Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Managing Multiple Research Projects

Alejandro Francetich

Working Paper n. 516

This Version: March 28, 2015

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

Managing Multiple Research Projects*

Alejandro Francetich †‡

Department of Decision Sciences and IGIER, Bocconi University, Italy

March 28, 2015

Abstract

A decision maker can experiment on up to two tasks or projects *simultaneously* over time. One and only one of these projects can produce successes, according to a Poisson process with known arrival rate; but there is uncertainty as to which project is the profitable one. The decision maker only observes the outcomes of the projects undertaken, and undertaking each project entails a cost. Simultaneous experimentation involves higher costs but can produce more data. At the same time, since the projects are correlated, the outcomes of either one are informative about the other. If the costs are high and she is sufficiently impatient, the decision maker never experiments on both projects at once. Otherwise, if she starts working on a single project that produces no successes, she becomes gradually pessimistic and eventually takes on the other project *while keeping the first one* — despite the higher costs and the negative correlation.

Keywords: Experimentation, two-armed bandits, multi-choice bandits, negatively correlated arms, Poisson process

JEL Classification Numbers: D83, D90

*A previous version of this paper has circulated under the title "Experimentation With Menus."

†Via Röntgen 1, 20136 Milan, Italy. Email address: alejandro.francetich@unibocconi.it

‡This paper borrows inspiration from Chapter 3 of my dissertation, submitted to the Stanford Graduate School of Business. I am deeply indebted to David Kreps for continuously enlightening me, and to Alejandro Manelli, Pierpaolo Battigali, and Massimo Marinacci for their support and guidance. I am grateful to Sven Rady for his encouragement and his insightful and helpful comments. The paper has also benefited from comments by Amanda Friedenberg, Marco Ottaviani, attendants of the 2014 Canadian Economic Theory Conference and of the 25th International Game Theory Conference at Stony Brook, seminar participants at UC Davis, UW at Bothell, University of Sheffield, and Universidad Carlos III de Madrid, and anonymous referees. I gratefully acknowledge financial support from ERC advanced grant 324219. Any remaining errors and omissions are all mine.

1 Introduction

Consider the following treasure-hunting problem. There is a cluster of islands, in one (and only one) of which a treasure is buried. The problem is to discover which of the islands hides this treasure, and also where exactly in the island, or how deep underground, the treasure is buried. This is a standard search problem. But now imagine that an explorer can organize multiple expeditions to multiple locations at once. There is a cost to exploring each island, but simultaneous expeditions can cover more ground faster.

This version of the treasure-hunting problem has at least two different applications: research and team recruiting.

A researcher has a conjecture for a new theoretical result. If the conjecture is true, and if she can provide a proof, she can write a paper with her result. If the conjecture is false, and she produces a counterexample, she can write a different paper with the “modified conjecture,” or perhaps a (shorter) paper with the counterexample. She can divide her time between trying to come up with a proof and trying to come up with a counterexample; but she can also use part of her research funds to hire a research assistant to work on the counterexample *while* she works on the proof, or viceversa. Here, the two tasks represent the different islands, while the option of simultaneous expeditions is represented by hiring the assistant.

Alternatively, a lab is conducting clinical research on different new treatments for a disease. Each of the treatments is based on different hypotheses regarding the cause of the disease. The lab director can have her staff experiment on one treatment, or she can hire additional researchers and have different teams working side by side on different treatments. This way, the successful treatment may be identified faster; but the additional researchers must be compensated for their work. Finding the treasure here is identifying the right treatment, and the cost of multiple expeditions is the cost of hiring the additional researchers.

Similarly, a principal is recruiting agents from a pool of job applicants. The different agents come from opposing schools or training, and there is uncertainty about which of their profiles is better suited to the company’s operation. The principal is interested in identifying the “star candidate.” To get data on their performance, the candidates have to be hired. The firm can hire one candidate at a time, or can hire multiple candidates at once. The different agents are represented by the different islands; the “treasure” is the star applicant; and the cost of exploring is the agents’ wage.

The present paper studies such experimentation problems, where a decision maker faces a finite set of tasks or projects (research projects, treatments, job applicants). Successes arrive over time according to a known Poisson process, but there is uncertainty as to which one of the projects can produce these successes. One and only one of the projects is fruitful. Thus, the projects are negatively correlated: A success from one is conclusive evidence that the others cannot produce value. Over any time interval, the decision maker can choose any *subset* of projects to work on, including the empty set (namely, not working at all). She bears a (constant) flow cost for each project undertaken, and she only observes successes *for those projects selected*.

While the problem can be formulated for an arbitrary number of finitely many projects, at such level of generality it becomes intractable. Thus, most of the analysis in this paper focuses on the case of two projects (as in the example of academic research).

The basic results when *simultaneous experimentation* is allowed are summarized as follows:

(a) If the costs of the projects are low and/or the arrival rates are sufficiently high relative to the discount rate, the decision maker begins with both projects at once when her prior is diffuse. If, instead, her prior assesses that one project is sufficiently likely to be the fruitful one, she begins with that project alone. If enough time passes and she meets with no success, she takes on *both projects simultaneously* once she becomes sufficiently unsure about the state; she never abandons one project for the other.

(b) But if the costs of the projects are high, and if the discount rate is high (or the arrival rate is low), she either does no research at all — if her prior is sufficiently diffuse — or she works on one project only — if the prior that said project is the fruitful project is high enough —, abandoning research if, after a while, she does not meet with success.

(c) Imagine the researcher, if she conducts both projects at once and success is achieved, cannot tell which project was responsible for the success. The manager of a consulting firm may observe whether a team of experts collectively meets their clients' needs, but not exactly how much each of the experts contributes individually. For low costs and high discount rate, the decision maker starts with both projects if her prior is diffuse. Now, however, if she starts on a singleton, she sticks to the singleton for longer. If the costs are too high relative to the arrival rate, or if the discount rate is sufficiently low, she only experiments on singletons: Information can only come from singletons, which are cheaper than simultaneous experimentation.

(d) Of the two basic projects, one must eventually succeed; in this sense, they are "collectively safe." Sometimes, a researcher may also have other, separate projects to work on, projects that may fail. The decision maker postpones starting on these collectively-safe projects in favor of a riskier one if she is sufficiently optimistic about this third project, the less optimism required the more uncertainty there is about the two original projects.

The decision maker experiments simultaneously on both projects when she is sufficiently uncertain about the state. In this sense, there is "more experimentation" for mid-range beliefs, when information is the most valuable. This stands in (apparent) contrast with Moscarini and Smith (2001), who find that experimentation "accelerates" as the decision maker becomes more confident.

Moscarini and Smith (2001) assume that the cost of "experimentation" — in their model, buying signals and delaying final, irreversible choices — is strictly increasing and strictly convex. Moreover, observations and posteriors obey a diffusion process, so they always change gradually over time. Thus, experimentation is more costly when it takes longer for the posterior to reach decision thresholds. The same is true in the present paper: Experimentation is more costly when it has the smallest impact on beliefs. However, the flow cost of undertaking each project is constant, and observing an arrival from a Poisson process produces jumps in the posterior rather than gradual changes; thus, experimentation has the smallest impact on beliefs when the posterior is already close to the extremes.

While Moscarini and Smith (2001) represent experimentation as a type of Wald sequential hypothesis-testing problem, I represent experimentation as a bandit problem — here, a Poisson bandit with correlated arms.[1] Not being restricted to choosing at most a single arm at a time, it is in fact a *multi-choice* bandit problem. A special class of such problems is studied in Bergemann and Välimäki (2001). A decision maker faces countably many arms, and can choose up to some fixed number of them at a time, at no additional cost. A generalization of the Gittins index[2] applies if the arms are independent, ex-ante identical, and there are (countably) infinitely many of them; however, Bergemann and Välimäki (2001) show by example that this solution fails if there are only finitely many arms.

In Francetich and Kreps (2014), we study the following variation of the present multi-choice bandit problem. A finite set X of "tools" is given. Each time period $t = 0, 1, \ldots$, a decision maker chooses a "toolkit" $K_t \subseteq X$ to carry for that period. Each tool $x \in X$ has a "rental" cost $c_x > 0$ and value on date t given by $v_t(x)$, where the process $\{v_t \in R_+^X\}_{t \in \mathbb{N}}$ is independent and identically distributed according to some unknown distribution. On each date t, the decision maker only observes the values $v_t(x)$ for those x that are in the toolkit she has selected, K_t. As a bandit problem with non-independent arms, we cannot enlist the Gittins index, and the problem can become intractable. At this level of generality, the best we can hope for — aside from asymptotic results —, and the subject of Francetich and Kreps (2014), is to investigate the performance of various decision heuristics. We borrow from the machine-learning literature in computer science and operations research, which is concerned with developing algorithms that "perform well" in bandit problems.[3] But one can imagine special and restricted formulations of this problem that are amenable to analytical solution, and the present paper provides one such formulation; this allows us to build up our intuition regarding solutions to the more general problem.

While the spirit of the problem studied in this paper is closely related to Francetich and Kreps (2014), the formal techniques employed borrow heavily from Keller and Rady (2010) and Klein and Rady (2011), who study *strategic* experimentation with Poisson and exponential bandits, respectively. In Keller and Rady (2010), each player has an identical copy of a bandit with one risky arm and one safe arm; in Klein and Rady (2011), the risky arms of each player's bandit are negatively correlated *across players*. These players can choose only one arm at a time, but they can learn from each other. Like the decision maker in the present paper, they have more than one source of information. However, to them, this "extra" information is public and free; our decision maker can only exploit her additional source of information — choosing more than one

[1]In terms of behavior, the Wald approach decouples payoffs and learning, but makes decisions irreversible. In the bandit approach, choices yield both information and payoffs (assessed on the basis of said information), and such choices are typically reversible. I consider some very special forms of irreversibility in sections 4 and 6.

[2]See, for instance, Gittins and Jones (1974); Whittle (1980); Weber (1992).

[3]References to this literature are provided in Francetich and Kreps (2014). Part of this literature provides algorithms even for problems where the arms of the bandit are statistically independent under the prior, so that the Gittins index can be applied; this is because the computation of the index is typically complex as a practical matter.

arm at a time— by means of payoff-relevant actions, so she faces a different trade-off.

Nonetheless, the baseline problem of our decision maker can be mapped to the benchmark problem of the social planner in Klein and Rady (2011). From the point of view of this planner, our two projects are the two agents; our cost of undertaking the projects is the opportunity cost of having the agents neglect their safe arms; and, since actions and outcomes are public, simultaneous experimentation amounts to assigning both agents to their corresponding risky arms. Thus, the results in section 3 have exact parallels to results in Klein and Rady (2011), and the analysis in the later sections constitutes an extension of their efficiency benchmark.

The rest of the paper is organized as follows. Section 2 describes the basics of the formal framework. Section 3 analyzes the optimal research strategy. Section 4 analyzes the case when choices must be nested, while section 5 discusses the case of "imperfect monitoring" of successes, namely, when successes from simultaneous experimentation cannot be attributed to individual projects. Section 6 discusses the problem of more than two projects. Finally, section 7 concludes. Proofs are relegated to the appendix, as is the extension of the analysis in section 3 to the case of asymmetric research costs.

2 The Model

There is a finite set of n tasks or "projects" $X = \{x_0, \ldots, x_{n-1}\}$ on which a decision maker (henceforth, DM) can experiment. The DM allocates her time between the different subsets of X, representing research agendas. The set of allocations of a divisible unit of time between the subsets of X is $\mathcal{A} := S^{2^n-1}$, the $(2^n - 1)$-dimensional simplex. Given some labelling of the elements of the power set of X, $2^X = \{A_j \subseteq X : j = 0, \ldots, 2^n - 1\}$, the j-th component α_j of vector $\alpha \in \mathcal{A}$ denotes the fraction of time spent on A_j.

There is a (flow) research cost $c > 0$ to undertaking each project; this cost represents, for instance, wages or fees. Successes yield a gross reward of 1, and they arrive over time for project $i = 0, \ldots, n-1$ according to a Poisson processes with arrival rate $\lambda I(\omega = \omega_i)$, where $I(\cdot)$ is the indicator function, $\lambda > c$ is the *known* arrival rate, and $\omega \in \Omega := \{\omega_0, \ldots, \omega_{n-1}\}$ is the ex-ante *unobserved* state of nature. In words, it is known that one and only one of these projects is profitable to undertake, and exactly how profitable it is, but there is uncertainty as to which one is the profitable one.[4] Payoffs are discounted at the subjective rate $\rho > 0$.

The DM starts with a prior π^0 over the states of nature; this prior is a point in $\Pi := S^{n-1}$, the $(n-1)$-dimensional simplex. If $\pi \in \Pi$ represents the beliefs of the DM, her expected immediate

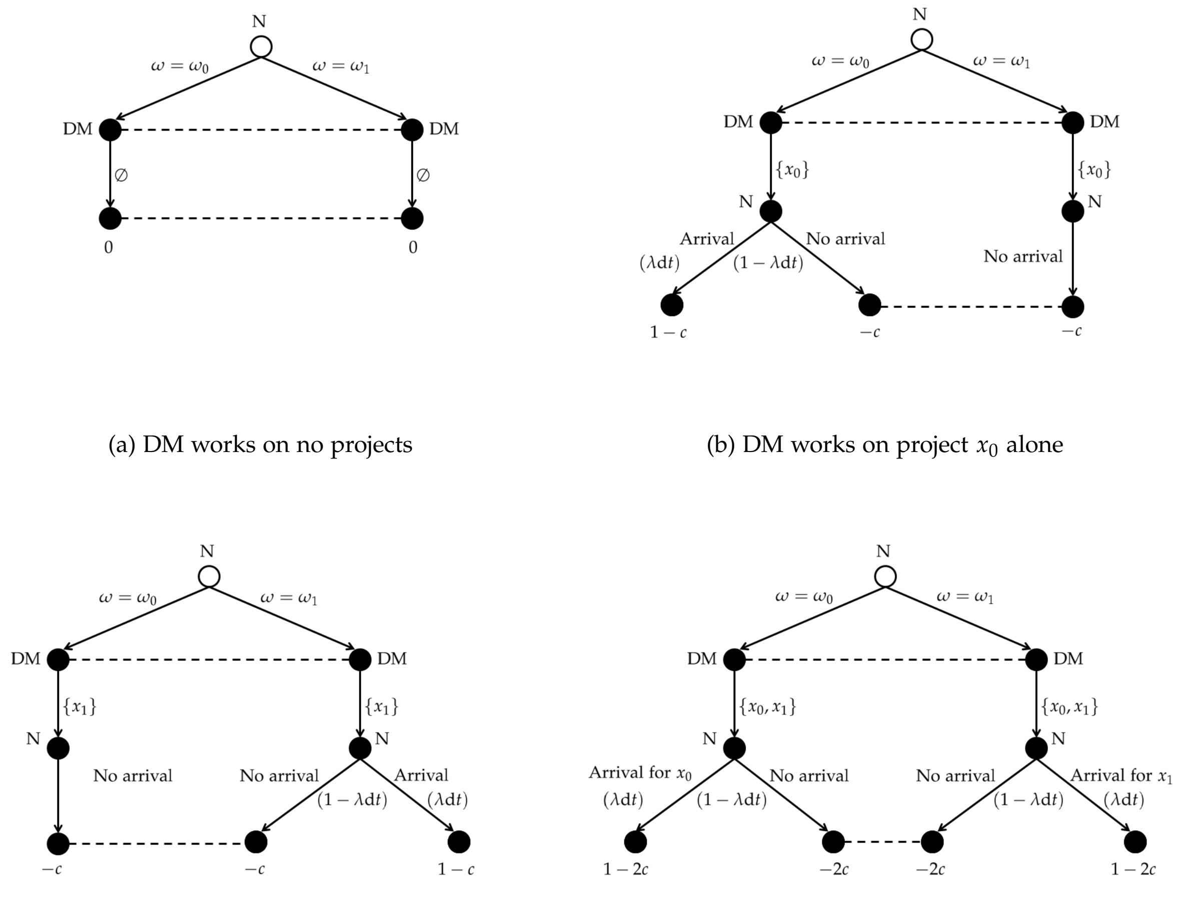
[4]A more flexible specification would allow for $\Omega = S^{n-1}$ and for arrival rates given by $\lambda\omega_i$ for project i, where $(\omega_0, \ldots, \omega_{n-1})$ is now a vector of probabilities. Under this alternative specification, successes arrive at rate λ and are "allocated" to project x_i with probability ω_i, independently of past arrivals and allocations; this yields a partitioning of the Poisson process of successes. But this additional flexibility comes at the cost of slowing down the learning process, without providing significant new insights. An arrival for a project ceases to be conclusive evidence that the project is the superior one. Instead, we would assess a project to be superior by observing a sufficiently larger frequency of arrivals for it relative to the other projects; a single observation of success no longer suffices.

payoff from experimenting on subset $A \subseteq X$ for a time interval of length $\Delta > 0$ is:

$$\lambda\Delta \sum_{i=0}^{n-1} \pi_i I(x_i \in A) - c\Delta\#A.$$

In addition, she observes whether any successes arrive over Δ for each of the projects $x \in A$. In particular, by working on a single project, she cannot distinguish between the event of an arrival for one of the other projects and the event of failure of arrival altogether. Figure 1 summarizes ex-post payoffs and data collected under each of the possible choices in the case $n = 2$.

Let $\pi_t = (\pi_{0,t}, \ldots, \pi_{n-1,t})$ denote the period-t posterior. At any moment, observing an arrival makes the posterior jump to 1 for the successful project and to 0 for the rest. By spending time on all projects, either nothing new is learned, or the model uncertainty is resolved immediately.



(a) DM works on no projects

(b) DM works on project x_0 alone

(c) DM works on project x_1 alone

(d) DM works on both projects

Figure 1: DM's observations and payoffs under each of her possible choices.

This is due to the symmetry in arrival rates; the event of failure of arrival is equally likely for all of the projects. This would not be the case if different projects had different arrival rates in their corresponding state: Failure of arrival over a given interval of time would speak louder about more productive projects than about less productive ones.[5]

The more interesting dynamics take place when the DM spends time on non-empty proper subsets of X, namely, when she works on some but not all projects. Given $\alpha \in \mathcal{A}$, let $\overline{\alpha}^i$ denote the fraction of time spent on project i, be it exclusively or as part of a larger set of projects: $\overline{\alpha}^i = \sum_{j:x_i \in A_j} \alpha_j$. If no arrival results over $[t, t+\Delta t)$, the posterior for project x_i is:

$$\pi_{i,t+\Delta t} = \frac{\pi_{i,t} e^{-\overline{\alpha}^i \lambda \Delta t}}{\pi_t e^{-\overline{\alpha}^i \lambda \Delta t} + 1 - \pi_t}.$$

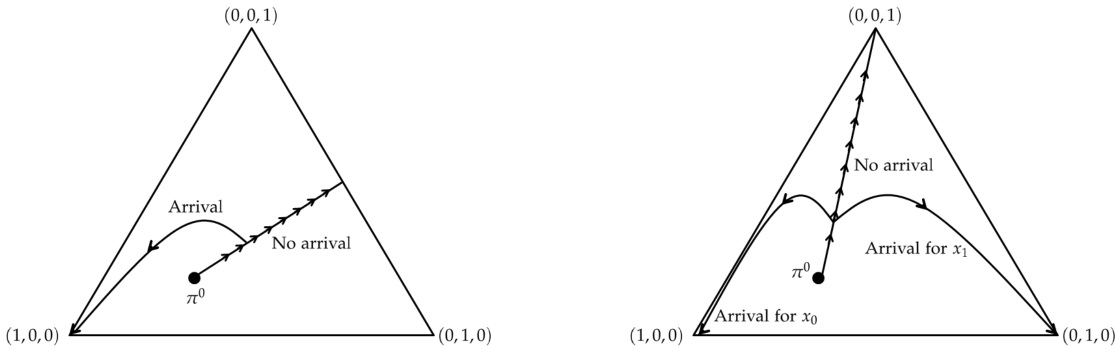
As Δt shrinks, we obtain:

$$\dot{\pi}_{i,t} = -\overline{\alpha}^i \lambda \pi_{i,t}(1 - \pi_{i,t}).$$

While working unsuccessfully on some but not all projects, the DM becomes progressively pessimistic about them while progressively optimistic about the neglected ones. See Figure 2.

The environment is stationary, and the state variable of the problem is the belief of the DM, $\pi \in \Pi$. Let $w : \Pi \to \mathbb{R}$ denote the (optimal, average) value function; w satisfies the Bellman equation:

$$w(\pi) = \max_{\alpha \in \mathcal{A}} \left\{ \sum_{j=0}^{2^n-1} \alpha_j \left(\lambda \Delta \sum_{i=0}^{n-1} \pi_i I(x_i \in A_j) - c\Delta \# A_j + \frac{E_{A_j,\pi}[C(w, \nabla w, \tilde{\pi})]}{\rho} \right) \right\},$$



(a) DM works on $A = \{x_0\}$

(b) DM works on $A = \{x_0, x_1\}$

Figure 2: Evolution of posteriors when the DM works on the projects in set $A \subseteq X$.

[5] Nonetheless, this case is similar to the case where bad projects have non-zero arrival rates; for a brief discussion on the latter, see footnote 4.

where C is the continuation value of the problem, and it depends on the distribution of posteriors and on the value function and its gradient.

Henceforth, consider the case $n = 2$. In this case, we can represent beliefs by a single number in $[0, 1]$, the belief that $\omega = \omega_0$. We can also label the sets of projects so that, for $\alpha \in \mathcal{A}$, α_0 represents the fraction of time spent on $\{x_0\}$, α_1, the fraction of time devoted to $\{x_1\}$, and α_2, the fraction of time spent on both, namely, on *simultaneous research*. The Bellman equation becomes:

$$
\begin{aligned}
w(\pi) = \max_{\alpha \in \mathcal{A}} \Bigg\{ & \alpha_0 \left[\lambda\pi - c + \frac{\lambda\pi \left(w\left(1\right) - w(\pi)\right) - \lambda\pi(1-\pi)w'(\pi)}{\rho} \right] \\
& + \alpha_1 \left[\lambda(1-\pi) - c + \frac{\lambda(1-\pi)\left(w\left(0\right) - w(\pi)\right) + \lambda\pi(1-\pi)w'(\pi)}{\rho} \right] \\
& + \alpha_2 \left[\lambda - 2c + \frac{\lambda\left(\lambda - c - w(\pi)\right)}{\rho} \right] \Bigg\}.
\end{aligned}
$$

Since the expression in braces in the Bellman equation is linear in α, optimal strategies will involve spending the full unit of time on the most promising set of projects, except perhaps at indifference points. Because of the stationarity of the problem, in looking for optimal strategies, we may restrict attention to *stationary strategies*, namely, strategies that recommend sets of projects as a function of beliefs.

By virtue of the next theorem, we shall focus on *cutoff strategies* in the sequel; these are stationary strategies with the following properties:

- If the strategy recommends $\{x_1\}$ for some $\pi \in [0, 1]$, then it also recommends $\{x_1\}$ for any $\pi' \in [0, \pi)$;
- If the strategy recommends $\{x_0\}$ for some $\pi \in [0, 1]$, then it also recommends $\{x_0\}$ for any $\pi' \in (\pi, 1]$.

Theorem 1 (Cutoff strategies)**.** *Let $\alpha^* : [0, 1] \to A$ be an optimal stationary strategy. Then, α^* is a cutoff strategy.*

3 Optimal Multi-Choice Strategy

The problem that our DM faces departs from standard multi-armed bandit problems in two ways. First, x_0 and x_1 are negatively correlated: A success for one project is conclusive evidence that the other is unproductive. Second, the DM is not restricted to choosing at most a single project at a time; hence, she faces a multi-choice multi-armed bandit problem. The multi-choice feature allows the DM to accumulate more data by experimenting with both projects simultaneously, while the correlation feature allows her to learn about both projects from the outcomes of any single one of them.

This multi-choice problem can be reinterpreted as the social-planner benchmark problem in an environment of strategic experimentation. The DM is the social planner; the different projects the DM can undertake are the different agents or players, each endowed with a single project of their own. If information about actions and outcomes is public, then the social planner learns from observing the agents just as much as our DM learns from working on her projects.

Thus, the baseline problem of our DM working on up to 2 projects can be mapped into the social-planner's problem in Klein and Rady (2011). While the results in this section can be borrowed from Klein and Rady (2011), further details and proofs are provided in the appendix, both for the sake of completeness and to be used as inputs in the proofs of results in the later sections.

To characterize the optimal strategy, we distinguish two different parameter regimes.

Case 1 (Costly research). $\rho(2c-\lambda) > \lambda(\lambda-c)$;

Case 2 (Beneficial research). $\rho(2c-\lambda) \leq \lambda(\lambda-c)$.

Under *costly research*, we have that research is *expensive*, namely $\lambda < 2c$, so simultaneous research is not profitable ex post; its only rationale is information. But information is not valuable to an *impatient* DM, one with $\rho > \frac{\lambda(\lambda-c)}{2c-\lambda}$. Conversely, we have *beneficial research* if research is *cheap*, $\lambda \geq 2c$, or if the DM is *patient*, namely, if $\rho \leq \frac{\lambda(\lambda-c)}{2c-\lambda}$.

These two cases correspond to the cases of low, intermediate, and high *stakes* in Klein and Rady (2011).[6] With the lump sum from an arrival normalized to 1, and with s denoting the flow payoff from the safe arm, the case of low stakes in Klein and Rady (2011) can be rewritten as $\rho(2s-\lambda) > \lambda(\lambda-s)$; this is exactly the inequality in case 1, with s playing the role of c. Figure 3 portrays the partition of the space of "objective parameters" $\{(\lambda,c) \in \mathbb{R}^2_+ : \lambda > c\}$ — excluding the "subjective" parameter ρ — according to whether they correspond to costly or beneficial research.

Let $\underline{\pi}, \overline{\pi} \in (0,1)$, $\overline{\pi} \geq \underline{\pi}$, be cutoffs such that the DM chooses $\{x_1\}$ for $\pi < \underline{\pi}$ and $\{x_0\}$ for $\pi > \overline{\pi}$. Optimal cutoffs are identified by means of the *value-matching* (VM) and *smooth-pasting* (SP) conditions. The (VM) conditions say that, at the cutoffs, the DM must be indifferent between the two corresponding recommendations from either side of the cutoffs. The (SP) conditions say that, at the cutoffs, the marginal value of learning from the corresponding recommended actions must be equal. Without these conditions, the DM would be switching actions "too early" or "too late."

The (VM) conditions are:

Condition (VM). $w(\underline{\pi}) = w(\overline{\pi}) = \begin{cases} 0 & \text{if } \emptyset \text{ is prescribed for } \pi \in (\underline{\pi},\overline{\pi}), \\ \lambda - c - \frac{\rho c}{\lambda+\rho} & \text{if } X \text{ is prescribed for } \pi \in (\underline{\pi},\overline{\pi}). \end{cases}$

[6] I thank Sven Rady for bringing this to my attention.

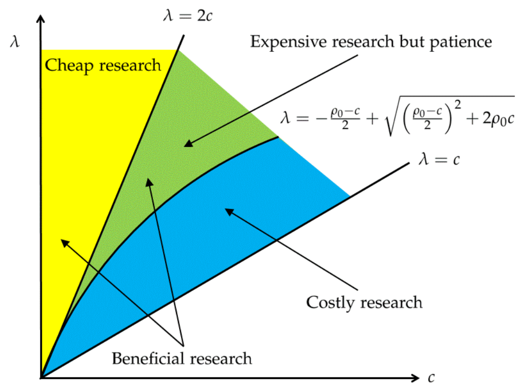


Figure 3: Objective-parameter space for a fixed $\rho = \rho_0$. The colored portion of the graph represents the parameter space for the problem. The yellow region represents cheap research. The middle curve is the level curve of the threshold $\lambda(\lambda - c) = \rho_0(2c - \lambda)$. In the green region, research is expensive but $\rho = \rho_0$ is "patience enough." These two regions combined represent case 2; the blue region represents case 1.

Regardless of whether $\varnothing$ or X is recommended for intermediate beliefs, the (SP) conditions are:

Condition (SP)**.** $w'(\underline{\pi}) = w'(\overline{\pi}) = 0$;

recall that there is no learning either from $\varnothing$ or from X.

The next result is the counterpart of Propositions 1 and 2 in Klein and Rady (2011).

Theorem 2 (Optimal strategy)**.** *The optimal strategy, α^*, is as follows. Under case 1,*

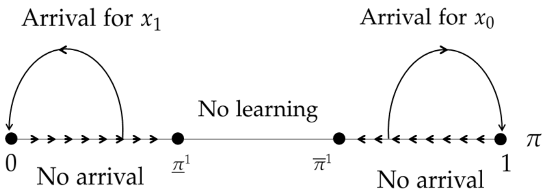
$$\alpha^*(\pi; \lambda, c) = \begin{cases} (0,1,0) & \pi \in \left[0, \underline{\pi}^1\right], \\ (0,0,0) & \pi \in \left(\underline{\pi}^1, \overline{\pi}^1\right), \\ (1,0,0) & \pi \in \left[\overline{\pi}^1, 1\right], \end{cases} \tag{1}$$

where $\underline{\pi}^1 := \frac{\lambda - c}{\lambda} \frac{\lambda + \rho}{\lambda + \rho - c} \in \left(0, \frac{1}{2}\right)$ and $\overline{\pi}^1 := 1 - \underline{\pi}^1$; under case 2,

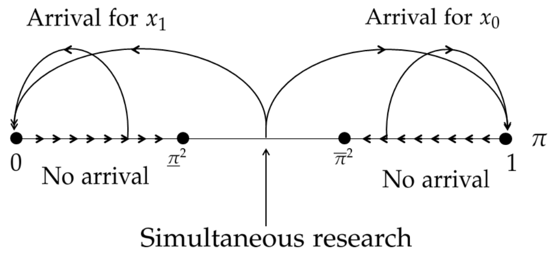
$$\alpha^*(\pi; \lambda, c) = \begin{cases} (0,1,0) & \pi \in \left[0, \underline{\pi}^2\right), \\ (0,0,1) & \pi \in \left[\underline{\pi}^2, \overline{\pi}^2\right], \\ (1,0,0) & \pi \in \left(\overline{\pi}^2, 1\right], \end{cases} \tag{2}$$

where $\underline{\pi}^2 := \frac{\lambda + \rho}{\lambda + \rho + c} \frac{c}{\lambda} \in \left(0, \frac{1}{2}\right)$ and $\overline{\pi}^2 := 1 - \underline{\pi}^2$.

Figure 4 portrays the dynamics of experimentation under the optimal strategy in Theorem 2. A sufficiently impatient DM who is unsure about the state of nature, one whose prior falls in



(a) Belief dynamics under strategy (1)



(b) Belief dynamics under strategy (2)

Figure 4: Belief dynamics under the optimal strategy

the mid range $\left(\underline{\pi}^1, \overline{\pi}^1\right)$, "gives up" if research is expensive; learning is simply too costly. If she is sufficiently confident in a project, the DM starts by working on it exclusively. If this project proves successful, then it is kept forever thereupon. While no arrivals occur for this project, the DM becomes progressively more pessimistic about it and progressively more optimistic about the other one. However, her posterior does not reach the point of being "optimistic enough" to switch to the other project: Eventually, her lost confidence leads her to drop the initial project and give up altogether, never giving the other project a chance.

If research is beneficial, and if her prior falls in the mid range $\left[\underline{\pi}^2, \overline{\pi}^2\right]$, the DM starts by working on simultaneous research. Otherwise, she starts working on a single project. Like before, while working on a single project unsuccessfully, she gradually becomes pessimistic about it and optimistic about the other one. But now, eventually, she takes on the other project as well, *without setting the "failing" project aside* — despite the higher cost and the negative correlation. The intuition is that simultaneous research gives the DM a better sense of why it is that she has been failing: Is the problem that the project is bad, or that she needs to work on it for a bit longer?

Compare this to the benchmark case where the DM is restricted to focusing on at most one project at a time. In Klein and Rady (2011), this would amount to preventing one agent to experiment if the other is experimenting. This constraint is only binding in case 2.

Theorem 3 (Single-choice benchmark)**.** *Assume that the DM can work on at most one project at a time. Under case 2, the optimal strategy, α^0, is given by:*

$$\alpha^0(\pi; \lambda, c) := \begin{cases} (0,1,0) & \pi \in \left[0, \frac{1}{2}\right), \\ \left(\frac{1}{2}, \frac{1}{2}, 0\right) & \pi = \frac{1}{2}, \\ (1,0,0) & \pi \in \left(\frac{1}{2}, 1\right]. \end{cases} \tag{3}$$

To the right of 1/2, the DM experiments on project x_1; while unsuccessful, her posterior gradually increases. To the left of 1/2, instead, the posterior gradually decreases. If we specify that the DM should choose x_1 at the cutoff, we run into the following problem: The posterior is strictly increasing at 1/2, but it switches sign and becomes strictly decreasing above 1/2.

To obtain a well-defined path of posteriors, beliefs must freeze at the cutoff 1/2.[7] A way to achieve this is to recommend the DM to *split* her time between x_0 and x_1. By dividing the intensity of experimentation equally between the two projects, beliefs (virtually) freeze at the threshold unless and until an arrival is observed: The gradual pessimism from working unsuccessfully on one project half of the time is compensated by the gradual pessimism from working unsuccessfully on the other project the other half of the time.

4 Nested Choices

In section 3, the DM has the option to take on a previously ignored project, and to resume work on a project that has been previously tried out and put aside. However, it may be that such projects "disappear": A research project that is set aside may be scooped by another researcher; an overlooked applicant or a dismissed employee may find another job and exit the market. In this section, I consider the extreme case where choices must be nested, so once a project is ignored or discarded it can never be chosen. This restriction introduces an option value to holding on to projects beyond what a less-restricted DM would consider optimal.

Now, the state space keeps track not only of the beliefs of the DM, but also of her feasible set of choices. For simplicity of the discussion, I restrict the DM to spending the entirety of each time interval on a single set of projects; that is, I consider the restricted action space $\mathcal{A}^r := \{\alpha \in \{0,1\}^3 : \alpha_0 + \alpha_1 + \alpha_2 \leq 1\}$.

Let $w^r : [0,1] \times 2^X \to \mathbb{R}$ represent the restricted value function. Clearly, $w^r(\pi, \emptyset) = 0$. The Bellman equation for $w^r(\pi, \{x_1\})$ is:

$$w^r(\pi,\{x_1\}) = \max\{0, w^r(\pi,\{x_1\}) + [\lambda(1-\pi) - c$$
$$+\frac{\lambda}{\rho}(1-\pi)\left(\lambda - c - w^r(\pi,\{x_1\}) + \pi(1-\pi)w^{r\prime}(\pi,\{x_1\})\right) - w^r(\pi,\{x_1\})\Big]\rho\mathrm{d}t\Big\};$$

either $w^r(\pi, \{x_1\}) = 0$, or $w^r(\pi, \{x_1\})$ solves the same differential equation as w does in section 3 (namely, equation 7 in the appendix). Looking for a cutoff strategy, the same (VM) and (SP) conditions relating the choice of $\{x_1\}$ and the choice of the empty set apply. Thus, we have:

$$w^r(\pi,\{x_1\}) = \begin{cases} \frac{\lambda c}{\lambda+\rho}\pi\left(\frac{\psi(\pi)}{\psi(\underline{\pi}^1)}\right)^{-\frac{\rho}{\lambda}} + \lambda(1-\pi) - c & \pi \in [0, \underline{\pi}^1], \\ 0 & \pi \in (\underline{\pi}^1, 1]. \end{cases} \tag{4}$$

[7] Klein and Rady (2011) define a strategy to be *admissible* if, starting from any prior, the strategy yields a well-defined path of posteriors $t \mapsto \pi_t$.

The same argument applies to $w^r(\pi, \{x_0\})$, leading to:

$$w^r(\pi, \{x_0\}) = \begin{cases} 0 & \pi \in [0, \overline{\pi}^1), \\ \frac{\lambda c}{\lambda+\rho}(1-\pi)\left(\frac{\psi(\pi)}{\psi(\overline{\pi}^1)}\right)^{\frac{\rho}{\lambda}} + \lambda\pi - c & \pi \in [\overline{\pi}^1, 1]. \end{cases} \tag{5}$$

Finally, for $w^r(\pi, X)$, we have:

$$w^r(\pi, X) = \max\{w^r(\pi, \{x_0\}), w^r(\pi, \{x_1\}), \\ \left(\lambda - 2c + \frac{\lambda(\lambda - c - w^r(\pi, X))}{\rho} - w^r(\pi, X)\right) \mathrm{d}t + w^r(\pi, X)\right\}.$$

As before, we look for an optimal cutoff strategy. Assume that $\rho(2c - \lambda) \leq \lambda(\lambda - c)$; hence, $\underline{\pi}^1 \geq 1/2 \geq \overline{\pi}^1$. (The case $\underline{\pi}^1 < \overline{\pi}^1$ is handled similarly.) On $[0, \overline{\pi}^1)$,

$$w^r(\pi, X) = \max\left\{w^r(\pi, \{x_1\}), \left(\lambda - 2c + \frac{\lambda(\lambda - c - w^r(\pi, X))}{\rho} - w^r(\pi, X)\right) \mathrm{d}t + w^r(\pi, X)\right\}.$$

We look for a cutoff $\underline{\pi} \in (0, \overline{\pi}^1)$ such that $w^r(\underline{\pi}, \{x_1\}) = \lambda - c - \frac{\rho c}{\lambda+\rho}$. Similarly, on $(\underline{\pi}^1, 1]$,

$$w^r(\pi, X) = \max\left\{w^r(\pi, \{x_0\}), \left(\lambda - 2c + \frac{\lambda(\lambda - c - w^r(\pi, X))}{\rho} - w^r(\pi, X)\right) \mathrm{d}t + w^r(\pi, X)\right\},$$

and we seek an analogous cutoff $\overline{\pi} \in (\underline{\pi}^1, 1)$ for $w^r(\pi, \{x_0\})$.

The existence of these cutoffs is established in Lemma A2 in the appendix. The remaining details of the optimal strategy are provided in the following theorem.

Theorem 4 (Nested choices)**.** *There exist two unique cutoffs denoted by $\underline{\pi}^N \in (0, \min\{\overline{\pi}^1, \underline{\pi}^1\})$, $\overline{\pi}^N \in (\max\{\overline{\pi}^1, \underline{\pi}^1\}, 1)$ such that the optimal strategy when choices must be nested, α^N, is as follows:*

- $\alpha^N(\pi, \varnothing; \lambda, c) = (0, 0, 0)$;
- $\alpha^N(\pi, \{x_1\}; \lambda, c) = \begin{cases} (0,1,0) & \pi \in [0, \underline{\pi}^1], \\ (0,0,0) & \pi \in (\underline{\pi}^1, 1]; \end{cases}$
- $\alpha^N(\pi, \{x_0\}; \lambda, c) = \begin{cases} (0,0,0) & \pi \in [0, \overline{\pi}^1), \\ (1,0,0) & \pi \in [\overline{\pi}^1, 1]; \end{cases}$
- *Under case 1, $\alpha^N(\pi, X; \lambda, c)$ is as in* (1) *in Theorem 2;*
- *Under case 2,* $\alpha^N(\pi, X; \lambda, c) = \begin{cases} (0,1,0) & \pi \in [0, \underline{\pi}^N), \\ (0,0,1) & \pi \in [\underline{\pi}^N, \overline{\pi}^N], \\ (1,0,0) & \pi \in (\overline{\pi}^N, 1]. \end{cases}$

At the outset, the feasible set for the DM is all of X. If research is expensive, a sufficiently impatient DM behaves as her unrestricted counterpart does. However, under cheap research or sufficient patience, the DM has to be more certain about the state to go with a singleton: If $\rho(2c-\lambda) \leq \lambda(\lambda-c)$, then $\underline{\pi}^N < \overline{\pi}^1 \leq \underline{\pi}^2 \leq \overline{\pi}^2 \leq \underline{\pi}^1 < \overline{\pi}^N$. Intuitively, by starting with a singleton, she is giving up the option value of being able to switch to the other project at a later point in time — after having gathered more information.

5 Imperfect Monitoring of Successes

The basic setup assumes that the DM can observe from which project successes come when doing simultaneous research. In other words, she can "monitor" progress on both projects closely and identify the source of a success when they are simultaneously undertaken. In recruiting, it may be the case that the output of a team can only be measured with respect to the team, and the individual contributions of the team members cannot be readily assessed. This section considers the case where the DM, when doing simultaneous research, can only observe the occurrence of arrivals but not their "precedence."

Figure 5 describes ex-post payoffs to the DM and the data she now gathers from simultaneous research. In this variation of the problem, simultaneous research is as uninformative as performing no research whatsoever; the only difference between the two is that the former yields an immediate payoff of $\lambda - 2c$. To learn about the state, the DM must give individual projects a chance to stand *on their own*. Unlike under the main specification, "experimentation" now entails focusing on singletons.

The relevant cases of costly versus beneficial research are now as follows.

Case 3 (Costly research). $\rho(\lambda - 2c) > \lambda c$;

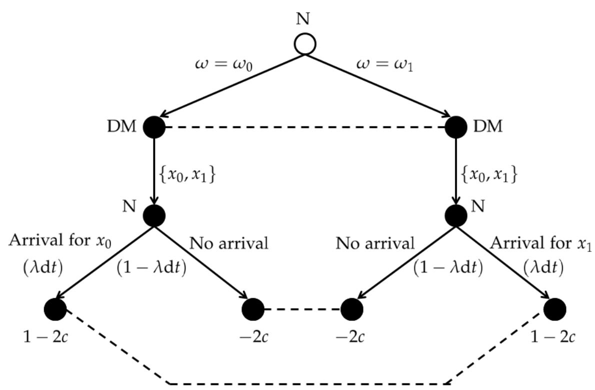


Figure 5: Observations and payoffs from simultaneous research under imperfect monitoring.

Case 4 (Beneficial research)**.** $\rho(\lambda - 2c) \leq \lambda c$.

Under case 3, we have that $\lambda > 2c$, so research is *cheap*, and $\rho > \frac{\lambda c}{\lambda - 2c}$, so the DM is *impatient*. Thus, the temptation to forgo information and exploit the projects simultaneously is high. Under case 4, either research is *expensive*, or the DM is sufficiently *patient*. Thus, working on a single project is an appealing proposition. Notice that case 4 is consistent both with cases 1 and 2: c may be high or ρ may be low. Figure 6 is the counterpart of Figure 3.

The next theorem presents the optimal strategy under imperfect monitoring of successes.

Theorem 5 (Imperfect monitoring)**.** *The optimal strategy under imperfect monitoring of successes, α^{*IM}, is as follows. Under case 3,*

$$\alpha^{*IM}(\pi;\lambda,c) = \begin{cases} (0,1,0) & \pi \in \left[0, \underline{\pi}^{IM}\right), \\ (0,0,1) & \pi \in \left[\underline{\pi}^{IM}, \overline{\pi}^{IM}\right], \\ (1,0,0) & \pi \in \left(\overline{\pi}^{IM}, 1\right], \end{cases} \tag{6}$$

where $\underline{\pi}^{IM} := \frac{\lambda+\rho}{\rho+c}\frac{c}{\lambda} \in \left(0, \frac{1}{2}\right)$ and $\overline{\pi}^{IM} := 1 - \underline{\pi}^{IM}$. Under case 4, α^{*IM} is equal to (1) *if $-\lambda(\lambda - c) > \rho(\lambda - 2c)$ (case 1) and to* (3) *if $-\lambda(\lambda - c) \leq \rho(\lambda - 2c)$ (case 2).*

If research is cheap, an impatient DM engages in simultaneous research and enjoys a constant expected payoff when she is sufficiently unsure about the state; not appreciating the additional information, she is not willing to give up the higher instant surplus to learn about the projects. Conversely, when $\lambda < 2c$, research is too expensive for the DM to ever want to undertake both projects at once. The same is true if they are cheap but the DM is sufficiently patient: While simultaneous research may be attractive, more so is the information that only singletons can

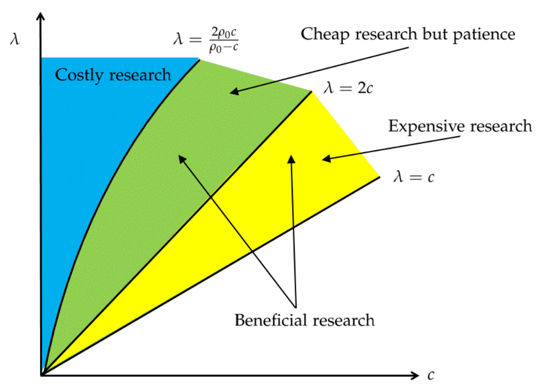


Figure 6: Objective-parameter space for a fixed $\rho = \rho_0$ under imperfect monitoring.

provide. Under case 4/2, she never gives up; under case 4/1, however, she gives up after having worked on a single project unsuccessfully for a sufficiently long time.

6 More Than Two Projects

Thus far, the DM has been given only two projects to choose from, each of which is equally appealing in its corresponding state. There are two related complications to generalizing the analysis in these directions. With n projects, the total number of possible sets of projects to undertake is 2^n. Thus, the size of the choice set grows exponentially in n. At the same time, even under the extreme negative-correlation specification, the state space — the simplex of posteriors — is multidimensional. Thus, the Bellman equation and (SP) conditions involve partial differential equations, and "cutoffs" are surfaces rather than points.

These limitations, both computational and technical, are severe. Nonetheless, the analysis of the case $n = 2$ suggests the structure of the solution to the more general problem. But before discussing this further, the next subsection analyzes a problem with three projects and a different correlation structure: Two negatively-correlated projects as before, and a third, independent and riskier one.

6.1 A Third, Risky Project

Imagine that the DM has a third project on which she can work. If this third project is productive, its arrival rate is higher than that of the other two. But this new project may be unproductive, while one of the original two projects is certainly productive; this new project may never flourish, while one of the others eventually will. For simplicity, I assume that this third project is "incompatible" with the other two in the sense that it requires the full attention of the DM while she is working on it; and that it must be forsaken once ignored or abandoned. Thus, the problem is to determine for how long to experiment on the riskier project, if at all, before switching to one or both of the original two.

The set of projects is now $X = \{x_0, x_1, y\}$. The DM allocates her time between the different subsets of $\{x_0, x_1\}$ and $\{y\}$. Undertaking the new project involves a flow cost of $c_y > 0$. There is a new state of nature, $\theta \in \{0, 1\}$, the realization of which is unobserved; the new project produces successes at a rate $\lambda_y \theta$, where λ_y is *known*, and $\lambda_y > \lambda$. Thus, the DM knows that, if this new project proves successful, it is more appealing than any of the others; otherwise, she is better off with the original ones.

Let $\mu \in [0, 1]$ denote the assessment of the DM that $\theta = 1$; $\mu^0 \in [0, 1]$ denotes the corresponding prior. Immediate rewards are $\lambda_y \mu - c_y$. Representing separate projects, I assume that ω and θ are independent. Thus, since the DM updates her beliefs about each state from independent data, the posteriors are also independent. In the event of an arrival from y, the posterior on $\theta = 1$ jumps to 1; after spending a fraction $\gamma \in [0, 1]$ of time on $\{y\}$ without observing any success, her posterior on $\theta = 1$ gradually decreases according to $\dot{\mu}_t = -\gamma \lambda_y \mu_t (1 - \mu_t)$.

Under irreversibility, once the DM switches away from y, she is back to the problem analyzed in section 3; she divides her time among x_0, x_1, or both, according to the corresponding optimal strategy, and she enjoys a continuation payoff of $w(\pi)$ if π is the posterior that $\omega = \omega_0$ at the time of switching. Thus, while she has not yet switched away from y, the Bellman equation for the new value function w_y is:

$$w_y(\pi, \mu) = \max \left\{ w(\pi), \lambda_y \mu - c_y + \frac{\lambda_y \mu \left(w_y(\pi, 1) - w_y(\pi, \mu) \right) - \lambda_y \mu (1 - \mu) w'_{y2}(\pi, \mu)}{\rho} \right\}.$$

Consider a strategy such that, for each $\pi \in [0, 1]$, there is some $\underline{\mu}(\pi) \in [0, 1]$ such that the DM starts by working on project y if $\mu \geq \underline{\mu}(\pi)$, and follows the optimal strategy among $\{x_0, x_1\}$ otherwise. The (VM) and (SP) conditions are:

Condition (VM). $w_y\left(\pi, \underline{\mu}(\pi)\right) = w(\pi)$ for all $\pi \in [0, 1]$.

Condition (SP). $w'_{y2}\left(\pi, \underline{\mu}(\pi)\right) = 0$ for all $\pi \in [0, 1]$.

Combining these two conditions gives

$$\underline{\mu}(\pi) = \frac{\rho}{\lambda_y} \frac{w(\pi) + c_y}{\lambda_y + \rho - w(\pi) - c_y}.$$

Theorem 6 (Third project). *The optimal strategy consists of starting on $\{y\}$ provided that $\mu^0 \geq \underline{\mu}(\pi^0)$, and sticking to it while the posterior (π, μ) satisfies $\mu \geq \underline{\mu}(\pi)$; if $\mu < \underline{\mu}(\pi)$, switch to the optimal strategy of Theorem 2 on the set $\{x_0, x_1\}$.*

Figure 7 represents the threshold $\underline{\mu}(\pi)$. If the DM is sufficiently sure about ω, she has to be sufficiently confident that $\theta = 1$ to start on project y. While experimenting unsuccessfully on y, her beliefs about $\theta = 1$ gradually decline. If a success occurs before she switches, however, she learns that $\theta = 1$, and sticks to y thereupon. If she is unsure about ω, she works on project y for a wider range of beliefs on θ: If she switches to the original projects, she will work on both x_0 and x_1 at once and bear a higher total research cost, or give up altogether.

6.2 More Than 2 Projects: a Conjecture and Beyond

Even under the extreme negative-correlation structure, the problem becomes intractable very quickly for $n > 2$. Nonetheless, the analysis of the 2-project problem suggests the following conjecture for the optimal strategy for the full problem.

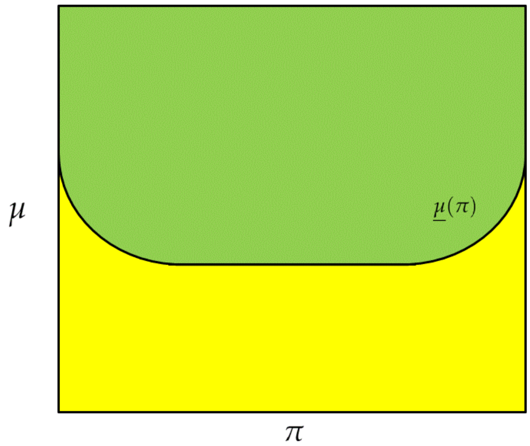
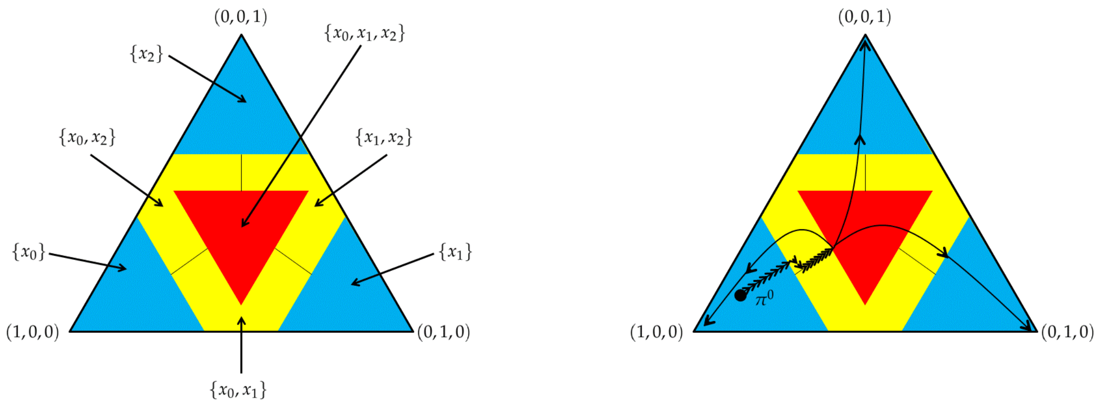


Figure 7: Cutoff $\underline{\mu}(\pi)$. Above the cutoff, the optimal strategy recommends $\{y\}$; below the cutoff, it recommends following the optimal strategy for $\{x_0, x_1\}$.

Conjecture. *We partition the parameter space into n regions on which only up to $k = 1, \ldots, n$ out of the n projects are undertaken at a time. If the DM is sufficiently confident in a project, she starts working on it exclusively. And she progressively takes on one additional project at a time as she becomes gradually pessimistic.*

Figure 8 represents this conjecture for the case $n = 3$, when research is beneficial enough that up to all 3 projects are undertaken.



(a) Conjectured strategy

(b) Belief dynamics under the conjectured strategy

Figure 8: Conjecture for $n \geq 3$

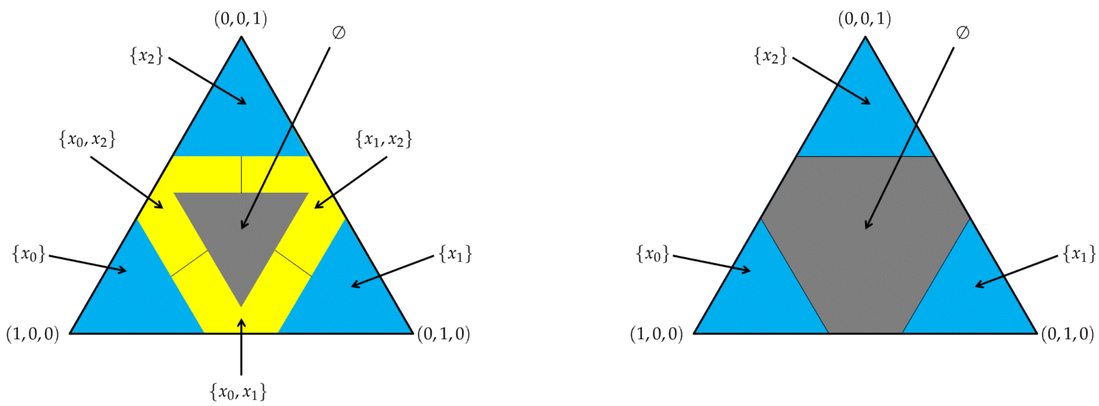
Consider the path of posteriors depicted in panel 8b. The prior falls in the region where the DM starts working on x_0 alone. While working unsuccessfully on it, her posterior starts moving towards the northwestern yellow region; eventually, she takes on project x_2 as well. Continued failure now pushes the posterior gradually in the direction of the x_1 corner; when beliefs are in the frontier of the two yellow regions, the DM holds on to x_0 and splits her time between x_1 and x_2 as in the manner of strategy (3). Eventually, if no successes are observed, she becomes sufficiently unsure and takes on all three projects.

Figure 9 represents the conjecture when the DM takes on at most two projects — panel 9a — and at most one project — panel 9b — at a time.

As a means to overcome the technical and computational limitations of larger problems, Francetich and Kreps (2014) explores *heuristics*. This paper looks at a larger class of multi-choice experimentation problems and investigates the performance of a variety of heuristic decision rules. In the paper, we provide theoretical results about the long-run performance of different heuristics (as in the manner of the literature on bandit learning algorithms); but our main interest is in how the heuristics perform for discount factors that are bounded away from 1, where both the future and the present effectively matter.

Giving up hope of being able to characterize optimal research strategies, the kinds of questions that we can address are, for instance, in which kinds of environments simple or standard heuristics perform well or poorly, and what kind of desiderata they satisfy.

What we see in test-problem simulations is that more simple-minded heuristics—those that ignore any prior information the decision maker might have and just use empirical evidence—take much too long (at these discount rates) to make good decisions; for such heuristics, quick-and-sloppy decisions are surprisingly better than slow-and-careful decisions. Among the more sophisticated heuristics that make use of prior information, the winner is usually the one-period-



(a) At most two projects undertaken at once

(b) At most one project undertaken at once

Figure 9: Conjecture for $n \geq 3$, continued

look-ahead heuristic (based on approximate dynamic programming), so even a limited amount of foresight can lead us a long way.

When projects are negatively correlated, their failure to produce breakthroughs is informative about other projects. As a result, provided that research is not costly, the DM never gives up a failing project unless and until uncertainty is resolved. Now, for richer correlation structures, failures on some projects may teach us little or nothing about other projects. Thus, the DM may want to discard some failing projects even before uncertainty is resolved. Herein lies a key difficulty in richer problems: While it is easy for the decision maker to avoid ending end up with too many projects on her table (bad projects tried infinitely often will be eventually identified as such), it is hard to prevent giving up valuable projects prematurely. So, it can pay to combine a short-listing rule with periodical re-evaluation of past cuts, in case some good project was not given the chance it deserved.

7 Conclusion

This paper analyzes the experimentation problem faced by a decision maker who can work on multiple projects at once over time. One and only one of these projects can produce successes, but the decision maker does not know ex-ante which one. To learn about the projects, she may work on one at a time, exploiting their correlation, or on multiple projects at once, gathering more data albeit at a higher cost.

Due to technical and computational issues, most of the analysis focuses on the case of two projects. If experimentation is cheap, or if she is sufficiently patient, the decision maker starts by working on both projects at once if she is sufficiently uncertain about the state of nature; more so if projects ignored or discarded can be scooped. Working on both projects at once, she can identify the profitable one as soon as the first breakthrough occurs, learning nothing new in the meantime — lack of success on both projects is a "neutral" event.

If she is sufficiently sure about a project, she starts working on it exclusively. As long as she encounters no success, she becomes gradually pessimistic. Eventually, once she becomes sufficiently unsure, she takes on the other project as well. While bad news about one project is good news about the other one, the lack of success leaves her unsure about which project is better, rather than confident enough in the neglected one. At this point, she works on both projects at once until her uncertainty is resolved.

If costs are high and the decision maker is sufficiently impatient, she eventually gives up if there are no arrivals. In this case, the neglected project is never given a chance: Before her posterior reaches a level where it would be optimal to work on it, it reaches a level of uncertainty such that neither project is worth pursuing any further.

When the projects cannot be "individually monitored" unless they are studied in isolation, the only way for the decision maker to learn is to stick to singletons, to test the projects "on their own." In this case, it is the impatient decision maker the one who works on both projects at once when experimentation is cheap; she is not willing to give up the higher immediate surplus to learn about the state of nature.

The structure of the problem studied here is extremely simple. However, while some extensions are feasible, such as allowing for "interior" states $0 < \underline{\omega} < \overline{\omega} < 1$, the problem can become intractable or far too cumbersome very quickly. Allowing for a richer set of projects or of states would certainly be interesting, but it requires expanding the size of the problem and the dimensionality of the state space.[8] Thus, for a richer discrete-time problem in a similar vein, Francetich and Kreps (2014) explores heuristics.

A very different but interesting extension is motivated by the recruiting example. The researcher is a principal who can hire multiple agents at once, and there is uncertainty about the dexterity or productivity of the job applicants. But imagine these agents must exert unobservable effort to produce output or breakthroughs. Now, we have a moral-hazard problem: For poor performance to be informative of productivity, rather than being simply a reflection of shirking, wages and compensations must incentivize the agents to work hard. The cost of experimentation is the cost of incentivizing the agents to put in the target effort, which determines the quality of information. The novel questions here are How much effort should the agents be induced to exert, that is, how much data does the principal want to collect, and how should the labor contracts be designed to provide the right incentives?

Notice that this problem is different from standard problems of delegating experimentation; it is rather an experimentation problem with endogenous costs of experimentation, given by the costs of providing incentives, and with the quantity of information also given endogenously by the target level of effort. This is the subject of ongoing work.

References

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

Forand, J.G. (2013). Keeping your options open.

Francetich, A. and Kreps, D. (2014). Choosing a good toolkit: An essay in behavioral economics. Working Paper.

Gittins, J. and Jones, D. (1974). A dynamic allocation index for the sequential design of experiments. In: *Progress in Statistics*, pp. 241–266. Amsterdam: North Holland.

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

[8]Some work has been done dealing with vector-valued states, such as Klein and Rady (2011) and Forand (2013); however, their analysis is catered to their formulation, and does not readily extend to the present setup.

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

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

Nyarko, Y. (1994). On the convexity of the value function in bayesian optimal control problems. *Economic Theory* 2(2):303–309.

Weber, R. (1992). On the gittins index for multiarmed bandits. *The Annals of Applied Probability* 2(4).

Whittle, P. (1980). Multi-armed bandits and the gittins index. *Journal of the Royal Statistical Society. Series B (Methodological)* 42(2):143–149.

A Proofs

Proof of Theorem 1. It is well known that the value function in Bayesian control problems is convex (see Nyarko, 1994 for the discrete-time case). Since $w(0) = w(1)$, convexity means that w is U-shaped. Therefore, we can find $0 < \underline{\pi} \leq \overline{\pi} < 1$ such that w is non-increasing on $[0, \underline{\pi}]$, constant on $[\underline{\pi}, \overline{\pi}]$, and non-decreasing on $[\overline{\pi}, 1]$. Higher values of π are good news for project x_0 and bad news for x_1. Thus, the non-increasing portion of w corresponds to beliefs that recommend $\{x_1\}$, while the non-decreasing region, to beliefs that recommend $\{x_0\}$. □

In what follows, consider strategies $\alpha(\cdot; \lambda, c) : [0, 1] \to \mathcal{A}$ with the following properties:

- There is some $\underline{\pi} \in (0, 1)$ such that, for all $\pi \in [0, \underline{\pi})$, $\alpha(\pi; \lambda, c) = (0, 1, 0)$;
- There is some $\overline{\pi} \in (0, 1)$, $\overline{\pi} > \underline{\pi}$, such that, for all $\pi \in (\overline{\pi}, 1]$, $\alpha(\pi; \lambda, c) = (1, 0, 0)$.

On $(0, \underline{\pi})$, we have:

$$-\lambda\pi(1-\pi)w'(\pi) + (\rho + \lambda(1-\pi))w(\pi) = \lambda(1-\pi)(\rho + \lambda - c) - \rho c. \tag{7}$$

This equation is similar to equation (1) in Keller and Rady (2010). Up to a constant of integration C_1, the solution to this differential equation is $w(\pi) = C_1 \pi \psi(\pi)^{-\frac{\rho}{\lambda}} + \lambda(1-\pi) - c$. On $(\overline{\pi}, 1)$, we have:

$$\lambda\pi(1-\pi)w'(\pi) + (\rho + \lambda\pi)w(\pi) = \lambda\pi(\rho + \lambda - c) - \rho c. \tag{8}$$

This equation is almost identical to equation (1) in Keller and Rady (2010); up to a constant of integration C_0, the solution is $w(\pi) = C_0(1-\pi)\psi(\pi)^{\frac{\rho}{\lambda}} + \lambda\pi - c$. We pin down $\underline{\pi}$, $\overline{\pi}$, C_1, and C_0 by means of the (VM) and (SP) conditions.

For $\alpha(\pi;\lambda,c) = (0,0,0)$ for all $\pi \in (\underline{\pi},\overline{\pi})$, the (VM) and (SP) conditions lead to $\underline{\pi} = \frac{\lambda-c}{\lambda}\frac{\lambda+\rho}{\lambda+\rho-c} \in (0,1)$, $\overline{\pi} = 1-\underline{\pi} \in (0,1)$, $C_1 = \frac{\lambda c}{\lambda+\rho}\psi(\underline{\pi})^{\frac{\rho}{\lambda}}$, and $C_0 = C_1\psi(\underline{\pi})^{-2\frac{\rho}{\lambda}}$. We have that $\overline{\pi} > \underline{\pi}$ if and only if $\rho(2c-\lambda) > \lambda(\lambda-c)$, namely, under case 1. Thus, we identify the following value function:

$$w(\pi) = \begin{cases} \frac{\lambda c}{\lambda+\rho}\pi\left(\frac{\psi(\pi)}{\psi(\underline{\pi})}\right)^{-\frac{\rho}{\lambda}} + \lambda(1-\pi) - c & \pi \in [0,\underline{\pi}]; \\ 0 & \pi \in (\underline{\pi},\overline{\pi}); \\ \frac{\lambda c}{\lambda+\rho}(1-\pi)\left(\frac{\psi(\pi)}{\psi(\overline{\pi})}\right)^{\frac{\rho}{\lambda}} + \lambda\pi - c & \pi \in [\overline{\pi},1]. \end{cases} \tag{9}$$

For $\alpha(\pi;\lambda,c) = (0,0,1)$ for all $\pi \in (\underline{\pi},\overline{\pi})$, we have:

$$\underline{\pi} = \frac{\lambda+\rho}{\lambda+\rho+c}\frac{c}{\lambda} \in (0,1), \quad \overline{\pi} = 1-\underline{\pi} \in (0,1),$$

$C_1 = \frac{\lambda c}{\lambda+\rho}\psi(\underline{\pi})^{\frac{\rho}{\lambda}+1}$, and $C_0 = C_1\psi(\underline{\pi})^{-2\left(\frac{\rho}{\lambda}+1\right)}$. Notice that $\overline{\pi} \geq \underline{\pi}$ if and only if we have that $\rho(2c-\lambda) \leq \lambda(\lambda-c)$, namely, under case 2. The value function is now:

$$w(\pi) = \begin{cases} \frac{\lambda c\psi(\underline{\pi})}{\lambda+\rho}\pi\left(\frac{\psi(\pi)}{\psi(\underline{\pi})}\right)^{-\frac{\rho}{\lambda}} + \lambda(1-\pi) - c & \pi \in [0,\underline{\pi}); \\ \lambda - c - \frac{\rho c}{\lambda+\rho} & \pi \in [\underline{\pi},\overline{\pi}]; \\ \frac{\lambda c}{(\lambda+\rho)\psi(\overline{\pi})}(1-\pi)\left(\frac{\psi(\pi)}{\psi(\overline{\pi})}\right)^{\frac{\rho}{\lambda}} + \lambda\pi - c & \pi \in (\overline{\pi},1]. \end{cases} \tag{10}$$

Finally, in the single-choice benchmark and under case 2, setting $\underline{\pi} = \overline{\pi} = 1/2$, we have:

$$w(\pi) = \begin{cases} \frac{(\lambda)^2}{\lambda+2\rho}\pi\psi(\pi)^{-\frac{\rho}{\lambda}} + \lambda(1-\pi) - c & \pi \in \left[0,\frac{1}{2}\right]; \\ \frac{(\lambda)^2}{\lambda+2\rho}(1-\pi)\psi(\pi)^{\frac{\rho}{\lambda}} + \lambda\pi - c & \pi \in \left(\frac{1}{2},1\right]. \end{cases} \tag{11}$$

Lemma A1. *The functions $w : [0,1] \to \mathbb{R}$ in* (9), (10), *and* (11) *are continuously differentiable, strictly decreasing below $\underline{\pi}$ (1/2 for* (11)*), and strictly increasing above $\overline{\pi}$ (1/2 for* (11)*).*

Proof. We prove the lemma for (9); the other cases are analogous. Continuous differentiability follows from value matching and smooth pasting. On $[0,\underline{\pi})$,

$$w'(\pi) = \frac{\lambda c\psi(\underline{\pi})}{\lambda+\rho}\left(\frac{\psi(\pi)}{\psi(\underline{\pi})}\right)^{-\frac{\rho}{\lambda}}\frac{\lambda(1-\pi)+\rho}{\lambda(1-\pi)} - \lambda < \frac{\lambda c\psi(\underline{\pi})}{\lambda+\rho}\frac{\lambda(1-\underline{\pi})+\rho}{\lambda(1-\underline{\pi})} - \lambda = 0.$$

Finally, on $(\overline{\pi},1]$,

$$w'(\pi) = -\frac{\lambda c}{(\lambda+\rho)\psi(\overline{\pi})}\left(\frac{\psi(\pi)}{\psi(\overline{\pi})}\right)^{\frac{\rho}{\lambda}}\frac{\lambda\pi+\rho}{\lambda\pi} + \lambda > -\frac{\lambda c}{(\lambda+\rho)\psi(\overline{\pi})}\frac{\lambda\overline{\pi}+\rho}{\lambda\overline{\pi}} + \lambda = 0.$$

This concludes the proof. □

Proof of Theorem 2. Consider case 1, $\rho(2c-\lambda) > \lambda(\lambda-c)$. We want to verify that w in (9) solves the Bellman equation on page 8. To this end, define:

$$R_w^0(\pi) := \lambda\pi - c + \frac{\lambda\pi\left[\lambda - c - w(\pi)\right] - \lambda\pi(1-\pi)w'(\pi)}{\rho};$$
$$R_w^1(\pi) := \lambda(1-\pi) - c + \frac{\lambda(1-\pi)\left[\lambda - c - w(\pi)\right] + \lambda\pi(1-\pi)w'(\pi)}{\rho}.$$

We must check the following conditions:

a). On $\left[0, \underline{\pi}^1\right)$, $R_w^1(\pi) > \max\left\{R_w^0(\pi), 0, \lambda - c - \frac{\rho c}{\lambda+\rho}\right\}$;

b). On $\left(\underline{\pi}^1, \overline{\pi}^1\right)$, $0 > \max\left\{R_w^0(\pi), R_w^1(\pi), \lambda - c - \frac{\rho c}{\lambda+\rho}\right\}$;

c). Finally, on $\left(\overline{\pi}^1, 1\right]$, $R_w^0(\pi) > \max\left\{R_w^1(\pi), 0, \lambda - c - \frac{\rho c}{\lambda+\rho}\right\}$.

Notice that, under case 1, $\lambda - c - \frac{\rho c}{\lambda+\rho} < 0$.
a). Start with $\pi \in (0, \underline{\pi}^1)$. Using (7), we can write $R_w^1(\pi) - R_w^0(\pi)$ as:

$$\begin{aligned}
R_w^1(\pi) - R_w^0(\pi) &= \lambda(1-2\pi) + \frac{\lambda(1-2\pi)\left[\lambda - c - w(\pi)\right] + 2\lambda\pi(1-\pi)w'(\pi)}{\rho} \\
&= \lambda - \frac{(\lambda+2\rho)\left[\lambda - c - w(\pi)\right]}{\rho} \\
&> \lambda - \frac{(\lambda+2\rho)\left[\lambda - c - w(\underline{\pi}^1)\right]}{\rho} \\
&= \lambda - \frac{(\lambda+2\rho)}{\rho}(\lambda - c) > 0,
\end{aligned}$$

where the first strict inequality follows from the fact that w is strictly decreasing on $[0, \underline{\pi}^1)$. Similarly,

$$\begin{aligned}
R_w^1(\pi) &= \lambda(1-\pi) - c + \frac{\lambda(1-\pi)\left[\lambda - c - w(\pi)\right] + \lambda\pi(1-\pi)w'(\pi)}{\rho} \\
&= \lambda(1-\pi) - c - \left[\lambda - c - w(\pi)\right] + \lambda\pi = w(\pi) > w(\underline{\pi}^1) = 0.
\end{aligned}$$

b). Next, consider $\pi \in \left(\underline{\pi}^1, \overline{\pi}^1\right)$. In this region, $w(\pi) = w'(\pi) = 0$. Now,

$$R_w^0(\pi) = \lambda\pi\left(\frac{\rho + \lambda - c}{\rho}\right) - c < \lambda\overline{\pi}^1\left(\frac{\rho + \lambda - c}{\rho}\right) - c = 0;$$
$$R_w^1(\pi) = \lambda(1-\pi)\left(\frac{\rho + \lambda - c}{\rho}\right) - c < \lambda\left(1 - \underline{\pi}^1\right)\left(\frac{\rho + \lambda - c}{\rho}\right) - c = 0.$$

c). Finally, for $\pi \in \left(\overline{\pi}^1, 1\right]$, using (8), we have:

$$R_w^0(\pi) - R_w^1(\pi) = \lambda(2\pi - 1) + \frac{\lambda(2\pi-1)\left[\lambda - c - w(\pi)\right] - 2\lambda\pi(1-\pi)w'(\pi)}{\rho}$$

$$
\begin{aligned}
&= \lambda - \frac{\lambda + 2\rho}{\rho}\left[\lambda - c - w(\pi)\right] \\
&> \lambda - \frac{\lambda + 2\rho}{\rho}\left[\lambda - c - w\left(\overline{\pi}^1\right)\right] \\
&= \lambda - \frac{(\lambda + 2\rho)}{\rho}(\lambda - c) > 0,
\end{aligned}
$$

where the (first) strict inequality follows because w is strictly increasing on this region; and:

$$
R_w^0(\pi) = \lambda\pi - c + \frac{-\rho(\lambda - c - w(\pi)) + \lambda\rho(1-\pi)}{\rho} = w(\pi) > w\left(\overline{\pi}^1\right) = 0.
$$

Consider the case $\rho(2c - \lambda) < \lambda(\lambda - c)$. To verify w in (10) solves the Bellman equation, define:

$$
R_w^2(\pi) := \lambda - 2c + \frac{\lambda\left[\lambda - c - w(\pi)\right]}{\rho}.
$$

By assumption, $R_w^2(\pi) > 0$ for all $\pi \in [0, 1]$. We must check that:

a). On $(0, \underline{\pi}^2)$, $R_w^1(\pi) - R_w^0(\pi) > 0$ and $R_w^1(\pi) - R_w^2(\pi) > 0$.

b). On $(\underline{\pi}^2, \overline{\pi}^2)$, $R_w^2(\pi) - R_w^0(\pi) > 0$ and $R_w^2(\pi) - R_w^1(\pi) > 0$.

c). Finally, on $(\overline{\pi}^2, 1)$, $R_w^0(\pi) - R_w^1(\pi) > 0$ and $R_w^0(\pi) - R_w^2(\pi) > 0$.

a). Start with $\pi \in (0, \underline{\pi}^2)$. The first inequality in 1 is established in the same way as before. As for the second,

$$
R_w^1(\pi) - R_w^2(\pi) = c - \frac{(\lambda + \rho)\left[\lambda - c - w(\pi)\right]}{\rho} > c - \frac{(\lambda + \rho)\left[\lambda - c - w(\underline{\pi}^2)\right]}{\rho} = 0.
$$

b). Next, take $\pi \in (\underline{\pi}^2, \overline{\pi}^2)$. Now,

$$
\begin{aligned}
R_w^2(\pi) - R_w^0(\pi) &= \lambda(1-\pi)\frac{\lambda + \rho + c}{\lambda + \rho} - c > \lambda(1 - \overline{\pi}^2)\frac{\lambda + \rho + c}{\lambda + \rho} - c = 0; \\
R_w^2(\pi) - R_w^1(\pi) &= \lambda\pi\frac{\lambda + \rho + c}{\lambda + \rho} - c > \lambda\overline{\pi}^2\frac{\lambda + \rho + c}{\lambda + \rho} - c = 0.
\end{aligned}
$$

c). Finally,

$$
\begin{aligned}
R_w^0(\pi) - R_w^2(\pi) &= -\left[\lambda(1-\pi) - c\right] - \frac{\lambda(1-\pi)\left[\lambda - c - w(\pi)\right] + \lambda\pi(1-\pi)w'(\pi)}{\rho} \\
&= c - \frac{(\lambda + \rho)\left[\lambda - c - w(\pi)\right]}{\rho} \\
&> c - \frac{(\lambda + \rho)\left[\lambda - c - w(\overline{\pi}^2)\right]}{\rho} = 0.
\end{aligned}
$$

This conclues the proof. $\square$

Proof of Theorem 3. The proof that w in (11) solves the Bellman equation is analogous to the counterpart proof for w in (9); further details are omitted. Under case 2, the DM cannot profit from giving up, as $w(\pi) \geq w(1/2) \geq 0$ if $\rho(2c-\lambda) \leq \lambda(\lambda - c)$. $\square$

Lemma A2. *There exists a unique* $\underline{\pi}^N \in \left(0, \min\left\{\overline{\pi}^1, \underline{\pi}^1\right\}\right)$ *such that:*

$$\frac{\lambda c}{\lambda+\rho}\underline{\pi}^N \left(\frac{\psi(\underline{\pi}^N)}{\psi\left(\max\left\{\overline{\pi}^1, \underline{\pi}^1\right\}\right)}\right)^{-\frac{\rho}{\lambda}} + \lambda(1-\underline{\pi}^N) - c = \lambda - c - \frac{\rho c}{\lambda+\rho};$$

similarly, there exists a unique $\overline{\pi}^N \in \left(\max\left\{\overline{\pi}^1, \underline{\pi}^1\right\}\right), 1$ *such that:*

$$\frac{\lambda c}{\lambda+\rho}(1-\overline{\pi}^N) \left(\frac{\psi(\overline{\pi}^N)}{\psi\left(\min\left\{\overline{\pi}^1, \underline{\pi}^1\right\}\right)}\right)^{\frac{\rho}{\lambda}} + \lambda\overline{\pi}^N - c = \lambda - c - \frac{\rho c}{\lambda+\rho}.$$

Proof. Consider the case $\rho(2c-\lambda) \leq \lambda(\lambda-c)$; the other case is handled analogously. Define the following function $h : [0,1] \to \mathbb{R}$, given by:

$$h(x) := \frac{\lambda c}{\lambda+\rho} x \left(\frac{\psi(x)}{\psi\left(\underline{\pi}^1\right)}\right)^{-\frac{\rho}{\lambda}} + \lambda(1-x) - c - (\lambda - c) + \frac{\rho c}{\lambda+\rho}.$$

By Lemma A1, h is differentiable and strictly decreasing on $\left[0, \underline{\pi}^1\right)$. Moreover, this function satisfies:

$$\begin{aligned} h(0) &= \frac{\rho c}{\lambda+\rho} > 0; \\ h\left(\overline{\pi}^1\right) &< \frac{\lambda c}{\lambda+\rho}\overline{\pi}^1 + \lambda\underline{\pi}^1 - c - (\lambda-c) + \frac{\rho c}{\lambda+\rho} = 0. \end{aligned}$$

Thus, there exists a unique $x^* \in \left(0, \overline{\pi}^1\right)$ such that $h(x^*) = 0$. A similar argument as above establishes that there exists a unique $x^{**} \in \left(\underline{\pi}^1, 1\right)$ such that $g(x^{**}) = 0$, where $g : [0,1] \to \mathbb{R}$ is given by:

$$g(x) := \frac{\lambda c}{\lambda+\rho}(1-x) \left(\frac{\psi(x)}{\psi\left(\overline{\pi}^1\right)}\right)^{\frac{\rho}{\lambda}} + -c - (\lambda-c) + \frac{\rho c}{\lambda+\rho}.$$

Set $\underline{\pi}^N = x^*$ and $\overline{\pi}^N = x^{**}$. $\square$

Proof of Theorem 4. There is nothing to show if the feasible set is the empty set. The portions of the theorem corresponding to singletons being the feasible sets follow as in the proof of Theorem 2. (The only difference is that, here, we do not need to worry about having $\underline{\pi}^1 < \overline{\pi}^1$; the two cutoffs apply to different states.) As for the last two cases, it suffices to compare the value functions in (9) and (10) to the value corresponding to simultaneous research. Start with the case $\rho(2c-\lambda) \le \lambda(\lambda-c)$. We have $\underline{\pi}^2 > \overline{\pi}^1$; thus, on $[0, \overline{\pi}^1]$, $w(\pi) - (\lambda - c) + \frac{\rho c}{\lambda+\rho} = h(\pi)$, where h is as in the proof of Lemma A2. Thus, for all $\pi < \underline{\pi}^N$, $w(\pi) > \lambda - c - \frac{\rho c}{\lambda+\rho}$. Similarly, on $[\underline{\pi}^1, 1]$, we have $w(\pi) - (\lambda-c) + \frac{\rho c}{\lambda+\rho} = g(\pi)$, and (the proof of) Lemma A2 establishes that $w(\pi) > \lambda - c - \frac{\rho c}{\lambda+\rho}$ for all $\pi > \overline{\pi}^N$. Finally, if $\lambda(\lambda-c) < \rho(2c-\lambda)$, the desired result follows from the fact that $\lambda - c - \frac{\rho c}{\lambda+\rho} < 0$. □

Proof of Theorem 5. The (SP) and (VM) conditions for strategies recommending simultaneous research for mid-range beliefs lead to $\underline{\pi} = \frac{\lambda+\rho}{\rho+c}\frac{c}{\lambda} \in (0,1)$ and $\overline{\pi} = 1 - \underline{\pi}$. We have $\overline{\pi} > \underline{\pi}$ if and only if $\rho(\lambda - 2c) > \lambda c$. The solution candidate in this case is:

$$w(\pi) = \begin{cases} \frac{\lambda(\lambda-c)}{\lambda+\rho}\pi\left(\frac{\psi(\pi)}{\psi(\underline{\pi})}\right)^{-\frac{\rho}{\lambda}} + \lambda(1-\pi) - c & \pi \in [0, \underline{\pi}); \\ \lambda - 2c & \pi \in [\underline{\pi}, \overline{\pi}]; \\ \frac{\lambda(\lambda-c)}{\lambda+\rho}(1-\pi)\left(\frac{\psi(\pi)}{\psi(\overline{\pi})}\right)^{\frac{\rho}{\lambda}} + \lambda\pi - c & \pi \in (\overline{\pi}, 1]. \end{cases}$$

The proof that this function solves the Bellman equation is entirely analogous to the corresponding proof in Theorem 2. The argument for the case $\rho(\lambda-2c) \le \lambda c$ is analogous to the argument behind Theorems 2 and 3; notice that:

$$w\left(\frac{1}{2}\right) = \frac{\lambda(\rho+\lambda-c) - 2\rho c}{\lambda+2\rho} \ge \lambda - 2c$$

if and only if $-\lambda(\lambda-c) \le \rho(\lambda-2c) \le \lambda c$, while $w(1/2) < 0$ if and only if $-\lambda(\lambda-c) > \rho(\lambda-2c)$. □

Proof of Theorem 6. On the region of the state space where the DM experiments with y, we have:

$$w_y(\pi,\mu) = \lambda_y\mu - c_y + \frac{\lambda_y\mu\left(w_y(\pi,1) - w_y(\pi,\mu)\right) - \lambda_y\mu(1-\mu)w'_{y2}(\pi,\mu)}{\rho}.$$

By assumption, $w_y(\pi,1) = \lambda_y - c_y$. Thus, $w(\pi,\mu) = C(\pi)(1-\mu)\psi(\mu)^{\frac{\rho}{\lambda_y}} + \lambda_y\mu - c_y$, where $C(\cdot)$ is some continuously differentiable function. From the (VM) and (SP) conditions, we find

$$\underline{\mu}(\pi) = \frac{\rho}{\lambda_y}\frac{w(\pi) + c_y}{\lambda_y + \rho - w(\pi) - c_y}$$

and $C(\pi) = \frac{\lambda_y}{\lambda_y+\rho}\psi\left(\underline{\mu}(\pi)\right)^{-\frac{\rho}{\lambda_y}}$. The value function of the proposed strategy is:

$$w_y(\pi,\mu) = \begin{cases} w(\pi) & \mu < \underline{\mu}(\pi); \\ \frac{\lambda_y}{\lambda_y+\rho}(1-\mu)\left(\frac{\psi(\mu)}{\psi(\underline{\mu}(\pi))}\right)^{\frac{\rho}{\lambda_y}} + \lambda_y\mu - c_y & \mu \ge \underline{\mu}(\pi). \end{cases}$$

Fix $\pi \in [0,1]$. By the same argument as in Lemma A1, $w_y(\pi, \mu)$ is strictly increasing in μ on $[\underline{\mu}(\pi), 1]$, and attains the value $w(\pi)$ at $\mu = \underline{\mu}(\pi)$. Thus, this function attains the maximum in the Bellman equation. $\square$

B Asymmetric Costs

This section of the appendix discusses the case where the projects have different costs. Let $c_0, c_1 > 0$ denote the research costs of x_0, x_1, respectively; assume that $c_1 < c_0 < \lambda$.

The value function of the cutoff strategy that recommends giving up for mid-range beliefs is:

$$w(\pi) = \begin{cases} \frac{\lambda c_1}{\lambda+\rho} \pi \left(\frac{\psi(\pi)}{\psi(\underline{\pi}^1)} \right)^{-\frac{\rho}{\lambda}} + \lambda(1-\pi) - c_1 & \pi \in \left[0, \underline{\pi}^1\right); \\ 0 & \pi \in \left[\underline{\pi}^1, \overline{\pi}^1\right]; \\ \frac{\lambda c}{\lambda+\rho} (1-\pi) \left(\frac{\psi(\pi)}{\psi(\overline{\pi}^1)} \right)^{\frac{\rho}{\lambda}} + \lambda\pi - c_0 & \pi \in \left(\overline{\pi}^1, 1\right]; \end{cases} \tag{12}$$

now,

$$\underline{\pi}^1 = \frac{\lambda+\rho}{\lambda} \frac{\lambda - c_1}{\lambda+\rho-c_1}, \quad \overline{\pi}^1 = \frac{\rho c_0}{\lambda(\lambda+\rho-c_0)}.$$

We have that $\overline{\pi}^1 > \underline{\pi}^1$ if and only if:

$$\lambda < \frac{\rho}{\lambda+\rho} \frac{\lambda+\rho-c_1}{\lambda+\rho-c_0} c_0 + c_1;$$

this condition reduces to $\lambda(\lambda - c) < \rho(2c - \lambda)$ (case 1) when $c_0 = c_1 =: c$, and it is satisfied if the costs are sufficiently high and the DM is sufficiently impatient.

The counterpart of (10) is:

$$w(\pi) = \begin{cases} \frac{\lambda c_0 \psi(\underline{\pi})}{\lambda+\rho} \pi \left(\frac{\psi(\pi)}{\psi(\underline{\pi})} \right)^{-\frac{\rho}{\lambda}} + \lambda(1-\pi) - c_1 & \pi \in [0, \underline{\pi}^2); \\ \lambda - c_1 - \frac{\rho c_0}{\lambda+\rho} + \frac{\lambda(c_1-c_0)}{\lambda+\rho}\pi & \pi \in [\underline{\pi}^2, \overline{\pi}^2]; \\ \frac{\lambda c_1}{(\lambda+\rho)\psi(\overline{\pi})} (1-\pi) \left(\frac{\psi(\pi)}{\psi(\overline{\pi})} \right)^{\frac{\rho}{\lambda}} + \lambda\pi - c_0 & \pi \in (\overline{\pi}^2, 1]; \end{cases} \tag{13}$$

now,

$$\underline{\pi}^2 = \frac{\lambda+\rho}{\lambda+\rho+c_1} \frac{c_0}{\lambda}, \quad \overline{\pi}^2 = \frac{(\lambda+\rho)(\lambda-c_1)+\lambda c_0}{\lambda(\lambda+\rho+c_0)}.$$

We have that $\overline{\pi}^2 \geq \underline{\pi}^2$ if and only if:

$$\lambda(\lambda - c_1) + \lambda c_0 \left[\frac{c_1 - c_0}{\lambda+\rho+c_1} \right] \geq \rho \left[\frac{\lambda+\rho+c_0}{\lambda+\rho+c_1} c_0 + c_1 - \lambda \right].$$

This condition reduces to case 2, $\lambda(\lambda - c) \geq \rho(2c - \lambda)$, when $c_0 = c_1 =: c$, and holds if the costs are sufficiently low or if the DM is sufficiently patient. The function is non-negative provided that its minimizer, which falls in the range $(\overline{\pi}^2, 1]$, is at least at large as $\frac{\rho c_0}{\lambda(\lambda+\rho-c_0)} = \overline{\pi}^1$.

Lemma B1. *The function $w : [0,1] \to \mathbb{R}$ given in* (13) *is continuously differentiable and satisfies $w'(\pi) < \frac{\lambda(c_1-c_0)}{\lambda+\rho}$ on $[0, \underline{\pi}^2)$ and $w'(\pi) > \frac{\lambda(c_1-c_0)}{\lambda+\rho}$ on $(\overline{\pi}^2, 1]$.*

Proof. Continuous differentiability follows from value matching and smooth pasting. On $[0, \underline{\pi}^2)$,

$$\begin{aligned} w'(\pi) &= \frac{\lambda c_0 \psi(\underline{\pi}^2)}{\lambda+\rho} \left(\frac{\psi(\pi)}{\psi(\underline{\pi}^2)} \right)^{-\frac{\rho}{\lambda}} \left(1 + \frac{\rho}{\lambda(1-\pi)} \right) - \lambda \\ &< \frac{\lambda c_0 \psi(\underline{\pi}^2)}{\lambda+\rho} \left(1 + \frac{\rho}{\lambda(1-\overline{\pi}^2)} \right) - \lambda = \frac{\lambda(c_1 - c_0)}{\lambda+\rho}; \end{aligned}$$

on $(\overline{\pi}^2, 1]$,

$$\begin{aligned} w'(\pi) &= -\frac{\lambda c_1}{(\lambda+\rho)\psi(\overline{\pi}^2)} \left(\frac{\psi(\pi)}{\psi(\overline{\pi}^2)} \right)^{\frac{\rho}{\lambda}} \left(1 + \frac{\rho}{\lambda\pi} \right) + \lambda \\ &> -\frac{\lambda c_1}{(\lambda+\rho)\psi(\overline{\pi}^2)} \left(1 + \frac{\rho}{\lambda\overline{\pi}^2} \right) + \lambda = \frac{\lambda(c_1 - c_0)}{\lambda+\rho}. \end{aligned}$$

This concludes the proof. □

Under symmetric costs, the case $\overline{\pi}^2 < \underline{\pi}^2$ corresponds to the case $\overline{\pi}^1 > \underline{\pi}^1$. This need not be the case under asymmetric costs. A sufficient condition for such correspondence is the following:

$$c_0 > \lambda \frac{\lambda+\rho}{\lambda+2\rho}; \quad c_1 > \frac{\lambda+\rho}{\rho}(\lambda - c_0).$$

These conditions state that, while both costs must be lower than λ, they should not be "too low."

Theorem B1 (Asymmetric costs)**.** *Assume that $c_0 > \lambda \frac{\lambda+\rho}{\lambda+2\rho}$ and $c_1 > \frac{\lambda+\rho}{\rho}(\lambda - c_0)$. If*

$$\lambda(\lambda - c_1) + \lambda c_0 \left[\frac{c_1 - c_0}{\lambda+\rho+c_1} \right] < \rho \left[\frac{\lambda+\rho+c_0}{\lambda+\rho+c_1} c_0 + c_1 - \lambda \right],$$

the optimal strategy is the same as in (1), *with $\underline{\pi}^1 = \frac{\lambda+\rho}{\lambda} \frac{\lambda-c_1}{\lambda+\rho-c_1}$ and $\overline{\pi}^1 = \frac{\rho c_0}{\lambda(\lambda+\rho-c_0)}$. If*

$$\lambda(\lambda - c_1) + \lambda c_0 \left[\frac{c_1 - c_0}{\lambda+\rho+c_1} \right] \geq \rho \left[\frac{\lambda+\rho+c_0}{\lambda+\rho+c_1} c_0 + c_1 - \lambda \right],$$

and if the minimum of (13) *is at least as high as:*

$$\frac{(\lambda+\rho)(\lambda-c_1)(\lambda+\rho+c_1) - \lambda c_0(c_0 - c_1) - \rho c_0(\lambda+\rho+c_0)}{(2\rho+\lambda)(\lambda+\rho+c_0)} \geq 0,$$

the optimal strategy is given by (2)*, with* $\underline{\pi}^2 = \frac{\lambda+\rho}{\lambda+\rho+c_1}\frac{c_0}{\lambda}$ *and* $\overline{\pi}^2 = \frac{(\lambda+\rho)(\lambda-c_1)+\lambda c_0}{\lambda(\lambda+\rho+c_0)}$.

Proof. Start with the portion of theorem regarding the counterpart of Theorem 2. Some special care needs to be taken compared to the argument behind Theorem 2, as the value function is non-monotonic on $(\overline{\pi}^2, 1]$. Define:

$$R_w^0(\pi) := \lambda\pi - c_0 + \frac{\lambda\pi\left[\lambda - c_0 - w(\pi)\right] - \lambda\pi(1-\pi)w'(\pi)}{\rho};$$
$$R_w^1(\pi) := \lambda(1-\pi) - c_1 + \frac{\lambda(1-\pi)\left[\lambda - c_1 - w(\pi)\right] + \lambda\pi(1-\pi)w'(\pi)}{\rho};$$
$$R_w^2(\pi) := \lambda - c_0 - c_1 + \frac{\lambda\left[\lambda - c_1 - \pi(c_0-c_1) - w(\pi)\right]}{\rho}.$$

We must check the following conditions:

a). On $[0, \underline{\pi}^2)$, $R_w^1(\pi) - R_w^0(\pi) > 0$ and $R_w^2(\pi) - R_w^1(\pi) < 0$.

b). On $(\overline{\pi}^2, 1]$, $R_w^1(\pi) - R_w^0(\pi) < 0$ and $R_w^2(\pi) - R_w^0(\pi) < 0$.

c). Finally, on $(\underline{\pi}^2, \overline{\pi}^2)$, $R_w^2(\pi) - R_w^0(\pi) > 0$ and $R_w^2(\pi) - R_w^1(\pi) > 0$.

a). Start with $\pi \in (0, \underline{\pi}^2)$; the counterpart of (7) is

$$-\lambda\pi(1-\pi)w'(\pi) + [\lambda(1-\pi) + \rho]w(\pi) = \lambda(1-\pi)(\rho + \lambda - c_1) - \rho c_1.$$

Proceeding as in the symmetric case, we can write:

$$R_w^1(\pi) - R_w^0(\pi) = \lambda + c_0 + c_1 - \frac{(\lambda + 2\rho)\left[\lambda - w(\pi)\right]}{\rho} + \frac{\lambda c_1 + \lambda\pi(c_0 - c_1)}{\rho}.$$

By Lemma B1, as $c_0 > c_1$, the right-hand side of the equation above is strictly decreasing in π (despite the last term being strictly increasing in π). In this case,

$$R_w^1(\pi) - R_w^0(\pi) > \lambda + c_0 + c_1 - \frac{(\lambda + 2\rho)\left[\lambda - w(\underline{\pi}^2)\right]}{\rho} + \frac{\lambda c_1 + \lambda\underline{\pi}^2(c_0 - c_1)}{\rho} = w(\underline{\pi}^2);$$

by assumption, $w(\underline{\pi}^2) > 0$. Next, take:

$$R_w^1(\pi) - R_w^2(\pi) = c_0 - \frac{(\lambda+\rho)(\lambda - c_1 - w(\pi)) - \lambda\pi(c_0 - c_1)}{\rho}.$$

From Lemma B1, it follows that the right-hand side is strictly decreasing in π. Thus,

$$R_w^1(\pi) - R_w^2(\pi) > c_0 - \frac{(\lambda+\rho)(\lambda - c_1 - w(\underline{\pi})) - \lambda\underline{\pi}(c_0 - c_1)}{\rho} = 0.$$

b). Next, consider $\pi \in (\underline{\pi}^2, \overline{\pi}^2)$. Here, we have:

$$R_w^2(\pi) - R_w^0(\pi) = \lambda(1-\pi)\frac{\lambda+\rho+c_0}{\lambda+\rho} - c_1 > \lambda(1-\overline{\pi}^2)\frac{\lambda+\rho+c_0}{\lambda+\rho} - c_1 = 0;$$
$$R_w^2(\pi) - R_w^1(\pi) = \lambda\frac{\lambda+\rho+c_1}{\lambda+\rho}\pi - c_0 > \lambda\frac{\lambda+\rho+c_1}{\lambda+\rho}\underline{\pi}^2 - c_0 = 0.$$

c). Finally, take $\pi \in (\overline{\pi}^2, 1]$; (8) is now $\lambda\pi(1-\pi)w'(\pi) + (\rho+\lambda\pi)w(\pi) = \lambda\pi(\rho+\lambda-c_0) - \rho c_0$. We have:

$$R_w^0(\pi) - R_w^1(\pi) = \lambda + c_0 - c_1 - \frac{\lambda+2\rho}{\rho}[\lambda - c_1 - w(\pi)] + \frac{\lambda(c_0-c_1)}{\rho}\pi.$$

Unlike in the symmetric case, this expression is not monotonic over the range $(\overline{\pi}^2, 1]$. However, if $\underline{w}^0$ denotes the minimum of w, we have:

$$R_w^0(\pi) - R_w^1(\pi) > \lambda + c_0 - c_1 - \frac{\lambda+2\rho}{\rho}[\lambda - c_1 - \underline{w}^0] + \frac{\lambda(c_0-c_1)}{\rho}\overline{\pi}^2;$$

by assumption, this last expression is non-negative. As for $R_w^0(\pi) - R_w^1(\pi)$, we have:

$$R_w^0(\pi) - R_w^2(\pi) = c_0 - \frac{\rho+\lambda}{\lambda}(\lambda - c_1 - w(\pi)) + \frac{\lambda}{\rho}(c_0-c_1)\pi.$$

By Lemma B1, this expression is indeed strictly increasing in π; hence,

$$R_w^0(\pi) - R_w^2(\pi) > c_0 - \frac{\rho+\lambda}{\lambda}(\lambda - c_1 - w(\overline{\pi}^2)) + \frac{\lambda}{\rho}(c_0-c_1)\overline{\pi}^2 = 0.$$

As for the rest of the proof, it remains to check that $R_w^1(\pi) > 0$ on $[0, \underline{\pi}^1)$, that $R_w^0(\pi) > 0$ on $(\underline{\pi}^1, \overline{\pi}^1)$, and that $R_w^0(\pi), R_w^1(\pi) < 0$ on $(\overline{\pi}^1, 1]$. The argument here is completely analogous to the symmetric case, since the regions on which the value function is strictly decreasing, constant, and strictly increasing, correspond to the ranges over which the strategy recommends choosing $\{x_1\}$, $\emptyset$, and $\{x_0\}$, respectively. Further details are omitted. □

The lower bound on the minimum, which is non-negative in the corresponding case of low costs or sufficient patience, helps handle the non-monotonicity of the value function on the range in which the optimal strategy recommends $\{x_0\}$. In the case where research is expensive and the DM is sufficiently impatient, this lower bound is negative, while the corresponding value function has 0 as its minimum. In this case, the additional condition is redundant.[9]

[9] No such additional condition is needed under symmetric costs.