p_A^t}\right)^{\frac{r}{G}}(1-p_A^t)$ captures the option value of either experimenting with project B or quitting experimentation altogether. The constant of integration $C_A(p_B^t)$ depends in general on p_B^t, since the experimenter's belief about project B can affect its payoffs when exiting the A-assignment region. Equation (3) establishes a useful necessary condition for optimal payoffs when project A is put to trial and project B is maintained: payoffs evolve as though the experimenter was facing an optimal stopping problem with cost $\overline{k}+\underline{k}$ for the risky project, with the value to the stopping region adjusted to incorporate the payoffs to future experimentation with project B.

If optimal experimentation calls for exiting an A-assignment region at time $t^* > t$, then $p_A^{t^*}$ and $C_A(p_B^t)$ satisfy the value-matching and smooth-pasting properties

$$\begin{aligned} v(p_A^{t^*}; C_A(p_B^t)) &= u(p_A^{t^*}, p_B^t), \text{ and} \\ \frac{\partial}{\partial p_A} v(p_A^{t^*}; C_A(p_B^t)) &= \frac{\partial}{\partial p_A} u(p_A^{t^*}, p_B^t). \end{aligned} \tag{4}$$

In general, $u(p_A^{t^*}, p_B^t)$ is endogenous and depends on the experimenter's continuation strategy once exit from the A-assignment region occurs. An important special case is the stopping problem with a single risky arm A, which in my model corresponds to the experimentation problem once arm B has been discarded. In that case, $t \geq t_B^*$, $p_B^t = 0$ and $t_A^* = t^*$, and we have that $u(p_A^{t^*}, 0) = 0$ and $\frac{\partial}{\partial p_A} u(p_A^{t_B^*}, 0) = 0$, which yields, through (4), $p_A^{t_A^*} = \frac{\overline{k}}{G}$ and $C_A(0) = \left(\frac{\overline{k}}{G-\overline{k}}\right)^{\frac{r}{G}} \frac{G\overline{k}}{r(r+G)}$.

3 Benchmark: Stay-With-The-Winner

To highlight the impact of maintenance costs on optimal experimentation, a useful benchmark is the case in which risky projects cannot be discarded individually. In this problem, the experimenter can only coarsely manage project retention: it can either run trials at the cost of maintaining both risky projects or quit experimentation. Specifically, suppose that the experimenter is restricted to maintenance rules such that $t_A = t_B \equiv t_{ND}$. Note that this problem is equivalent to a standard three-armed bandit problem with direct experimentation flow cost $(\overline{k} + \underline{k})$. The next result shows that when projects cannot be discarded individually, optimal experimentation follows a 'stay-with-the-winner' rule.

Proposition 1. *Fix (p_A^0, p_B^0) with $p_A^0 \geq p_B^0$. When projects cannot be discarded individually, the*

following strategy is optimal

$$\alpha^*_{ND}(t) = \begin{cases} 1 & \text{if } p^t_A > p^t_B \text{ and } p^t_A > \frac{\overline{k}+\underline{k}}{G}, \\ \frac{1}{2} & \text{if } p^t_A = p^t_B > \frac{\overline{k}+\underline{k}}{G}, \\ Q & \text{if } p^t_A \leq \frac{\overline{k}+\underline{k}}{G}. \end{cases}$$

$$t^*_{ND} = \sup\left\{t : p^t_A \geq \frac{\overline{k}+\underline{k}}{G}\right\}.$$

Any optimal assignment strategy is identical to $\alpha^*_{ND}(t)$ for almost all t.[12]

Conditional on continuing experimentation at t, the experimenter always puts the project with the highest belief to trial. When beliefs are such that $p^t_A > p^t_B$, this means using project A. When beliefs are such that $p^t_A = p^t_B$, then both projects are 'winning' and 'staying-with-the-winner' entails sharing experimentation effort equally between them. The result of Proposition 1 mimics the Gittins index representation of the optimal experimentation policy, where in this case a project's belief is taken to be the index.[13] In fact, a key step in its proof is essentially a simplified version of the original 'interchange argument' in Gittins and Jones (1974) and Gittins (1979) that establishes the optimality of the Gittins index for standard bandit problems.[14] Starting with an assignment in which a project with a non-maximal Gittins index is chosen before the project with the maximal index, the argument shows that interchanging the order in which both projects are put to trial, keeping expected continuation assignments following these (random) periods of experimentation fixed, increases the experimenter's payoff.

Continuing experimentation is optimal as long as one project's belief is above the cutoff $\frac{\overline{k}+\underline{k}}{G}$, which, since $p^t_A \geq p^t_B$ for all t, holds as long as $p^t_A > \frac{\overline{k}+\underline{k}}{G}$. Figure 2 illustrates belief dynamics consistent with the optimal experimentation policy $(\alpha^*_{ND}, t^*_{ND})$. Given initial belief (p^0_A, p^0_B) with $p^0_A > \frac{\overline{k}+\underline{k}}{G} > p^0_B$, only project A is ever put to trial, until t^*_{ND} such that $p_A^{t^*_{ND}} = \frac{\overline{k}+\underline{k}}{G}$ (and $p_B^{t^*_{ND}} = p^0_B$). For initial beliefs with this property, the optimal payoff function can be derived through conditions (4) with $u(p_A^{t^*_{ND}}, p_B^{t^*_{ND}}) = \frac{\partial}{\partial p_A} u(p_A^{t^*_{ND}}, p_B^{t^*_{ND}}) = 0$. From initial belief (p'^0_A, p'^0_B) with $p'^0_A > p'^0_B > \frac{\overline{k}+\underline{k}}{G}$, experimenting with project A followed by shared experimentation until t^*_{ND} such that $(p_A^{t^*_{ND}}, p_B^{t^*_{ND}}) = (\frac{\overline{k}+\underline{k}}{G}, \frac{\overline{k}+\underline{k}}{G})$ is optimal. In the Appendix, I derive optimal payoffs under shared experimentation, using the fact that under shared experimentation with shared belief $p^t_A = p^t_B = p^t$, a differential equation like (2), which depends only on p^t, can be derived. However, since each project receives half of the experimentation allocation, the decay in the shared belief p^t is slower, although the quitting belief is the same as in the one-project case.

[12] All proofs are in the Appendix.

[13] This is a special feature of exponential bandits, so that the myopically optimal allocation is also dynamically optimal. This was first shown for discrete time Bernoulli bandits by Berry and Fristedt (1985). Their result was generalised in Banks and Sundaram (1992a), who show that dynamically optimal play is myopic for a class of two-type symmetric bandit problems.

[14] For a concise presentation of the original proof, see Frostig and Weiss (1999).

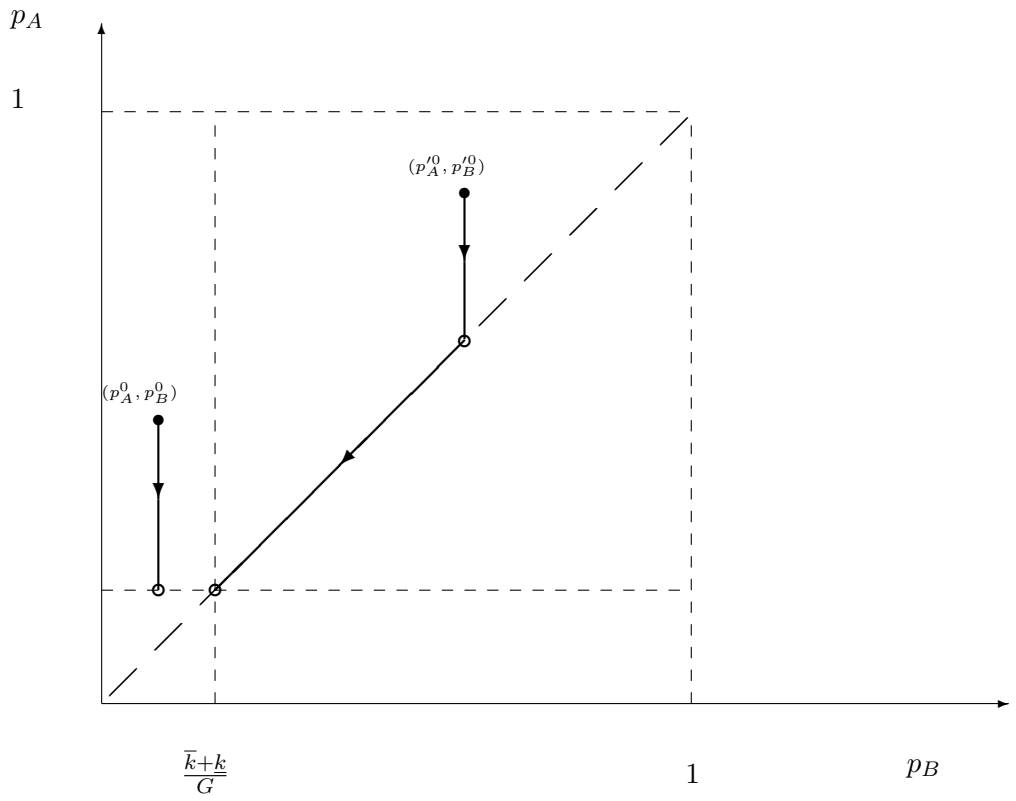


Figure 2: Optimal experimentation without discarding individual projects.

4 Optimal Experimentation with Maintenance Costs

In this section, I allow the experimenter to discard projects individually and present the main results of the paper. Under the 'stay-with-the-winner' rule, the experimenter can avoid accumulating maintenance costs on the inactive losing project B only by discarding this project. However, discarding project B is irreversible and carries the opportunity cost attached to the eventuality in which project B becomes valuable following repeated failures by project A. This tension generates an incentive to abandon the 'stay-with-the-winner' rule in order to bring the option value of project B forward.

4.1 A Culling Rule

As a first step, I provide necessary conditions for the losing project B to be put to trial under optimal experimentation, showing that optimal assignment strategies when projects can be discarded individually follow simple rules.

Proposition 2. *Fix (p_A^0, p_B^0) with $p_A^0 \geq p_B^0$. An optimal strategy (α^*, t_A^*, t_B^*) satisfies*

(i). $t_A^* \geq t_B^*$ (without loss of generality).

(ii). $t_B^* = \inf\{t' : \alpha^*(t) \in \{1, Q\}$ for almost all $t \geq t'\}$.

(iii). Suppose that there exist $t' < t'' < t_B^*$ such that $p_A^t > p_B^t$ for all $t \in [t', t'')$ and that $\alpha^*(t) < 1$ for almost all $t \in [t', t'')$. Then $\alpha(t) = 0$ for almost all $t \in [t', t_B^*)$.

In words, Proposition 2 shows that (i) if a project is ever discarded while trials are still being run it is the losing project B, and (ii) it is never discarded in the middle of a phase in which project A is put to trial but rather as soon as the experimenter's assessment strategy no longer intends to put it to trial, while (iii) whenever project A has a strictly higher belief than project B but it is not exclusively on trial, then it must be that project B is exclusively on trial until t_B^* at which it is discarded. This last property establishes that experimentation with the losing project B must take the form of a simple but powerful culling rule: a losing project can be given priority only in the form of a 'last chance' to produce a success. Continued failure in this period of reprieve leads to the abandonment of the project.

Proving parts (i) and (ii) of Proposition 2 is straightforward. First, suppose that the winning project A (which is without loss of generality) is discarded before B and that B is put to trial after having discarded A. Then an alternative assignment strategy inverting the roles of projects A and B would increase the experimenter's payoff. Second, suppose that project B is maintained from t' to $t'' > t'$ but not put to trial. Then under an alternative strategy in which project B is discarded at t', successes on A would occur with the same probability and maintenance costs would be avoided for an amount of time of positive expected length. The proof of part (iii) of Proposition 2 relies on Lemma 2, stated in the Appendix, which shows that, conditional on both projects being maintained, optimal experimentation requires that the project with the highest belief be put to trial.[15] Hence, a necessary condition for the existence of a period of experimentation in which the losing project B is put to trial while the winning project A is maintained is that this period is terminated by project B being discarded.

[15] This lemma, along with Lemma 3, also stated in the Appendix, is key in establishing the 'stay-with-the-winner' property in Proposition 1.

4.2 The Discarding Decision

In this section, I characterise the experimenter's decision to discard project B following a culling phase of experimentation. To this end, suppose that $p_A^t > p_B^t$ and that t lies in an open interval in which project B is put to trial. Then, by Proposition 2, project B is put to trial exclusively until $t_B^* > t$ at which it is discarded in the absence of a success. The experimenter's payoff at t satisfies

$$u(p_A^t, p_B^t) = v(p_B^t; C_B(p_A^t)),$$

for some constant of integration $C_B(p_A^t)$ where $p_A^{t_B^*} = p_A^t$. The experimenter's payoff at time t_B^* once project B has been discarded is given by $u(p_A^t, 0) = v(p_A^t, C_A(0))$ and is independent of the belief about project B. Hence value-matching and smooth-pasting conditions at time t_B^* yield

$$v(p_B^{t_B^*}; C_B(p_A^t)) = v(p_A^t, C_A(0)), \text{ and} \tag{5}$$

$$\begin{aligned} \frac{\partial}{\partial p_B} v(p_B^{t_B^*}; C_B(p_A^t)) &= \frac{\partial}{\partial p_B} v(p_A^t, C_A(0)) \\ &= 0, \end{aligned} \tag{6}$$

where $p_A^t = p_A^{t_B^*}$. Rearranging (6) yields

$$C_B(p_A^t) \left(\frac{1 - p_B^{t_B^*}}{p_B^{t_B^*}} \right)^{\frac{r}{G}} = \frac{G(r + \overline{k} + \underline{k})}{r(r+G)} \frac{p_B^{t_B^*} G}{p_B^{t_B^*} G + r},$$

which, along with (5), yields that $p_B^{t_B^*}$ solves

$$v(p_A^t, C_A(0)) = \frac{p_B^{t_B^*} G - (\overline{k} + \underline{k})}{p_B^{t_B^*} G + r}. \tag{7}$$

Equation (7) is a necessary condition: it does not determine whether project B is actually ever put to trial if $p_A^t > p_B^t$, just when it should be discarded if it is used. Note that the right-hand side in (7) is the experimenter's payoff to a project that is known to be of type *Good* but has a success rate $p_B^{t_B^*} G$ and associated experimentation cost $\overline{k} + \underline{k}$. Hence (7) states that at a cutoff belief $(p_A^t, p_B^{t_B^*})$ at which project B is discarded, the experimenter is indifferent between its payoff to risky project A in the absence of project B and a risk-less project with a flow payoff equal to project B's flow payoff at the belief $p_B^{t_B^*}$, which includes the maintenance costs associated to project A.

It can be verified that the mapping $p_B \mapsto \frac{p_B G - (\overline{k} + \underline{k})}{p_B G + r}$ is increasing and concave. It follows that, given p_A^t, if a solution $p_B^{t^*}$ to (7) exists, it is unique. Hence, define mapping $p_B^* : [0,1] \to [0,1]$ such that $p_B^*(p_A^t)$ is the unique solution to (7) if it exists, and is equal to p_A^t otherwise. The locus $(p_A, p_B^*(p_A))_{p_A \in [0,1]}$ defines a *discarding boundary* in belief-space. Note that a necessary condition

for project B to be put to trial before project A under some initial beliefs is that there exists belief p_A such that $p_B^*(p_A) < p_A$, which occurs whenever, for fixed p_A, there exists p_B such that $v(p_A, C_A(0)) < \frac{p_B G - (\overline{k}+\underline{k})}{p_B G + r}$, for which a necessary and sufficient condition is that there exists p_A such that

$$\begin{aligned} z(p_A) &\equiv v(p_A, C_A(0)) - \frac{p_A G - (\overline{k}+\underline{k})}{p_A G + r} \\ &< 0. \end{aligned} \tag{8}$$

Since $v(p_A, C_A(0))$ is increasing and convex, $z(p_A)$ is convex and the set $\{p_A : z(p_A) \leq 0\}$ is convex. If this set is nonempty, it can be written as $[\underline{p}_A, \overline{p}_A]$, for beliefs $\underline{p}_A \leq \overline{p}_A$. In particular, note that

$$\begin{aligned} \lim_{p_A \nearrow 1} \left[v(p_A^t, C_A(0)) - \frac{p_A G - (\overline{k}+\underline{k})}{p_A G + r} \right] &= \frac{G - \overline{k}}{G + r} - \frac{G - (\overline{k}+\underline{k})}{G + r} \\ &> 0, \end{aligned}$$

showing that, if it exists, $\overline{p}_A < 1$. That is, if the probability that project A is of type *Good* is close to 1, the experimenter's payoff to a single risky project with cost $\overline{k}$ is close to the payoff to a project known to be of type *Good* with cost $\overline{k}$ and success rate $p_A G \approx G$. This dominates the payoff to a project known to be of type *Good* with cost $\overline{k} + \underline{k}$ and success rate G. Intuitively, the experimenter has a strict incentive to discard project B when project A is almost sure to be good. Furthermore,

$$\begin{aligned} \lim_{p_A \searrow \frac{\overline{k}+\underline{k}}{G}} \left[v(p_A, C_A(0)) - \frac{p_A G - (\overline{k}+\underline{k})}{p_A G + r} \right] &= v(\frac{\overline{k}+\underline{k}}{G}; C_A(0)) \\ &> 0, \end{aligned}$$

showing that, if it exists, $\underline{p}_A > \frac{\overline{k}+\underline{k}}{G}$. That is, if p_A is close to the quitting belief $\frac{\overline{k}+\underline{k}}{G}$ (for a risky project with flow cost $\overline{k} + \underline{k}$), the experimenter's payoff to a project known to be of type *Good* with flow cost $\overline{k} + \underline{k}$ and success rate $p_A G \approx \overline{k} + \underline{k}$ is close to 0, while the payoff to a risky project with flow cost $\overline{k}$ is strictly positive, since the quitting belief attached to this stopping problem is $\frac{\overline{k}}{G}$. Comparing this to the result of Proposition 1, the experimenter has a strict incentive to discard arm B before beliefs (p_A^t, p_B^t) reach $(\frac{\overline{k}+\underline{k}}{G}, \frac{\overline{k}+\underline{k}}{G})$, at which an experimenter that cannot discard arms individually discards both projects.

Figure 3 illustrates the discarding boundary in belief-space when the condition from (8) obtains. Define

$$\mathcal{P}_M = \{(p_A, p_B) : p_A \geq p_B, p_B \geq p_B^*(p_A)\},$$

which is the maintenance region: the set of beliefs inside which project B is never discarded. Further define

$$\mathcal{P}_D = \{(p_A, p_B) : p_A \geq p_B\} \setminus \mathcal{P}_M,$$

which is the discarding region: the set of beliefs inside which project B can be discarded immediately or maintained but never put to trial. From initial belief (p_A^0, p_B^0) in the interior of $\mathcal{P}_M$, if it is optimal to experiment with project B, then B must be put to trial until time $t_B^* > 0$ such that$(p_A^{t_B^*}, p_B^{t_B^*}) = (p_A^0, p_B^*(p_A^0))$, after which B is discarded and A is put to trial until time t_A^* at which $p_A^{t_A^*} = \frac{\overline{k}}{G}$, the quitting belief for the stopping problem with a single risky project.

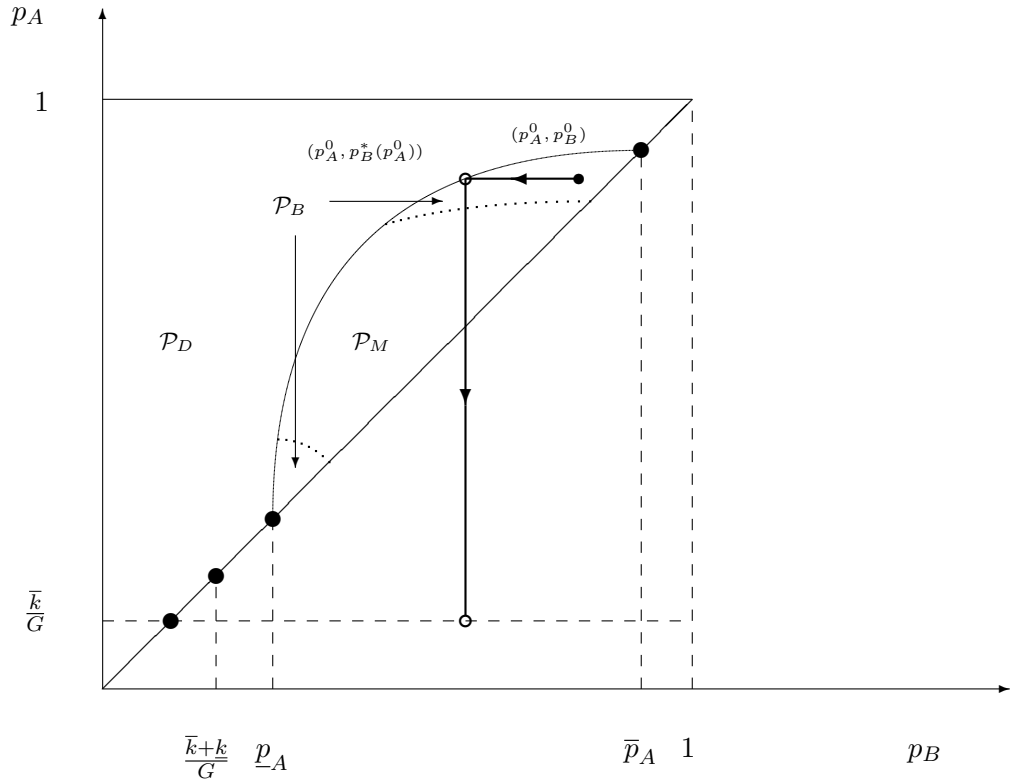


Figure 3: The discarding boundary and belief dynamics in a culling phase.

4.3 'Going-with-the-Loser'

The simple necessary condition from (8) for the optimality of a culling phase, that there exists a belief p_A such that $p_B^*(p_A) < p_A$, is also sufficient to guarantee the existence of an open set of beliefs in which project B is put to trial before project A. The following proposition contains the main result of the paper: 'going-with-the-loser' is optimal whenever it is not the case that maintenance costs are so high that the losing project is immediately discarded at all initial beliefs.

Proposition 3. *There exists a cutoff maintenance cost* $\underline{k}^*(\overline{k}, G, r)$ *such that*

(i). If $\underline{k} \geq \underline{k}^*(\overline{k}, G, r)$ *then* $t_B^* = 0$ *for all initial beliefs* (p_A^0, p_B^0) *such that* $p_A^0 \geq p_B^0$.

(ii). If $\underline{k} < \underline{k}^*(\overline{k}, G, r)$, *then there exist* $\overline{p}_A > \underline{p}_A$ *such that, for almost all* $t < t_B^*$, $\alpha^*(t) = 0$ *only if* $p_A^t \in [\underline{p}_A, \overline{p}_A]$ *and* $p_B^t \in [p_B^*(p_A^t), p_A^t]$. *Furthermore, there exists an open set* I *of initial beliefs with* $p_A^0 \geq p_B^0$ *such that, given any* $(p_A^0, p_B^0) \in I$, $t_B^* > 0$ *and* $\alpha^*(t) = 0$ *for almost all* $t \in [0, t_B^*)$.

If maintenance costs are sufficiently high (part (i)), then having multiple projects yields no value to the experimenter, who immediately discards project B given any initial beliefs. Otherwise (part (ii)), maintenance costs are sufficiently low and there exist initial beliefs for which a culling phase for project B is optimal. The region $\mathcal{P}_B$, illustrated in Figure 3, contains some initial beliefs for which the argument in the proof of part (ii) applies, which are those beliefs close to $(\underline{p}_A, \underline{p}_A)$ and $(\overline{p}_A, \overline{p}_A)$. For initial beliefs (p_A^0, p_B^0) close to $(\overline{p}_A, \overline{p}_A)$, (7) yields that the stopping problem with project A after project B has been discarded yields a payoff close to the payoff to a project known to be of type *Good* with success rate $\overline{p}_A G$ and experimentation cost $\overline{k} + \underline{k}$. However, since the experimenter's beliefs only decrease following failures, the payoff to putting project A on trial (or to shared experimentation) and maintaining B is strictly less than the payoff to using a project known to be of type *Good* with success rate $\overline{p}_A G$ and experimentation cost $\overline{k} + \underline{k}$. Hence, discarding project B immediately yields strictly higher payoffs than using project A and maintaining project B (or shared experimentation). An intuitive interpretation is that for initial beliefs around $(\overline{p}_A, \overline{p}_A)$, the experimenter has decided that it does not wish to maintain both projects in the long run. It remains that if a single project is to be maintained it should be project A. Yet, for initial beliefs in the interior of $\mathcal{P}_M$, using project B until the discarding boundary yields strictly higher payoffs than discarding it immediately. That is, the option represented by project B has enough value that the experimenter wants to exploit it before discarding it, which means that project B must take precedence over project A.

The same argument applies around $(\underline{p}_A, \underline{p}_A)$, with an interesting difference in interpretation. For initial beliefs around $(\overline{p}_A, \overline{p}_A)$, the experimenter chooses to maintain a single project in the long run because it is sufficiently optimistic about the quality of both projects. Since both projects are likely to eventually yield a success, it puts the losing project B to trial in the hope of a quick success and then discards it if these trials fail and moves on to the winning project A. For initial beliefs around $(\underline{p}_A, \underline{p}_A)$, the experimenter chooses to maintain a single project in the long run because it is pessimistic about the quality of both projects. Neither project is likely to yield a success, so that exploiting some of the residual value of project B and pinning its hopes on developing project A at a reduced cost is optimal. This difference in the two regions of beliefs composing $\mathcal{P}_B$ will be important in Section 5.2 when I examine the case in which project successes can be accumulated. In that case, 'going-with-the-loser' when the experimenter

is optimistic about both projects need no longer be optimal, but it remains optimal when the experimenter is pessimistic about both projects.

4.4 Optimal Experimentation Policies

My previous results pin down the possible dynamics of trial allocations among projects. Given $p_A^0 > p_B^0$, either (a) project B is discarded immediately, (b) project B is put to trial exclusively until it is discarded in favour of project A or (c) project A is put to trial exclusively until either a switch to B occurs and B is put to trial until it is discarded, or belief p_A^t drops to p_B^0, which leads to shared experimentation. Furthermore, a period of shared experimentation is always followed by a culling phase for project B. In this final section, I complete the characterisation of optimal experimentation by deriving when it entails these various patterns. The arguments are constructive and involve backward induction from the possible endpoints of experimentation dynamics.

A useful starting point are beliefs (p_A^t, p_B^t) such that $p_A^t = p_B^t$ and $t < t_B^*$. For such beliefs, either there is shared experimentation ('stay-with-the-winner') or project B enters a culling phase ('go-with-the-loser').

Lemma 1. *There exists $\underline{p}, \overline{p} \in (\underline{p}_A, \overline{p}_A)$ with $\underline{p} < \overline{p}$ such that, if $p_A^t = p_B^t$ and $\alpha^*(t) = \frac{1}{2}$ for almost all $t \in [t', t'')$, then*

(i). $p_A^{t'} \leq \overline{p}$.

(ii). $p_A^{t''} \geq \underline{p}$ and $\alpha^*(t) = \frac{1}{2}$ for almost all $t \in [t', t''')$, where $p_A^{t'''} = \underline{p}$.

Hence, there exists a unique belief $\underline{p}$ at which a transition from shared experimentation to a culling phase for project B can occur.[16] Proposition 3 implies that (a) $\underline{p} > \underline{p}_A$ and that $\alpha^*(t) = 0$ for almost all t such that $p_A^t = p_B^t \in [\underline{p}_A, \underline{p})$, and that (b) if $\alpha^*(t) = \frac{1}{2}$ for almost all $t \in [t', t'')$, then there exists belief $\overline{p} < \overline{p}_A$ such that $p_A^{t'} \leq \overline{p}$ and $\alpha^*(t) = 0$ for almost all t such that $p_A^t = p_B^t \in (\overline{p}, \overline{p}_A]$. That is shared experimentation can be optimal only for those beliefs (p_A^t, p_B^t) with $p_A^t = p_B^t \in (\underline{p}, \overline{p}]$. However, the set of beliefs for which shared experimentation is optimal can be empty.

This simple characterisation of shared experimentation allows the definition of two sets of beliefs, $\mathcal{P}_L \subseteq \mathcal{P}_M$ and $\mathcal{P}_W$, which define an optimal assignment strategy. Both these sets contain only beliefs (p_A^t, p_B^t) such that $t < t_B^*$, that is, at which both projects are maintained. $\mathcal{P}_W$ corresponds to the set of beliefs at which 'staying-with-the-winner' (and maintaining the loser) is optimal whereas $\mathcal{P}_L$ corresponds to the set of beliefs at which 'going-with-the-loser' (and maintaining the winner) is optimal. The details, which follow from a backwards induction argument, are left to the Appendix.

[16] This belief is derived explicitly in the Appendix.

Proposition 4. *Fix (p_A^0, p_B^0) with $p_A^0 \geq p_B^0$. When projects can be discarded individually, there exist sets of beliefs $\mathcal{P}_W$ and $\mathcal{P}_L$ such that the following strategy is optimal.*

$$\alpha_D^*(t) = \begin{cases} 0 & \text{if } (p_A^t, p_B^t) \in \mathcal{P}_L, \\ 1 & \text{if } (p_A^t, p_B^t) \in \mathcal{P}_W \text{ and } p_A^t > p_B^t, \text{ or if } (p_A^t, p_B^t) \in \mathcal{P}_D \cap \mathcal{P}_W^c \text{ and } p_A^t \geq \frac{\overline{k}}{G}, \\ \frac{1}{2} & \text{if } (p_A^t, p_B^t) \in \mathcal{P}_W \text{ and } p_A^t = p_B^t, \\ Q & \text{otherwise.} \end{cases}$$

$$t_B^* = \sup\left\{t : (p_A^t, p_B^t) \in \mathcal{P}_W \cup \mathcal{P}_L\right\},$$

$$t_A^* = \sup\left\{t : p_A^t \geq \frac{\overline{k}}{G}\right\}.$$

Any optimal assignment strategy is identical to $\alpha_D^*(t)$ for almost all t.

Figure 4 provides an illustration of the sets $\mathcal{P}_W$ and $\mathcal{P}_L$.[17] The discarding region $\mathcal{P}_D$ is also depicted, and the maintenance region $\mathcal{P}_M$ corresponds to $\mathcal{P}_L \cup (\mathcal{P}_W \cap \mathcal{P}_D^c)$. Note that the set $\mathcal{P}_B$ from Figure 3 is contained in $\mathcal{P}_L$. When $\mathcal{P}_W$ contains a segment of the 45-degree line, the reversal of the 'stay-with-the-winner' property under maintenance costs exhibits a noteworthy non-monotonicity: for fixed p_B^t, 'going-with-the-loser' is optimal only if p_A^t is neither too high nor too low. In the absence of maintenance costs, the option value of project B only contains this project's usefulness once it is put to trial, which is captured by its belief p_B^t. With maintenance costs, the option value to project B also includes the likelihood with which it is exercised and the amount of time elapsed before doing so. Fix three pairs of initial beliefs such that $(p_A^0, p_B^0) \in \mathcal{P}_D$, $(p_A'^0, p_B^0) \in \mathcal{P}_L$ and $(p_A''^0, p_B^0) \in \mathcal{P}_W$. Given initial belief (p_A^0, p_B^0), project A is believed to be of type *Good*, and hence to succeed quickly, with high probability, and project B's option value is unlikely to be exercised. Hence, it is best for the experimenter to discard project B immediately and exploit project A only. 'Going-with-the-loser' is optimal only for intermediate beliefs for project A. At initial belief $(p_A'^0, p_B^0)$, 'going-with-the-winner' would put project A to trial until its belief dropped to p_B^0, after which experimentation would be shared. However, project A is both (a) not thought likely to succeed fast enough to dwarf the option value of project B but (b) of sufficient quality that paying to maintain the option value of project B under 'stay-with-the-winner' is too costly, since this value can be realised only after project A has failed for a long time. Hence, the experimenter opts for bringing the option value of project B forward by putting it to trial first. For initial beliefs $(p_A''^0, p_B^0)$, it is optimal to order project development as though there were no maintenance costs, since the realisation of project B's option value under 'stay-with-the-winner' occurs after a short amount of time. The belief dynamics in the absence of a success given initial beliefs $(p_A''^0, p_B^0)$ are depicted in figure 4.[18]

[17] In the figure I depict the case in which $\mathcal{P}_W$ is connected, which need not be the case in general.

[18] In the three-period example of Section 1.2, initial beliefs (p_A^0, p_B^0) were such that two trials on project A were always required before project B became the winning project. Hence, an increase in p_B^0 did not reduce the

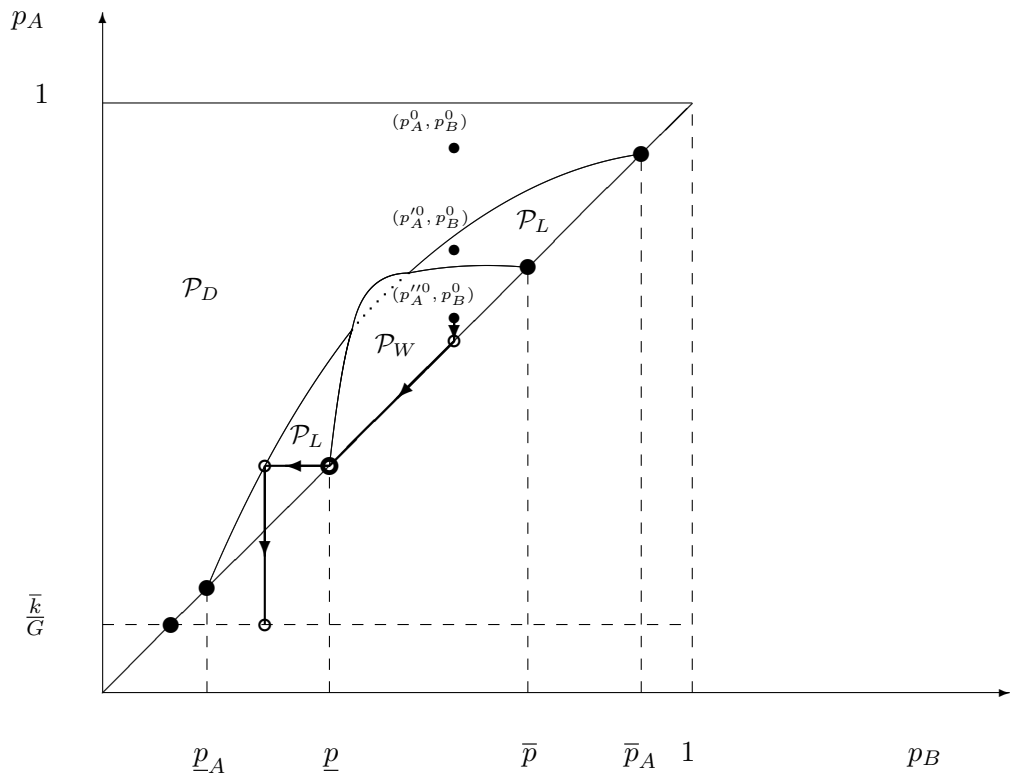


Figure 4: Optimal experimentation policy when individual projects can be discarded.

5 Extensions

5.1 A Culling Rule with More than Two Risky Projects

While the characterisation of optimal experimentation policies becomes more involved, no new conceptual difficulties arise if the experimenter has more than two risky projects. The key is that optimal trial allocation dynamics following the choice of a non-winning project are qualitatively similar to those uncovered by Proposition 2 for the two-project case. Consider the case in which there are three risky projects, A, B and C. Generalising my arguments to the case in which there are even more risky projects is straightforward. Without loss of generality, assume that

expected maintenance cost payments on the losing project required to realise its option value under the 'stay-with-the-winner' strategy. In that case, contrary to what is depicted in Figure 4 for the infinite horizon model, 'going-with-the-loser' was optimal for high values of p_B^0 and 'staying-with-the-winner' was optimal for intermediate values of p_B^0.

$p_A^t \geq p_B^t \geq p_C^t$ for all t. Given t, let $\alpha(t) = (\alpha_A(t), \alpha_B(t), \alpha_C(t))$, where $\sum_{J\in\{A,B,C\}} \alpha_J(t) = 1$ if and only if $\alpha(t) \neq Q$.

Proposition 5. *Fix (p_A^0, p_B^0, p_C^0) with $p_A^0 \geq p_B^0 \geq p_C^0$. Suppose that there exist $t' < t'' < t_C^*$ such that, for all $t \in [t', t'')$, $p_A^t > p_B^t > p_C^t$. An optimal strategy (α^*, t_A^*, t_B^*) satisfies*

(i). If $\alpha_C^*(t) > 0$ for almost all $t \in [t', t'')$, then $\alpha_C^*(t) = 1$ for almost all $t \in [t', t_C^*)$.

(ii). If $\alpha_B^*(t) > 0$ for almost all $t \in [t', t'')$, then $\alpha_A^*(t) = 0$ for almost all $t \in [t', t_B^*)$.

Part (i) is the result in three-project model corresponding to part (iii) of Proposition 2 in the two-project model: if the losing project (here project C) is ever put to trial, this takes the form of a culling rule. Part (ii) extends this rule to the 'middle' project B: it it is ever optimal to put this project to trial, then experimentation can proceed to the winning project A only after both non-winning projects B and C have been discarded. That is, the 'last chance' extended by the experimenter applies not only to project B but to all projects ranked lower than B, so that optimal experimentation will either put the winning project to trial or enter a targeted culling phase in which it puts to trial and then discards all sufficiently poor projects. Proposition 5 and my characterisation of the two-project case could lead, through an induction argument, to a full, although tedious, characterisation of optimal experimentation with three (and then more) projects.

5.2 Additive Project Sucesses

I have assumed that the outcomes of the projects are perfect substitutes in that the experimenter searches for a single success on either project. An alternative assumption is that a success on a given project retires that particular project but the experimenter obtains a payoff of 1 from all projects that succeed. This section shows that the result of Proposition 3, that the set of initial beliefs for which it is optimal to 'go-with-the-loser' is non-negligible, continues to hold with additive project successes. A version of Proposition 2 can be derived in this extension of the model, which, while both projects are maintained, allows the same types of optimal trial allocation dynamics as the model with substitutable successes. It is straighforward to show that, given t in an open interval of time in which project A is put to trial and project B is maintained, the experimenter's payoff must follow the following version of (2),

$$v_A(p_A^t; C_A(p_B^t))[r + p_A^t G] + \frac{\partial}{\partial p_A} v_A(p_A^t; C(p_B^t)) G p_A^t (1 - p_A^t) = p_A^t G[1 + v_B(p_B^t; C_B(0))] - (\overline{k} + \underline{k}),$$

where the difference with (2) is that when successes are additive, the success of project A leads to continued experimentation with project B. Hence, the experimenter's payoff to a success on

A is $1 + v_B(p_B^t; C_B(0))$. By replicating the arguments of Section 4.2, the relevant version of (7), which determines the discarding boundary, is

$$v_A(p_A^t; C_A(0)) = \frac{p_B^{t_B^*} G[1 + v_A(p_A^t; C_A(0))] - (\overline{k} + \underline{k})}{G p_B^{t_B^*} + r}. \tag{9}$$

Since $v_A(p_A^t; C_A(0)) > 0$ whenever project A is maintained, (9) shows that the maintenance region $\mathcal{P}_M$ is larger when successes are additive, as this enhances the value of holding on to project B. One difference with the discarding boundary (7) when projects are perfect substitutes is that, intuitively, project B need not be discarded when p_A^t is close to 1, which, recalling Section 4.2, means that a belief $(\overline{p}_A, \overline{p}_A)$ need not exist. However, as

$$\lim_{p_A^t \searrow \frac{\overline{k}}{G}} \left[v_A(p_A^t; C_A(0)) - \frac{p_A^t G[1 + v_A(p_A^t; C_A(0))] - (\overline{k} + \underline{k})}{p_A^t G + r} \right] > 0,$$

if $\mathcal{P}_M$ is nonempty, it must be that there exists a boundary belief $(\underline{p}_A, \underline{p}_A)$ with $\underline{p}_A > \frac{\overline{k}}{G}$ in the maintenance region $\mathcal{P}_M$. This allows the application of the argument in the proof of part (ii) of Proposition 3 for those beliefs (p_A^t, p_B^t) close to $(\underline{p}_A, \underline{p}_A)$. For such beliefs (which exist if a corresponding condition ruling out a case like that of part (i) of Proposition 3 is met), when discarding project B in favour of project A, the experimenter is guaranteed the payoff to a project known to be of type *Good* that succeeds at rate

$$p_B^{t_B^*} G \left[1 + v_A(p_A^t; C_A(0))\right] \Big|_{p_A^t \approx p_B^{t_B^*} \approx \underline{p}_A}.$$

Meanwhile, the experimenter's payoff from experimenting with project A and maintaining project B yields stricly less than the payoff to a project known to be of type *Good* that succeeds at rate

$$p_A^t G \left[1 + v_B(p_B^{t_B^*}; C_B(0))\right] \Big|_{p_A^t \approx p_B^{t_B^*} \approx \underline{p}_A}.$$

Hence, at initial beliefs close to $(\underline{p}_A, \underline{p}_A)$, the experimenter strictly prefers discarding project B immediately to 'going-with-the-winner', but for beliefs in the interior of $\mathcal{P}_M$, it strictly prefers 'going-with-the-loser' to discarding project B immediately, which yields the result.

When successes can be accumulated, the experimenter has no incentive to 'go-with-the-loser' in order to maintain a single project in the long run when it is optimistic about both projects. A quick success on the winning project A when B is maintained kills the value of project B and sinks the expended maintenance costs under perfectly substitutable successes, while it allows the experimenter to exploit project B's value promptly under additive successes. However, when the experimenter is pessimistic about both projects, the likelihood of obtaining two successes is small and the experimenter has an incentive to economise on maintenance costs by focusing its research efforts on the winning project.

6 Conclusion

The standard approach to experimentation has been to assume that when currently occupied by other projects, keeping the option of researching various alternatives at later dates is costless. However, that keeping options open can involve maintenance costs is natural in many settings. This paper shows that such costs generate new trade-offs for experimenters by giving them incentives to manage the timing of the realisation of inactive alternatives' option values and have important implications for optimal experimentation policies. In particular, my results provide a rationale for project development in which less promising alternative receive priority over more promising ones.

While I have focused on the simple and tractable exponential bandit framework, it is not unreasonable to expect that my main arguments extend to more general bandit settings. Note that more generally, the arguments used in the paper are based on finite backwards induction, where the recursion is on the set of maintained projects. At each step of the recursion, the arguments rely on maintained projects' Gittins indices. This is made clear by the common structure of Lemma 2 and the original 'interchange argument' of Gittins (1979) that establishes the optimality of index policies in standard bandit problems. Investigating the effects of maintenance costs on optimal experimentation in general bandits is an interesting avenue for future work.

A unified treatment of experimentation with both maintenance and switching costs also presents interesting possibilities. In an environment in which an experimenter can acquire new projects at a cost at any time, it is plausible that start-up costs for inactive projects are higher for those projects that have not been maintained. In that case, the experimenter would trade off (a) acquiring multiple projects early and paying to maintain inactive projects and (b) acquiring one project at a time and bearing switching costs whenever project development priorities change. A model along these lines could shed light on when a portfolio approach to experimentation is preferred to a more focused sequential approach. Experimentation dynamics would vary in these approaches, as an experimenter with a portfolio of projects purchases through maintenance costs the flexibility of re-prioritising between them easily, whereas an experimenter acquiring new projects as needed saves on maintenance costs by facing higher switching costs, and so should delay transitions between projects.

References

Banks, J. and R. Sundaram (1992a). A class of bandit problems yielding myopic optimal strategies. *Journal of Applied Probability 29*(3), 625–632.

Banks, J. and R. Sundaram (1992b). Denumerable-armed bandits. *Econometrica 60*(5), 1071–1096.

Banks, J. and R. Sundaram (1994). Switching costs and the Gittins index. *Econometrica 62*(3), 687–694.

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

Bergemann, D. and J. Välimäki (2001). Stationary multi-choice bandit problems. *Journal of Economic Dynamics and Control 25*(10), 1585–1594.

Bergemann, D. and J. Välimäki (2008). Bandit Problems. *The New Palgrave Dictionary of Economics, 2nd ed.*.

Berry, D. and B. Fristedt (1985). *Bandit problems*. London: Chapman and Hall London.

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

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

Frostig, E. and G. Weiss (1999). Four proofs of Gittins' multiarmed bandit theorem. *Applied Probability Trust*, 1–20.

Garland, H. (1990). Throwing good money after bad: The effect of sunk costs on the decision to escalate commitment to an ongoing project. *Journal of Applied Psychology 75*(6), 728–731.

Gittins, J. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society. Series B (Methodological)*, 148–177.

Gittins, J. and D. Jones (1974). A dynamic allocation index for the sequential design of experiments. *Progress in statistics 241*, 266.

Hörner, J. and L. Samuelson (2009). Incentives for Experimenting Agents. *Cowles Foundation Discussion Papers*.

Jun, T. (2004). A survey on the bandit problem with switching costs. *de Economist 152*(4), 513–541.

Keller, G. and S. Rady (2010). Strategic experimentation with poisson bandits. *Theoretical Economics 5*(2), 275–311.

Keller, G., S. Rady, and M. Cripps (2005). Strategic experimentation with exponential bandits. *Econometrica 73*(1), 39–68.

Klein, N. (2011). The importance of being honest.

Klein, N. and S. Rady (2011). Negatively correlated bandits. *The Review of Economic Studies 78*(2), 693–732.

McAfee, R., H. Mialon, and S. Mialon (2010). Do Sunk Costs Matter? *Economic Inquiry 48*(2), 323–336.

Staw, B. (1981). The escalation of commitment to a course of action. *Academy of management Review 6*(4), 577–587.

Staw, B. and J. Ross (1987). Knowing when to pull the plug. *Harvard Business Review 65*(2), 68–74.

Strulovici, B. (2010). Learning while voting: Determinants of collective experimentation. *Econometrica 78*(3), 933–971.

A Appendix

I first present two related lemmas that will be useful in many of my results. The first establishes that if the experimenter fails to 'stay-with-the-winner', then it must put the losing project to trial exclusively.

Lemma 2. *Fix (p_A^0, p_B^0) such that $p_A^0 \geq p_B^0$ and an optimal strategy (α^*, t_A^*, t_B^*). Suppose that there exists $t' < t'' < \min\{t_A^*, t_B^*\}$ such that $p_A^t > p_B^t$ for all $t \in [t', t'')$ and that $\alpha^*(t) < 1$ for almost all $t \in [t', t'')$. Then $\alpha^*(t) = 0$ for almost all $t \in [t', \min\{t_A^*, t_B^*\})$.*

Proof. Fix $t' < t'' < \min\{t_A^*, t_B^*\}$ and suppose that $p_A^t > p_B^t$ for all $t \in [t', t'')$ and that $\alpha^*(t) < 1$ for all $t \in [t', t'')$. Let $T_A(t', t'') = \int_{t'}^{t''} \alpha^*(t)dt$ and $T_B(t', t'') = \int_{t'}^{t''} (1 - \alpha^*(t))dt$. By assumption, $T_A(t', t'') < t'' - t'$. Towards a contradiction, assume that $T_A(t', t'') > 0$. Solving the differential equation for the evolution of beliefs yields that, for any $t \in [t', t'')$,

$$p_A^t = \frac{1}{1 + \frac{1-p_A(t')}{p_A(t')} e^{GT_A(t',t)}}, \text{ and}$$
$$p_B^t = \frac{1}{1 + \frac{1-p_B(t')}{p_B(t')} e^{GT_B(t',t)}}.$$

In words, belief p_J^t for depends only on the cumulative experimentation $T_J(t', t)$ on project $J \in \{A, B\}$ up to time t, and not on when this experimentation occurred in the interval $[t', t)$.

Consider an alternative strategy $(\hat{\alpha}, t_A^*, t_B^*)$ such that

$$\hat{\alpha}(t) = \begin{cases} 1 & \text{for all } t \in [t', t'' - T_A(t', t'')), \\ 0 & \text{for all } t \in [t'' - T_A(t', t''), t''), \\ \alpha^*(t) & \text{otherwise.} \end{cases}$$

Then $(\hat{p}_A^{t''}, \hat{p}_B^{t''}) = (p_A^{t''}, p_B^{t''})$, where $(\hat{p}_A^t, \hat{p}_B^t)$ is the belief at time $t \in [t', t'']$ under assignment $\hat{\alpha}$. Hence, the experimenter's payoffs following t'' are the same under (α^*, t_A^*, t_B^*) and $(\hat{\alpha}, t_A^*, t_B^*)$, or $V(\hat{\alpha}, t_A^*, t_B^*; t'') = U(t'')$.

Let τ_{α^*} (respectively $\tau_{\hat{\alpha}}$) be the random arrival time of a success under assignment α^* (respectively $\hat{\alpha}$) conditional on $(p_A^{t'}, p_B^{t'}) = (\hat{p}_A^{t'}, \hat{p}_B^{t'})$. Then $Pr[\tau_{\hat{\alpha}} \leq t] > Pr[\tau_{\alpha^*} \leq t]$ for all $t \in (t', t'')$, while $Pr[\tau_{\hat{\alpha}} \leq t] = Pr[\tau_{\alpha^*} \leq t]$ for all $t \geq t''$. Hence, the distribution of τ_{α^*} first-order dominates the distribution of $\tau_{\hat{\alpha}}$. By discounting, the experimenter's payoff is decreasing in the arrival time of a success, and hence $(\hat{\alpha}, t_A^*, t_B^*)$ yields a strictly higher expected payoff than (α^*, t_A^*, t_B^*) in $[t', t'')$, or

$$\int_{t'}^{t''} [1 - (\overline{k} + \underline{k})] e^{-r\tau_{\hat{\alpha}}} \hat{\mu}(d\tau_{\hat{\alpha}}) > \int_{t'}^{t''} [1 - (\overline{k} + \underline{k})] e^{-r\tau_{\alpha^*}} \mu^*(d\tau_{\alpha^*}),$$

where $\hat{\mu}$ and μ^* are the distributions of $\tau_{\hat{\alpha}}$ and τ_{α^*}, respectively. It follows that

$$\begin{aligned} V(\hat{\alpha}, t_A^*, t_B^*; t') &= \int_{t'}^{t''} [1 - (\overline{k} + \underline{k})] e^{-r\tau_{\hat{\alpha}}} \hat{\mu}(d\tau_{\hat{\alpha}}) + e^{-r(t''-t')} Pr[\tau_{\hat{\alpha}} > t''] U(t'') \\ &> \int_{t'}^{t''} [1 - (\overline{k} + \underline{k})] e^{-r\tau_{\alpha^*}} \mu^*(d\tau_{\alpha^*}) + e^{-r(t''-t')} Pr[\tau_{\alpha^*} > t''] U(t'') \\ &= U(t'), \end{aligned}$$

contradicting the optimality of (α^*, t_A^*, t_B^*). Hence it must be that $\alpha^*(t) = 0$ for almost all $t \in [t', t'')$. This implies that $p_A^{t''} > p_B^{t''}$. Fix any $t''' \in (t'', \min\{t_A^*, t_B^*\})$. If it is the case that $T_A(t'', t''') > 0$, then the previous argument can be applied on the time interval $[t', t''')$ to derive a contradiction to the optimality of (α^*, t_A^*, t_B^*). Hence, we have that $\alpha^*(t) = 0$ for all $t \in [t', \min\{t_A^*, t_B^*\})$ □

The next lemma, mirroring Lemma 2, establishes that 'going-with-the-winner' entails shared experimentation for those beliefs (p_A^t, p_B^t) such that $p_A^t = p_B^t$ and $t < \min\{t_A^*, t_B^*\}$.

Lemma 3. *Fix (p_A^0, p_B^0) such that $p_A^0 \geq p_B^0$ and an optimal strategy (α^*, t_A^*, t_B^*). Suppose that there exists $t' < t'' < \min\{t_A^*, t_B^*\}$ such that $p_A^{t'} = p_B^{t'}$ and $p_A^{t''} = p_B^{t''}$. Then $\alpha^*(t) = \frac{1}{2}$ for almost all $t \in [t', t'')$.*

Proof. Fix $t' < t'' < \min\{t_A^*, t_B^*\}$ such that $p_A^{t'} = p_B^{t'}$ and $p_A^{t''} = p_B^{t''}$, and, towards a contradiction, suppose that there exists $S \subseteq [t', t'')$ with positive Lebesgue measure such that $\alpha^*(t) \neq \frac{1}{2}$ for all $t \in S$. Hence, there exists $t''' \in (t', t'')$ such that $p_A^{t'''} > p_B^{t'''}$. Hence, $T_A(t', t''') < T_B(t', t''')$ and $T_A(t''', t'') > T_B(t'', t''')$, while $T_A(t', t'') = T_B(t', t'')$. Consider an alternative strategy $(\hat{\alpha}, t_A^*, t_B^*)$ such that

$$\hat{\alpha}(t) = \begin{cases} \frac{1}{2} & \text{if } t \in [t', t''), \\ \alpha^*(t) & \text{otherwise.} \end{cases}$$

We have that $V(\hat{\alpha}, t_A^*, t_B^*; t'') = U(t'')$, and following an argument as in Lemma 2, that $V(\hat{\alpha}, t_A^*, t_B^*; t') > U(t')$, yielding the desired contradiction. Hence $\alpha^*(t) = \frac{1}{2}$ for almost all $t \in [t', t'')$. □

Proof of Proposition 1. Fix $t' < t'' < t^*$ and suppose that $p_A^t > p_B^t$ for all $t \in [t', t'')$, and that $\alpha^*(t) < 1$ for all $t \in [t', t'')$. By substituting $t^* = \min\{t_A^*, t_B^*\}$, Lemma 2 applies directly, and it must be that $\alpha^*(t) = 0$ for all $t \in [t', t^*)$. Consider an alternative strategy $(\hat{\alpha}, t^*)$ such that

$$\hat{\alpha}(t) = \begin{cases} 1 & \text{for all } t \in [t', t^*), \\ \alpha^*(t) & \text{otherwise.} \end{cases}$$

Then,

$$\begin{aligned} V(\hat{\alpha}, t^*; t') &= v(p_A^{t'}; C_A(0)) \\ &> v(p_B^{t'}, C_B(0)) \\ &= u(p_A^{t'}, p_B^{t'}), \end{aligned}$$

contradicting the optimality of (α^*, t^*), where, since $t^* = t_A^* = t_B^*$, $C_A(0) = C_B(0)$ is the constant of integration for the optimal stopping problem with a single risky project and experimentation flow cost $\overline{k} + \underline{k}$. Hence, it must be that $\alpha^*(t) = 1$ for almost all $t \in [t', t'')$. The previous argument establishes the 'stay-with-the-winner' property when project A has a strictly higher belief than project B. Fix t' and suppose that $\alpha^*(t) = 1$ for all $t \in [t', t^*)$. Then conditions (4) yield that t^* is such that $p_A^{t^*} = \frac{\overline{k}+\underline{k}}{G}$.

Suppose now that t' is such that $p_A^{t'} = p_B^{t'}$. By Lemma 3, if there exists $t'' \in (t', t^*)$ such that $p_A^{t''} = p_B^{t''}$, then $\alpha^*(t) = \frac{1}{2}$ for all $t \in [t', t'')$. Hence, we can let $t' = \sup\{t : p_A^t = p_B^t\}$. Suppose, towards a contradiction, that $t' < t^*$. Then, for all $t \in (t', t^*)$, $p_A^t > p_B^t$ and hence, by Lemma 2, $\alpha^*(t) = 0$. However, by the previous argument, this contradicts the optimality of (α^*, t^*). Hence, $t' = t^*$, as desired.

To derive the optimal stopping time t^* under shared experimentation, I first obtain an expression for optimal payoffs under shared experimentation. Given t' such that, for all $t \in [t', t^*)$, $p_A^t = p_B^t = p^t$ and that $\alpha^*(p^t, p^t) = \frac{1}{2}$, the optimal payoff $u(p_A^t, p_B^t)$ must satisfy $\frac{\partial}{\partial p_A} u(p_A^t, p_B^t) = \frac{\partial}{\partial p_B} u(p_A^t, p_B^t)$. Define $u_{AB}(p^t) \equiv u(p^t, p^t)$, then it follows that $u'_{AB}(p^t) = 2\frac{\partial}{\partial p_A} u(p^t, p^t)$ and u_{AB} solves

$$u_{AB}(p^t)(r + p^t G) = p^t G - (\overline{k} + \underline{k}) - \frac{1}{2} G p^t (1 - p^t) u'_{AB}(p^t). \tag{10}$$

The differential equation (10) has solution

$$\begin{aligned} v_{AB}(p^t; C_{AB}(0)) &\equiv u_{AB}(p^t) \\ &= C_{AB}(0) \left(\frac{1 - p^t}{p^t}\right)^{\frac{2r}{G}} (1 - p^t)^2 + (p^t)^2 \frac{G - (\overline{k} + \underline{k})}{r + G} \\ &\quad + 2p^t(1 - p^t) \frac{\frac{G}{2} - (\overline{k} + \underline{k})}{r + \frac{G}{2}} - (1 - p^t)^2 \frac{\overline{k} + \underline{k}}{r}. \end{aligned}$$

If optimal experimentation leads from shared experimentation to Q at belief p^{t^*}, the constant of integration $C_{AB}(0)$ and cutoff belief $p^{t^*} = \frac{\overline{k}+\underline{k}}{G}$ are determined by value-matching and smooth-pasting conditions.

$$v_{AB}(p^{t^*}; C_{AB}(0)) = 0, \text{ and}$$
$$v'_{AB}(p^{t^*}; C_{AB}(0)) = 0.$$

The previous results establish that any optimal assignment strategy must be identical to strategy $\alpha^*_{ND}(t)$ for almost all t. What remains to be verified is that, for all t, the value function $V(\alpha^*, t^*; t)$ is a solution to the Bellman equation

$$ru(p^t_A, p^t_B) = \max\Bigg\{0,\ p^t_A G - (\overline{k}+\underline{k}) - \frac{\partial u(p^t_A, p^t_B)}{\partial p_A} G p_A (1-p^t_A) - u(p^t_A, p^t_B) G p^t_A,$$
$$p^t_B G - (\overline{k}+\underline{k}) - \frac{\partial u(p^t_A, p^t_B)}{\partial p_B} G p^t_B (1-p^t_B) - u(p^t_A, p^t_B) G p^t_B \Bigg\}. \tag{11}$$

Suppose that t is such that $p^t_A > p^t_B > \frac{\overline{k}+\underline{k}}{G}$. Then

$$\begin{aligned} V(\alpha^*, t^*; t) &= v(p^t_A; C_A(p^t_B)) \\ &= C_A(p^t_B)\left(\frac{1-p^t_A}{p^t_A}\right)^{\frac{r}{G}}(1-p^t_A) + p^t_A \frac{G-(\overline{k}+\underline{k})}{r+G} - (1-p^t_A)\frac{\overline{k}+\underline{k}}{r}, \end{aligned}$$

with the constant of integration determined at the switch to shared experimentation at time t' such that $p^{t'}_A = p^{t'}_B$. Since $V(\alpha^*, t^*; t) > V(\alpha^*, t^*; t') > 0$, it is optimal to continue experimentation for all $t > t'$. The derivative of the third term of (11) with respect to p_A is

$$\frac{\partial}{\partial p_A}\left[\frac{\partial}{\partial p_B} C_A(p^t_B)\left(\frac{1-p^t_A}{p^t_A}\right)^{\frac{r}{G}}(1-p^t_A) - v(p^t_A; C_A(p^t_B)) G p^t_B\right] < 0,$$

since $C_A(p^t_B)$ is independent of p_A and $v(p^t_A; C_A(p^t_B))$ is increasing in p_A. This last fact also implies that the second term of (11) is increasing in p_A. If $t \approx t'$, the difference in these two terms is given by

$$G p^{t'}_B (1-p^{t'}_B)\left[\frac{\partial v(p^t_A; C_A(p^t_B))}{\partial p_A} - \frac{\partial v(p^t_A; C_A(p^t_B))}{\partial p_B}\right]\Big|_{p^{t'}_A = p^{t'}_B} = 0,$$

by smooth-pasting since at $(p^{t'}_A, p^{t'}_B)$ a transition occurs to shared experimentation. Hence, the second term of (11) is larger than the third term for all $t > t'$, as required. If t is such that $p^t_A > \frac{\overline{k}+\underline{k}}{G} > p^t_B$, then $V(\alpha^*, t^*; t) = v(p^t_A; C_A(0))$, while if t is such that $p^t_A = p^t_B > \frac{\overline{k}+\underline{k}}{G}$, then $V(\alpha^*, t^*; t) = v_{AB}(p^t_A; C_{AB}(0))$. In both these cases, it can be verified that value-matching and smooth-pasting imply that $V(\alpha^*, t^*; t) > 0$, as desired.

□

Proof of Proposition 2. For part (i), suppose that $t_A^* < t_B^*$. Recall that, without loss of generality, we can restrict attention to assignments such that, for all t, $p_A^t \geq p_B^t$. If $p_A^{t_A^*} > p_B^{t_A^*}$, then consider an alternative strategy $(\hat{\alpha}, \hat{t}_A, \hat{t}_B)$ such that $(\hat{t}_A, \hat{t}_B) = (t_B^*, t_A^*)$ and

$$\hat{\alpha}(t) = \begin{cases} 1 & \text{for all } t \in [t_A^*, t_B^*), \\ \alpha^*(t) & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} U(t_A^*) &= v(p_A^{t_B^*}; C_A(0)) \\ &< v(p_A^{t_A^*}; C_A(0)) \\ &= V(\alpha^*, \hat{t}_A, \hat{t}_B; t_A^*), \end{aligned}$$

contradicting the optimality of (α^*, t_B^*, t_A^*). If instead $p_A^{t_A^*} = p_B^{t_A^*}$, then the strategy $(\hat{\alpha}, \hat{t}_A, \hat{t}_B)$ yields the same payoffs to the experimenter as (α^*, t_A^*, t_B^*). Hence, $t_A^* \geq t_B^*$ without loss of generality.

For part (ii), suppose that $t_B^* > \tilde{t} = \inf\{t : \alpha^*(s) \in \{1, Q\} \text{ for almost all } s \geq t\}$. Consider an alternative strategy $(\alpha^*, t_A^*, \tilde{t})$. The difference between the experimenter's payoffs under (α^*, t_A^*, t_B^*) and $(\alpha^*, t_A^*, \tilde{t})$ corresponds to the additional expected costs incurred while maintaining project B from $\tilde{t}$ to t_B^*, and we have that

$$\begin{aligned} V(\alpha^*, t_A^*, \tilde{t}; \tilde{t}) - U(\tilde{t}) &\equiv c(\tilde{t}) \\ &= C \left(\frac{1 - p_B^{\tilde{t}}}{p_B^{\tilde{t}}} \right)^{\frac{r}{G}} (1 - p_B^{\tilde{t}}) + p_B^{\tilde{t}} \frac{\underline{k}}{r + G} + (1 - p_B^{\tilde{t}}) \frac{\underline{k}}{r} \\ &> 0, \end{aligned}$$

where the constant of integration $C < 0$ is determined by $c(t_B^*) = 0$. This contradicts the optimality of (α^*, t_A^*, t_B^*). Hence, $t_B^* = \tilde{t}$.

Part (iii) follows from directly from Lemma 2 since, by part (i), $t_B^* = \min\{t_A^*, t_B^*\}$. □

Proof of Proposition 3. From (2), we can rewrite

$$z(p_A) = \frac{1}{p_A G + r} \left[\underline{k} - \frac{\partial}{\partial p_A} v(p_A, C_A(0)) p_A (1 - p_A) \right].$$

Let $\hat{p}_A(\overline{k}, G, r) \in \arg\max_{p_A \in [\frac{\overline{k} + \underline{k}}{G}, 1]} \frac{\partial}{\partial p_A} v(p_A, C_A(0)) p_A (1 - p_A)$, and let $\underline{k}^*(\overline{k}, G, r)$ be such that

$$\underline{k}^*(\overline{k}, G, r) = \frac{\partial}{\partial p_A} v(\hat{p}_A(\overline{k}, G, r), C_A(0)) \hat{p}_A(\overline{k}, G, r)(1 - \hat{p}_A(\overline{k}, G, r)).$$

To prove part (i), note that if $\underline{k} \geq \underline{k}^*(\overline{k}, G, r)$, then $z(p_A) \geq 0$ for all p_A. Otherwise, the condition of part (ii) holds and there exists p_A such that $z(p_A) < 0$, and $\underline{p}_A$ and $\overline{p}_A$ are the only two solutions to $z(p_A) = 0$ by the convexity of z.

When the conditions of part (ii) obtain, a first claim is that if $p_A^0 = p_B^0 = \overline{p}_A$, then $t_B^* = 0$. In particular, discarding project B at time 0 is strictly preferred to shared experimentation. By an argument as in the proof of Proposition 1, if $t_B^* > 0$ then there exists t' such that, for almost all $t \in [0, t')$, $\alpha^*(t) = \frac{1}{2}$ and $p_A^{t'} = p_B^{t'}$. The experimenter's payoffs are given by $U(0) = v_{AB}(\overline{p}_A; C_{AB}(p_A^{t'}))$, where by (10),

$$\begin{aligned} v_{AB}(\overline{p}_A; C_{AB}(p_A^{t'})) &= \frac{\overline{p}_A G - (\overline{k} + \underline{k})}{(r + \overline{p}_A G)} - \frac{G\overline{p}_A(1 - \overline{p}_A)}{2(r + \overline{p}_A G)} v'_{AB}(\overline{p}_A; C_{AB}(p_A^{t'})) \\ &< \frac{\overline{p}_A G - (\overline{k} + \underline{k})}{(r + \overline{p}_A G)} \\ &= v(\overline{p}_A; C_A(0)). \end{aligned}$$

Hence, by continuity, for initial beliefs (p_A^0, p_B^0) with $p_A^0 = p_B^0 < \overline{p}_A$ sufficiently close to $\overline{p}_A$, discarding B at time 0 is strictly preferred to shared experimentation. A very similar argument shows that discarding B at time 0 is strictly preferred to using project A and maintaining project B for an open set of states of initial beliefs (p_A^0, p_B^0) with $p_A^0 > p_B^0$ and $p_A^0 < \overline{p}_A$ sufficiently close to $(\overline{p}_A, \overline{p}_A)$. However, within the discarding boundary using project B at time 0 until time t_B^* is preferred to discarding it and hence there exists an open set of initial beliefs (p_A^0, p_B^0) with $p_A^0 < \overline{p}_A$ around $(\overline{p}_A, \overline{p}_A)$ in which using project B is optimal. A very similar argument demonstrates the same result for a region of initial beliefs around $(\underline{p}_A, \underline{p}_A)$.

□

Proof of Lemma 1. By Proposition 3, there exists a belief $\hat{p} > \underline{p}_A$ such that $\alpha^*(t) = 0$ for almost all (p_A^t, p_B^t) such that $p_A^t = p_B^t$ and $p_A^t \in [\underline{p}_A, \hat{p}]$. Suppose that $p_A^t = p_B^t$ and $\alpha^*(t) = \frac{1}{2}$ for almost all $t \in [t', t'')$ and let $t''' = \inf\left\{t : \alpha^*(s) \neq \frac{1}{2} \text{ for almost all } s \geq t\right\}$. That is, the experimenter switches from shared experimentation to project B at belief $p_A^{t'''} = p_B^{t'''} \geq \hat{p}$. Hence the smooth-pasting condition at belief $p^{t'''}$ is

$$\begin{aligned} \frac{\partial}{\partial p_B} u(p_A^{t'''}, p_B^{t'''}) &= \frac{1}{2} \cdot \frac{\partial}{\partial p} v_{AB}(p_B^{t'''}; C_{AB}(p_B^{t'''})) \\ &= \frac{\partial}{\partial p_B} v_B(p_B^{t'''}, C_B(p_B^{t'''})), \end{aligned}$$

which, with manipulations, yields that

$$C_{AB}(p_B^{t'''}) = \frac{C_B(p_B^{t'''})}{\left(\frac{1-p}{p}\right)^{\frac{r}{G}}(1-p)} + \frac{p_B^{t'''}\left[\frac{G-(\overline{k}+\underline{k})}{r+G} - \frac{\frac{G}{2}-(\overline{k}+\underline{k})}{r+\frac{G}{2}}\right] + (1-p_B^{t'''})\left[\frac{\frac{G}{2}-(\overline{k}+\underline{k})}{r+\frac{G}{2}} - \frac{\overline{k}+\underline{k}}{r}\right] - \frac{G(r+\overline{k}+\underline{k})}{r(r+G)}}{\left(\frac{1-p_B^{t'''}}{p_B^{t'''}}\right)^{\frac{2r}{G}+1}\frac{Gp_B^{t'''}+r}{G}}.$$

Meanwhile, the value matching condition is

$$v_{AB}(p_B^{t'''}; C_{AB}(p_B^{t'''})) = v_B(p_B^{t'''}, C_B(p_B^{t'''})),$$

which, with manipulations, yields that

$$C_{AB}(p_B^{t'''}) = \frac{C_B(p_B^{t'''})}{\left(\frac{1-p_B^{t'''}}{p_B^{t'''}}\right)^{\frac{r}{G}}(1-p_B^{t'''})} - \frac{\frac{G^2}{2}(r+\overline{k}+\underline{k})}{\left(\frac{1-p_B^{t'''}}{p_B^{t'''}}\right)^{\frac{2r}{G}+1} r(r+G)(r+\frac{G}{2})}.$$

Together, these yield that

$$p_B^{t'''} = \frac{2(\overline{k}+\underline{k})(r+G)(r+\frac{G}{2})}{2(\overline{k}+\underline{k})(r+G)(r+\frac{G}{2}) + \frac{G^2}{2}(\overline{k}+\underline{k}+r)}, \tag{12}$$

which is uniquely defined and such that $p_B^{t'''} \in [0,1]$. Define $\underline{p}$ to be the unique solution to (12). □

Proof of Proposition 4. First, I construct the sets $\mathcal{P}_L$ and $\mathcal{P}_W$. In the following, assume that the conditions of Lemma 1 are met and that there exists a (unique) portion of the 45-degree line $(\underline{p}, \overline{p})$ for which shared experimentation is optimal if $p_A^t = p_B^t \in (\underline{p}, \overline{p})$. The arguments that follow apply in a straightforward way when this is not the case.

First, let

$$\mathcal{P}_L^1 = \left\{(p_A, p_B) \in \mathcal{P}_M : p_A \in [\underline{p}_A, \underline{p}] \cup [\overline{p}, \overline{p}_A]\right\}.$$

By Proposition 2, it must be that given $p_A^{t'} = p_B^{t'} \in [\underline{p}_A, \underline{p}] \cup [\overline{p}, \overline{p}_A]$, $\alpha^*(t) = 0$ for almost all $t \in [t', t_B^*]$, where $p_B^{t_B^*} = p_B^*(p_A^{t'})$. That is, from beliefs on the 45-degree line, if there is no shared experimentation then project B enters a culling phase.

Second, consider

$$\mathcal{P}_M \setminus \mathcal{P}_L^1 = \left\{(p_A, p_B) \in \mathcal{P}_M : p_A \in (\underline{p}, \overline{p}), p_B \in [p_B^*(\underline{p}), p_B^*(\overline{p})]\right\},$$

the set of beliefs in the maintenance region that have not been attributed to $\mathcal{P}_L^1$. By Proposition 2, from beliefs $(p_A^t, p_B^t) \in \mathcal{P}_M \setminus \mathcal{P}_L^1$, an optimal policy will either (a) put project B to trial immediately until it is discarded, or (b) put project A to trial until time t' at which either (bi) $p_A^{t'} = p_B^t$ and shared experimentation ensues, or (bii) $p_A^{t'} > p_B^t$ and a culling phase for project B ensues. Define $v(p_A^t; C_A(p_A^{t'}, p_B^t))$ to be the payoff to the experimenter at t were it to put project A to trial until t' with belief $p_A^{t'} \in [\max\{p_B^t, \underline{p}\}, p_A^t)$, and then switch to project B until time t'' with discarding belief $p_B^{t''} = p_B^*(p_A^{t'})$. Hence the constant of integration $C_A(p_A^{t'}, p_B^t)$ depends on the belief p_B^t and on the switching belief $p_A^{t'}$, but not on p_A^t. Similarly, if $p_B^t > \underline{p}$, define $v(p_A^t; \tilde{C}_A(p_B^t))$ to be the payoff to the experimenter at t were it to put project A to trial until t' at which $p_A^{t'} = p_B^t$, followed by shared experimentation until t'' at which $p_A^{t''} = p_B^{t''} = \underline{p}$. If $p_B^t \leq \underline{p}$, then define $v(p_A^t; \tilde{C}_A(p_B^t)) = v(p_A^t; C_A(\underline{p}, p_B^t))$. Note that $v(p_A^t; C_A(p_A^{t'}, p_B^t)) \geq v(p_A^t; C_A(p_A^{t''}, p_B^t))$ if and only if $C_A(p_A^{t'}, p_B^t) \geq C_A(p_A^{t''}, p_B^t)$ and $v(p_A^t; C_A(p_A^{t'}, p_B^t)) \geq v(p_A^t; \tilde{C}_A(p_B^t))$ if and only if $C_A(p_A^{t'}, p_B^t) \geq \tilde{C}_A(p_B^t)$. Hence if $C_A(p_A^t, p_B^t) = \max_{\{p_A^{t'} \in [\max\{p_B^t, \underline{p}\}, p_A^t]\}} C_A(p_A^{t'}, p_B^t)$, then the experimenter has no incentive to put project A to trial. If, on the other hand, there exists a $p_A^{t'}$ such that $C_A(p_A^{t'}, p_B^t) > C_A(p_A^t, p_B^t)$, then the experimenter gains by staying with project A until belief $p_A^{t'}$. Let

$$\mathcal{P}_L^2 = \left\{(p_A, p_B) \in \mathcal{P}_M \setminus \mathcal{P}_L^1 : \max\left\{\max_{p'_A \in [\max\{p_B, \underline{p}\}, p_A]} C_A(p'_A, p_B), \tilde{C}_A(p_B)\right\} \leq C_A(p_A, p_B)\right\},$$

and let $\mathcal{P}_L = \mathcal{P}_L^1 \cup \mathcal{P}_L^2$. Finally, let $\mathcal{P}_W^1 = \mathcal{P}_M \setminus \mathcal{P}_L$. Hence, all the beliefs in $\mathcal{P}_M$ have been attributed either to $\mathcal{P}_L$ or to $\mathcal{P}_W^1$.

Third, fix t' and beliefs $(p_A^{t'}, p_B^{t'}) \in \mathcal{P}_D$. By proposition 3, either $t_B^* = t'$ or $\alpha^*(s) = 1$ for almost all $t \in [t', t'')$, where t'' is such that $p_A^{t''} = {p_B^*}^{-1}(p_B^{t'})$. Let

$$\mathcal{R} = \left\{(p_A, p_B) \in \mathcal{P}_D : ({p_B^*}^{-1}(p_B), p_B) \in \mathcal{P}_W^1\right\}.$$

That is, $\mathcal{Q}$ is the set of beliefs in the discarding region such that were A to be used and B maintained until the discarding boundary, B would also be maintained when beliefs cross into $\mathcal{P}_M$. By Proposition 2, it must be that $t_B^* = t'$ for all $(p_A^{t'}, p_B^{t'}) \in \mathcal{P}_D \cap \mathcal{R}^c$. For any $(p_A, p_B) \in \mathcal{R}$, define $\tilde{p}_A(p_B) = \sup\{p_A : (p_A, p_B) \in \mathcal{P}_W^1\}$. Furthermore, define

$$\mathcal{P}_W^2 = \left\{(p_A, p_B) \in \mathcal{R} : v(p_A; C_A(\tilde{p}_A(p_B), p_B)) > v(p_A; C_A(0))\right\}.$$

Finally, let $\mathcal{P}_W = \mathcal{P}_W^1 \cup \mathcal{P}_W^2$.

The previous results establish that any optimal assignment strategy must be identical to strategy $\alpha_D^*(t)$ for almost all t. What remains is to verify that the strategy $(\alpha_D^*, t_A^*, t_B^*)$ is optimal, which follows from an argument very similar to that for Proposition 1 and is omitted. □

Proof of Proposition 5. Part (i) follows from an argument almost identical to that of Lemma 2. For part (ii), suppose that $\alpha_B^*(t) > 0$ for almost all $t \in [t', t'')$, but that, towards a contradiction,

there exists $S \subseteq [t', t_B^*)$ with positive Lebesgue measure such that $\alpha_A^*(t) > 0$ for all $t \in S$. Following the proof of Lemma 2, let $T_A(t', t_B^*) = \int_{t'}^{t_B^*} \alpha_A^*(t)dt$, and consider the alternative strategy $(\hat{\alpha}, t_A^*, t_B^*)$ such that

$$\hat{\alpha}(t) = \begin{cases} (1,0,0) & \text{if } t \in [t', t_B^*),\ \alpha_C^*(t) = 0 \text{ and } t - t' - \int_{t'}^{t} \alpha_C^*(t)dt < T_A(t', t_B^*), \\ (0,1,0) & \text{if } t \in [t', t_B^*),\ \alpha_C^*(t) = 0 \text{ and } t - t' - \int_{t'}^{t} \alpha_C^*(t)dt \geq T_A(t', t_B^*), \\ \alpha^*(t) & \text{otherwise.} \end{cases}$$

By part (i), we have that $\alpha_C^*(t) = 1$ for almost all $t \in [t', t_C^*)$ such that $\alpha_C^*(t) > 0$. Hence, $V(\hat{\alpha}, t_A^*, t_B^*, t_C^*; t_B^*) = U(t_B^*)$, and, again by an argument almost identical to that of Lemma 2, $V(\hat{\alpha}, t_A^*, t_B^*, t_C^*; t') > U(t')$, yielding the desired contradiction. Hence, $\alpha_A^*(t) = 0$ for all $t \in [t', t_B^*)$. □