The Boundary of Indexability in Locally Correlated Search

Dario Gori and Kourosh Khounsari

September 15, 2026

Abstract

We study when forward-looking search over locally correlated alternatives admits an index policy. A decision maker searches an ordered payoff landscape with stationary independent increments, choosing among known alternatives and unresolved regions. Our main result gives a sufficient condition for indexability. If each region is explored through a fixed local protocol, each opportunity can be valued independently, and an optimal Gittins-index policy exists. Without this restriction, indexability can fail: random-walk examples show that both the experiment chosen within a region and the ranking of regions can depend on outside opportunities. Thus, correlation need not prevent indexability; what matters is whether the decision maker can adjust local search to opportunities elsewhere. We apply the framework to R&D and career mobility. In a Brownian R&D model, discoveries generate threshold transitions among frontier research, recombination, commercialization, and return to earlier projects. Greater technological distance raises the value of recombination. In career search, correlation makes previously skipped occupations valuable after adverse outcomes, generating lateral moves that disappear in a benchmark with identical marginal job distributions but independent payoffs.

1 Introduction

This paper asks whether forward-looking search over locally correlated alternatives—such as those generated by Brownian motion (Jovanovic and Rob [1990], Callander [2011]) or random walks—admits an index policy. Indexability means that each exploratory opportunity can be assigned a priority based only on its local state. When indexability fails, the ranking of two opportunities can change after learning elsewhere. To study when this occurs, we develop a model of search over an ordered space with stationary independent increments.

Our main result shows that indexability depends on how much control the decision maker has over local search. To formalize this intuition, we introduce a *granularity constraint*: the decision maker chooses which region to explore, but the location sampled within that region is drawn from a fixed local protocol. The granularity constraint has a natural economic interpretation in some settings. In R&D, it captures imperfect short-run control over research outcomes. Management can choose which technological direction to pursue—for example, which pair of existing technologies to recombine or whether to push the frontier—but cannot perfectly choose the exact prototype produced in a single research round. The realized design is instead generated by a fixed research protocol, while the firm retains full flexibility to redirect subsequent R&D toward the most promising opportunities created by that experiment. In career mobility, the corresponding restriction arises from skill transferability. A worker can choose among career directions made accessible by past experience, but cannot move arbitrarily far in occupational space. Which jobs are currently accessible therefore depends on the worker's previous positions and the skills acquired along the way. In both settings, the decision maker chooses where to direct search, while the technology of experimentation or mobility constrains how that choice is implemented locally.

The restriction is less appropriate in settings where the decision maker has precise control over the alternative being tested. For example, a firm experimenting with prices can typically choose the price it wants to post, and a central bank choosing an interest rate can set the policy rate directly. In such applications, local flexibility is part of the economic problem. Our framework is therefore most useful in settings where the direction of search is chosen deliberately but its local implementation is constrained.

Under the granularity constraint, each search opportunity can be represented by an arm whose reward and descendants depend only on its local state. Unselected opportunities remain unchanged. The problem therefore has an Open Bandit representation. Because the problem features unbounded payoffs, Appendix A extends a bounded-rewards Open Bandit index result under a transversality condition. We prove the Indexability Theorem using this extension: each available search opportunity can be assigned a Gittins index based only on its local sufficient statistics, and an optimal policy selects an opportunity with maximal index.

Conversely, we demonstrate that the granularity constraint is effectively necessary for indexability. Once the decision maker is granted the flexibility to select optimal locations *within* an unexplored interval—essentially treating the interval as a sub-problem rather than a single arm—the resulting optimization problem is generally non-indexable. Through a series of counterexamples, we prove that under such flexibility, the optimal internal search location changes based on the state of the rest of the system, making it impossible to assign a local value to an unexplored interval.

To demonstrate the scope and utility of the framework, we apply our model to R&D and career mobility.

We first study R&D before market entry. A startup searches over a Brownian technological landscape and can either commercialize a product it has already developed, expand the technological frontier, or recombine two previously developed technologies by exploring the interval between them. The indexability result reduces this problem to a comparison among the indices of these op-

portunities. The model generates endogenous changes in the direction of R&D. Following a frontier discovery, sufficiently poor realizations lead the firm back to recombining old projects, or to enter the market with the best developed technology; sufficiently profitable realizations induce market entry with the newly developed product; while intermediate discoveries sustain further frontier R&D. Conditional on the economic value of the discovery, a sufficiently large technological jump makes the interval produced by frontier expansion the most valuable next project and induces recombination. A similar threshold structure governs what happens after a recombination experiment. The resulting dynamics provide a mechanism for cycles between broad exploration and more focused development within a single firm, and connect naturally to evidence on phased technological search and recombinant uncertainty. Indeed, these dynamics have close empirical counterparts. Randle and Pisano [2021] document a pattern in which firms producing breakthrough inventions first explore unfamiliar technological terrain and subsequently shift toward a more focused use of the knowledge accumulated during exploration. In our model, frontier expansion creates the technological anchors that make such a later recombination phase valuable. Fleming [2001] similarly shows that unfamiliar components and new combinations generate more dispersed invention outcomes and are less useful on average. Our model provides a selection-based interpretation of these findings. Greater technological distance increases uncertainty and therefore the option value of experimentation: the firm can abandon poor realizations while retaining the upside from favorable discoveries. As a result, the firm is willing to undertake more distant projects at lower average profitability. Thus, unfamiliar combinations may appear less useful on average not because distance mechanically lowers the expected payoff of a given project, but because greater option value lowers the profitability threshold at which such projects are selected. These papers do not identify the threshold mechanism developed here, but they document precisely the kinds of search dynamics that the model rationalizes.

Our second application studies a worker's career trajectory across a landscape of correlated occupations modeled as a discrete random walk. In this setting, the worker is subject to a local exploration constraint representing the friction of human capital transferability: a worker cannot leap across the skill space but must take "stepping-stone" jobs to unlock new career frontiers. The simplicity of the environment allows us to characterize the optimal policy completely. The worker expands the career frontier while its match quality remains close to the best match previously found; after sufficiently adverse news, she instead experiments with a previously skipped occupation whose value is informed by successful jobs on both sides; and she eventually settles when neither outward search nor such lateral options are sufficiently attractive. To isolate the role of correlation, we compare this policy with a benchmark that preserves the same occupational landscape, the same mobility constraint, and the same marginal distribution of match quality at every position, but makes jobs independent. In that benchmark, skipped occupations are never revisited: the worker either continues outward or settles. Correlation therefore creates a stock of latent lateral options that can become valuable after adverse career shocks, providing a mechanism for the path dependence and retraining toward related fields documented in the empirical career-mobility literature. Indeed, Humlum et al. [2025] show that adverse career shocks can induce workers to redirect human-capital investment toward fields related to their previous experience, while Bruhn et al. [2025] document persistent effects of early occupational assignments on subsequent employment in the same or related occupations. Our model provides one mechanism for these patterns. Successful past jobs make the worker optimistic about related but previously untried occupations, so that a deterioration at the current career frontier can make such a lateral move optimal. The independent-marginal benchmark isolates this mechanism: when jobs retain the same marginal payoff distributions but no longer convey information about one another, skipped occupations are never revisited. The empirical evidence does not establish correlated learning as the unique mecha-

nism, but it is consistent with the path dependence and shock-induced lateral adjustment generated by the model.

1.1 Literature review

The Brownian framework has been introduced in Jovanovic and Rob [1990] and has been later revived by Callander [2011]. Both papers examine the process of discovery when the realization of a Brownian motion generates payoffs over a continuous action space. In Callander [2011], a sequence of myopic players incrementally explores the policy space, with each agent choosing actions near previously tried ones and observing locally informative payoffs. This setting captures the idea of learning by trial and error in an unpredictable environment, but its complex structure makes the analysis of forward-looking behavior analytically challenging. Inspired by this problem, we propose a tractable analog for forward-looking search with an infinitely-lived player. Our framework also complements the dynamic social experimentation analysis of Garfagnini and Strulovici [2016]. In their model, a sequence of risk-neutral agents living for two periods explores the Brownian environment introduced by Callander [2011]. Crucially, moving from two-period players to an infinitely-lived decision maker changes what is relevant for optimal search. For instance, Garfagnini and Strulovici [2016] demonstrates that the value of an interval is determined solely by its intrinsic features and the maximum payoff previously encountered. This result is intuitive within a two-period horizon: an interval's utility is limited to its potential to yield a payoff exceeding the current maximum in a single step. However, as our examples illustrate, extending the planning horizon beyond two periods allows for the sequential combination of search across multiple intervals. Consequently, if no additional constraint is imposed, the value of a specific interval is no longer local but becomes a function of the global state of the system. Other contributions to forward-looking search in a locally correlated environment with an infinitely lived player are Urgun and Yariv [2025] and Wong [2025]. In these studies, agents are restricted to *contiguous* search. Their decision-making is twofold: they must determine whether to terminate or persist in their search and, conditional on continuing, select an optimal search intensity. Consequently, their analysis focuses on how history shapes the optimal search speed, rather than addressing the problem of optimal location selection. Our analysis is also related to Fershtman and Pavan [2026], who study experimentation when the decision maker can expand her consideration set as well as explore alternatives already available to her. Fershtman and Pavan interpret consideration-set expansion as a branching arm and develop a recursive index for the value of search. In our environment, branching instead arises from learning about a correlated payoff landscape: experimenting with an unresolved region replaces it with the local opportunities generated by the new observation. Thus, both papers exploit the machinery developed for branching bandits, but our focus is on the conditions under which correlation across primitive alternatives remains compatible with an index decomposition. Closest to our question of where a forward-looking decision maker should search are Malladi [2022] and Banchio and Malladi [2026], who study long-lived searchers that can freely choose where to experiment in an ordered space in which nearby alternatives have related payoffs. Their approach is robust rather than Bayesian: the decision maker evaluates continuation policies against the worst-case payoff landscape consistent with her observations. We instead study the problem under Bayesian beliefs generated by a stochastic payoff landscape.

2 The General Model of Locally Correlated Search

2.1 Search Environment

The decision maker (DM) searches over a one-sided spatial domain $\mathcal{X}$, where either $\mathcal{X} = \mathbb{R}_+$ or $\mathcal{X} = \mathbb{N}_0$. The continuous domain $\mathbb{R}_+$ will be used in the R&D application, whereas the discrete domain $\mathbb{N}_0$ will be used in the career-mobility application.

Each location $x \in \mathcal{X}$ has an unknown payoff Ψ_x. Before search begins, Nature draws an entire payoff landscape $\psi(x)$ from a stochastic process $\Psi = \{\Psi_x : x \in \mathcal{X}\}$.[1] Once drawn, the landscape remains fixed over decision time. The DM knows the probability distribution of Ψ and the initial payoff $\Psi_0 \equiv \bar{y}_0$, but not the realized landscape.

Assumption 1. *For every finite sequence $x_0 < x_1 < \cdots < x_n$ with $x_i \in \mathcal{X}$, the increments*

$$\Psi_{x_1} - \Psi_{x_0}, \ldots, \Psi_{x_n} - \Psi_{x_{n-1}}$$

are mutually independent.

Assumption 1 concerns increments over disjoint spatial segments, not payoff levels. Payoffs at nearby alternatives remain statistically correlated because they share the increments accumulated before their paths separate.

Crucially, this assumption implies that learning is local. Once the payoffs at the two endpoints of an unexplored interval have been observed, observations outside that interval do not affect beliefs about its unresolved interior. Similarly, the distribution beyond the rightmost observed location depends on earlier observations only through the payoff at that frontier. More formally, conditional on the observed endpoint payoffs, the remaining uncertainty in distinct intervals and beyond the frontier is independent.

Assumption 2. *For any $x < y$ and $x' < y'$ in $\mathcal{X}$ such that $y - x = y' - x'$, we have*

$$\Psi_y - \Psi_x \overset{d}{=} \Psi_{y'} - \Psi_{x'}.$$

Assumption 2 makes the landscape spatially homogeneous. Together with Assumption 1, it implies that an unexplored bounded interval can be summarized by its endpoint payoffs $y_L = \psi(x_L)$, $y_R = \psi(x_R)$, and its length $\ell = x_R - x_L$, rather than by the absolute coordinates x_L, x_R. Similarly, uncertainty beyond the frontier depends on the current frontier payoff and on the relative step taken beyond it, but not on the frontier's absolute location.

When $\mathcal{X} = \mathbb{N}_0$, assumptions 1 and 2 characterize a random walk. When $\mathcal{X} = \mathbb{R}_+$, adding stochastic continuity implies that $\Psi - \bar{y}_0$ is a Lévy process up to the usual càdlàg modification. Brownian motion is the continuous process used in the R&D application.[2]

[1] Formally, Ψ is defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and the map $(x, \omega) \mapsto \Psi_x(\omega)$ is $\mathcal{B}(\mathcal{X}) \otimes \mathcal{F}$-measurable.

[2] The indexability result presented below relies solely on the aforementioned properties of conditional independence and spatial homogeneity; it does not require sample-path continuity. Consequently, the theorem extends to discontinuous payoff functions, such as those governed by a Poisson process.

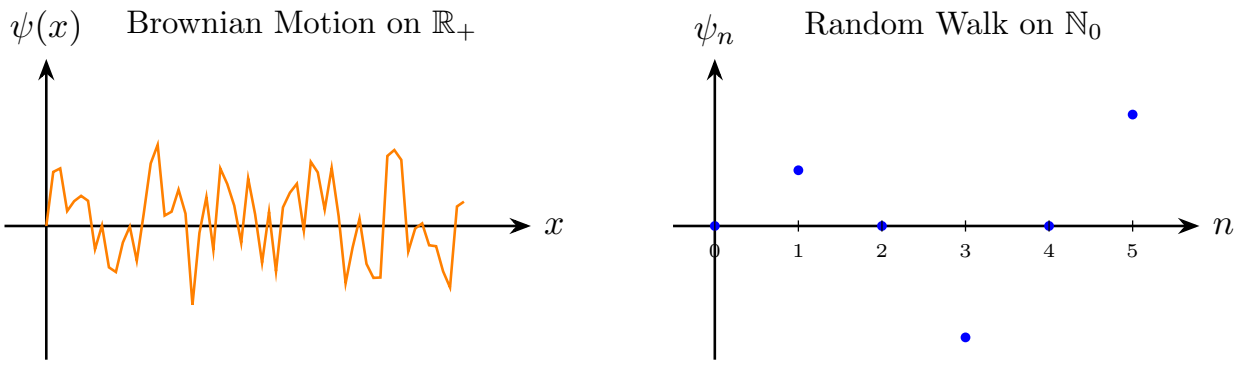

Figure 1: Examples of stationary-independent-increment payoff landscapes on the two domains used in the paper: Brownian motion on $\mathbb{R}_+$ and a random walk on $\mathbb{N}_0$.

2.2 Information Structure

Time is discrete and indexed by $t \in \mathbb{N}_0$. Before period 0, the DM observes the payoff at the origin, $\psi(0) = \bar{y}_0$. Set $(x_{-1}, y_{-1}) \equiv (0, \bar{y}_0)$. At the beginning of period t, a realized history is the ordered vector $h_{t-1} \equiv \big((x_s, y_s)\big)_{s=-1}^{t-1}$, where $x_s \in \mathcal{X}$ is the location observed in period s and $y_s = \psi(x_s)$ is its realized payoff.

The ordered history records the chronology of observations and may contain repeated visits to the same location. Let

$$X_{t-1} \equiv \{x_s : s = -1, \ldots, t-1\}$$

be the finite set of distinct locations observed before period t. The corresponding informational state is

$$S_{t-1} \equiv \{(x_s, y_s) : s = -1, \ldots, t-1\}$$

Because the payoff landscape is fixed over decision time, revisiting a location reveals the same payoff and leaves X_{t-1} and S_{t-1} unchanged.

Since X_{t-1} is nonempty and finite, the rightmost observed location

$$x^{\mathrm{r}}_{t-1} \equiv \max X_{t-1}$$

is well defined. We call x^{r}_{t-1} the *right frontier*, and denote its observed payoff by

$$y^{\mathrm{r}}_{t-1} \equiv \psi(x^{\mathrm{r}}_{t-1}).$$

The right frontier is a spatial boundary and need not be the location with the highest observed payoff. The associated unexplored tail is

$$(x^{\mathrm{r}}_{t-1}, \infty) \cap \mathcal{X}.$$

2.3 Granularity Constraint and Search Policy

We first define the search opportunities generated by a finite observed set X_{t-1}. The set of bounded unexplored intervals[3] is

$$\mathcal{I}_{t-1} = \mathcal{I}(X_{t-1}) \equiv \{(x_L, x_R) \in X_{t-1} \times X_{t-1} : x_L < x_R,\ X_{t-1} \cap (x_L, x_R) = \varnothing,\ \mathcal{X} \cap (x_L, x_R) \neq \varnothing\}.$$

[3]Throughout the paper, we use the terms *bounded interval* and *gap* interchangeably.

Thus, $(x_L, x_R) \in \mathcal{I}_{t-1}$ if its endpoints are consecutive among the observed locations and the interval contains at least one feasible unobserved location. The final condition is automatic when $\mathcal{X} = \mathbb{R}_+$, but excludes adjacent observed positions when $\mathcal{X} = \mathbb{N}_0$.

The feasible action set associated with X_{t-1} is thus

$$\mathcal{A}(X_{t-1}) = X_{t-1} \cup \mathcal{I}_{t-1} \cup \{(x^r_{t-1}, \infty)\}.$$

The three action types have the following interpretations:

$$\begin{aligned} a_t = x \in X_{t-1} \quad &: \text{ select a previously observed alternative,} \\ a_t = (x_L, x_R) \in \mathcal{I}_{t-1} \quad &: \text{ explore the bounded interval between } x_L \text{ and } x_R, \\ a_t = (x^r_{t-1}, \infty) \quad &: \text{ explore beyond the right frontier.} \end{aligned}$$

An admissible pure search policy is a sequence $\Sigma = (\Sigma_t)_{t \geq 0}$ of measurable decision rules such that, for every feasible history h_{t-1},

$$a_t = \Sigma_t(h_{t-1}) \in \mathcal{A}(X_{t-1})$$

for every $t \geq 0$.[4]

The coordinate explored after an action choice is sampled according to the following rule.

Assumption 3. *There exist fixed families of probability distributions $\sigma^{\mathrm{g}}\big(\cdot \;\big|\; \ell, y_L, y_R\big)$ and $\sigma^{\mathrm{f}}\big(\cdot \;\big|\; y\big)$ governing search within bounded intervals and beyond the frontier, with the following properties.*[5]

For every bounded-interval state (ℓ, y_L, y_R), $\sigma^{\mathrm{g}}\big((0,\ell) \cap \mathcal{X} \;\big|\; \ell, y_L, y_R\big) = 1$. If $a_t = (x_L, x_R) \in \mathcal{I}_{t-1}$, then a relative displacement

$$\Delta_t \sim \sigma^{\mathrm{g}}\left(\cdot \;\middle|\; \ell, y_L, y_R\right)$$

is drawn and $x_t = x_L + \Delta_t$.

For every frontier state y^r, $\sigma^{\mathrm{f}}\big((0,\infty) \cap \mathcal{X} \;\big|\; y^r\big) = 1$. If $a_t = (x^r, \infty)$, then a relative displacement

$$\Delta_t \sim \sigma^{\mathrm{f}}\left(\cdot \;\middle|\; y^{\mathrm{r}}\right)$$

is drawn and $x_t = x^{\mathrm{r}}_{t-1} + \Delta_t$.

If $a_t = x \in X_{t-1}$, then $x_t = x$.

The auxiliary randomizations used to generate exploratory draws are independent across periods and, conditional on the displayed local state, independent of the unresolved payoff landscape.

The pair $\sigma \equiv (\sigma^{\mathrm{g}}, \sigma^{\mathrm{f}})$ therefore describes the fixed local sampling rules of the environment. The DM chooses which bounded interval or frontier to explore, but not the exact location sampled within it. The sampling distribution may depend on the local state of the selected opportunity, but not on calendar time, on other available opportunities, or on any other feature of the history.

After the period-t location is determined, the payoff $y_t = \psi(x_t)$ is observed and the state evolves according to

$$X_t = X_{t-1} \cup \{x_t\}, \qquad S_t = S_{t-1} \cup \{(x_t, y_t)\}.$$

[4] In the stochastic construction used in the proof of Theorem 1, the policy and the granularity constraint introduced below induce random histories, random observed sets, and informational states. We introduce those random objects only when they are needed in the proof.

[5] Formally, σ^{g} and σ^{f} are Markov kernels. In particular, they are measurable in their conditioning arguments.

Under Assumption 3, every gap or frontier action reveals a previously unobserved location. Moreover, given S_{t-1} and x_t, the selected action is uniquely identified: an old location corresponds to selection of a known alternative, a new point below the frontier belongs to a unique bounded gap, and a new point above the frontier corresponds to frontier exploration. It is therefore unnecessary to record actions separately in the observational history.

The restriction in Assumption 3 is precisely what makes each exploratory opportunity locally self-contained: once an interval or the frontier is selected, the distribution of the resulting observation depends only on that opportunity's local state. This property will allow us to value the available opportunities separately in the indexability analysis below.

We retain the full class of history-dependent admissible policies. The indexability theorem below establishes that an optimal policy can be implemented by comparing the local states of the currently available opportunities.

2.4 Flow Payoffs and the Search Objective

The flow payoff may depend on the nature of the selected alternative: the DM may choose a previously observed alternative whose payoff is known, or an exploratory opportunity whose payoff is not yet known. The distinction has a different economic interpretation across applications. In the R&D application, selecting a known technology means putting it on the market, whereas selecting a bounded interval or the frontier means developing a new technology, which cannot be marketed in the same period. Define

$$\chi_t \equiv \mathbf{1}\{a_t \in X_{t-1}\}.$$

Thus, $\chi_t = 1$ when the decision maker selects a previously observed alternative and $\chi_t = 0$ when she selects a bounded gap or the frontier. Since

$$\mathcal{A}(X_{t-1}) = X_{t-1} \cup \mathcal{I}_{t-1} \cup \{(x^r_{t-1}, \infty)\},$$

we have

$$1 - \chi_t = \mathbf{1}\{a_t \in \mathcal{I}_{t-1} \cup \{(x^r_{t-1}, \infty)\}\}.$$

Let

$$u : \mathbb{R} \times \{0, 1\} \longrightarrow \mathbb{R}$$

be the measurable flow-utility function, where the second argument records whether the selected payoff was known before the action was taken. Every exploratory activation incurs the cost $c \geq 0$. The realized period payoff is therefore

$$r_t = u\big(y_t, \chi_t\big) - (1 - \chi_t)c.$$

The decision maker solves[6]

$$V(S_{-1}) = \sup_{\Sigma} \mathbb{E}^{\Sigma,\sigma}\left[\sum_{t=0}^{\infty} \delta^t \big(u(y_t, \chi_t) - (1 - \chi_t)c\big) \,\middle|\, S_{-1}\right], \qquad \delta \in (0, 1).$$

This formulation contains the two applications studied below. In the R&D application,

$$u(y, \chi) = \chi y, \qquad c > 0,$$

[6]The expectation is taken with respect to the probability law over search histories induced jointly by the stochastic payoff landscape, the policy Σ, and the sampling rules σ, starting from S_{-1}. The formal stochastic construction is given in the proof of Theorem 1.

so that operating a known product yields its flow profitability, whereas conducting an R&D experiment yields the current payoff $-c$.[7] In the career mobility application,

$$u(y, \chi) = y, \qquad c = 0,$$

because the worker receives the payoff of the selected occupation in the current period whether or not that occupation was previously visited.

The following regularity condition allows unbounded payoffs while controlling both total discounted rewards and their distant tails.

Assumption 4. *For every reachable finite informational state S,*

$$\sup_{\Sigma} \mathbb{E}^{\Sigma,\sigma}\left[\sum_{t=0}^{\infty} \delta^t |r_t| \,\middle|\, S\right] < \infty$$

and

$$\lim_{T\to\infty} \sup_{\Sigma} \mathbb{E}^{\Sigma,\sigma}\left[\sum_{t=T}^{\infty} \delta^t |r_t| \,\middle|\, S\right] = 0.$$

Both suprema are over all admissible continuation policies from S.[8]

The first requirement ensures that the global value and the values of the single-opportunity descendant problems are finite. The second makes rewards from sufficiently distant periods negligible uniformly over continuation policies. This uniformity permits the passage from bounded-reward index results to the unbounded rewards allowed here.

Remark 1. *A convenient sufficient condition for Assumption 4 is that, for every reachable finite state S, there exist constants $C_S < \infty$ and $q \geq 0$ such that*

$$\sup_{\Sigma} \mathbb{E}^{\Sigma,\sigma}[|u(y_t, \chi_t)| \,|\, S] \leq C_S(1+t)^q \qquad \text{for every } t \geq 0.$$

Indeed, for every $T \geq 0$,

$$\sup_{\Sigma} \mathbb{E}^{\Sigma,\sigma}\left[\sum_{t=T}^{\infty} \delta^t |r_t| \,\middle|\, S\right] \leq \sum_{t=T}^{\infty} \delta^t \big[C_S(1+t)^q + c\big].$$

The right-hand side is finite at $T = 0$ and converges to zero as $T \to \infty$, verifying both requirements. The R&D and career-mobility applications satisfy this condition with $q = 1$.

2.5 Indexability and Optimal Search

At the beginning of each period, the informational state induces a finite collection of known alternatives and exploratory opportunities. A direct dynamic-programming approach would assign a value to this entire collection. This is potentially cumbersome because exploring a bounded gap

[7] The common cost c can be replaced by type-specific or state-dependent costs without affecting the indexability result, provided those costs depend only on the local state and local realization of the selected opportunity.

[8] The requirements are imposed at every reachable finite state, rather than only at the initial state. They therefore also hold for each single-opportunity descendant problem: such a problem can be embedded in a global continuation problem by following the selected arm and its descendants while ignoring all outside opportunities. The next section clarifies why such single-opportunity descendant problems are relevant.

or the frontier can generate several descendant opportunities. The number and composition of the objects entering the global state may therefore grow over time.

The granularity constraint yields a stronger structure. Conditional on the local state of the selected opportunity, its expected current reward and the distribution of the opportunities that it generates do not depend on the states of the other available opportunities. Moreover, opportunities that are not selected remain unchanged. The constraint does not eliminate correlation across nearby alternatives. It ensures instead that correlation can be contained within the descendant process generated by the selected opportunity.

This structure is precisely the one exploited by an *Open Bandit Process*. In an open bandit problem, the decision maker faces a collection of arms and activates one of them in each period. Activation produces a current payoff and replaces the selected arm with a random collection of descendant arms, while all unselected arms remain unchanged. In our setting, the correlation among nearby physical alternatives is absorbed into the descendant process generated by the selected opportunity: once that opportunity is specified, its payoff and descendants can be described independently of the other opportunities currently available. We therefore rewrite the search problem in open-bandit terms and then apply the corresponding Gittins-index result. The first step is to associate each available search opportunity with a local arm state.

Local arm states. The actions defined in Section 2.3 identify the physical opportunities currently available to the DM: a known location, a bounded interval, or the frontier. To represent these opportunities as arms, we associate each action with a *local arm state*. The arm state records the information about that opportunity that determines both its current expected payoff and the distribution of the opportunities generated when it is explored. It therefore strips away features such as absolute location and the states of other available opportunities that are irrelevant for the local evolution of the selected arm. Distinct physical actions may consequently have the same arm state.

For a known alternative, the local state is simply its observed payoff. For a bounded interval, it consists of its length and the payoffs at its two endpoints. For the frontier, it is the payoff at the rightmost observed location. These are the objects to which the Gittins index will eventually be assigned.

Let

$$\mathcal{L}_{\mathcal{X}} \equiv \{x_R - x_L : x_L, x_R \in \mathcal{X},\ x_L < x_R,\ (x_L, x_R) \cap \mathcal{X} \neq \varnothing\}$$

denote the set of feasible lengths of bounded gaps. Thus,

$$\mathcal{L}_{\mathbb{R}_+} = (0, \infty), \qquad \mathcal{L}_{\mathbb{N}_0} = \{2, 3, \ldots\}.$$

The local arm-state space is the disjoint union

$$\mathcal{Z} = \mathcal{Z}_{\mathrm{s}} \sqcup \mathcal{Z}_{\mathrm{g}} \sqcup \mathcal{Z}_{\mathrm{f}},$$

where

$$\mathcal{Z}_{\mathrm{s}} \equiv \{\mathsf{S}(y) : y \in \mathbb{R}\}, \quad \mathcal{Z}_{\mathrm{g}} \equiv \left\{\mathsf{G}(\ell, y_L, y_R) : \ell \in \mathcal{L}_{\mathcal{X}},\ (y_L, y_R) \in \mathbb{R}^2\right\}, \quad \text{and} \quad \mathcal{Z}_{\mathrm{f}} \equiv \{\mathsf{F}(y) : y \in \mathbb{R}\}.^{9}$$

The three arm types have the following interpretations. A safe arm $\mathsf{S}(y)$ represents an available alternative whose payoff y is known. A gap arm $\mathsf{G}(\ell, y_L, y_R)$ represents a bounded unexplored region

[9] Each component is endowed with the Borel structure induced by its parameterization, and $\mathcal{Z}$ is endowed with the corresponding disjoint-union σ-algebra.

of length ℓ whose endpoint payoffs are y_L and y_R. A frontier arm $\mathsf{F}(y)$ represents the unexplored tail beyond a right-frontier location whose observed payoff is y.

More precisely, for a given informational state S, define, for every action $a \in \mathcal{A}(X)$, its associated local arm state by

$$z(a;S) = \begin{cases} \mathsf{S}\big(\psi(x)\big), & a = x \in X, \\ \mathsf{G}\big(x_R - x_L, \psi(x_L), \psi(x_R)\big), & a = (x_L, x_R) \in \mathcal{I}(X), \\ \mathsf{F}\big(\psi(x^r)\big), & a = (x^r, \infty). \end{cases}$$

The global portfolio of arms induced by S is the finite multiset[10]

$$\mathbf{M}(S) \equiv \biguplus_{a \in \mathcal{A}(X)} \{z(a;S)\}.$$

For every $z \in \mathcal{Z}$, let

$$R(z)$$

denote the *expected* one-period payoff from activating an arm in state z. In particular, $R(\mathsf{S}(y)) = u(y,1)$, whereas for a gap or frontier state, $R(z)$ is the conditional expectation of $u(y,0) - c$ under the corresponding granularity constraint and posterior distribution.

Let

$$P(\,\cdot\,|z)$$

denote the probability law of the finite multiset of descendant arms generated when an arm in state z is activated. A safe arm reproduces itself. A gap arm generates the newly revealed safe arm and any nonempty sub-gaps created by the observation. A frontier arm generates the newly revealed safe arm, the new frontier arm, and the intervening bounded gap whenever that gap contains a feasible unobserved alternative. All arms that are not activated remain unchanged.

Assumptions 1, 2, and 3 imply that both $R(z)$ and $P(\,\cdot\,|z)$ depend only on z and the model primitives. They do not depend on the remaining arms in the global portfolio.

The single-arm descendant problem. The Gittins index assigns a value to each available opportunity by considering only what can be generated from that opportunity itself. We therefore isolate the descendant family of a single initial arm and study the best discounted payoff obtainable before that family is abandoned.

Fix an initial arm state $z \in \mathcal{Z}$. To value this arm, consider an auxiliary process that begins with z as its only available arm:

$$\mathbf{M}^z_{-1} = \{z\}.$$

At local date $n \geq 0$, a policy π selects one arm $z^\pi_n \in \mathbf{M}^z_{n-1}$. The selected arm is removed and replaced by a random multiset $\mathbf{D}_n$ drawn according to $P\big(\,\cdot \;\big|\; z^\pi_n\big)$:

$$\mathbf{M}^z_n = \big(\mathbf{M}^z_{n-1} \ominus \{z^\pi_n\}\big) \uplus \mathbf{D}_n,$$

where $\ominus$ removes one copy of the selected arm.

[10] A multiset is a collection in which the same element may appear more than once. This is useful here because two distinct physical opportunities can share the same local arm state. The symbol $\biguplus$ denotes multiset union: it combines the arms associated with all available actions while preserving their multiplicities. Thus, if two different actions generate the same state z, then $\mathbf{M}(S)$ contains two copies of z rather than collapsing them into a single element.

Let $\Pi(z)$ denote the set of admissible policies for this descendant process. Thus, a policy in $\Pi(z)$ can select only the initial arm z and arms subsequently generated from it. For each $\pi \in \Pi(z)$, let $\mathcal{T}(z,\pi)$ denote the set of strictly positive stopping times with respect to the natural filtration of the descendant process under π.

The normalized Gittins index of z is[11]

$$I(z) = \sup_{\substack{\pi\in\Pi(z)\\ \tau\in\mathcal{T}(z,\pi)}} \frac{\mathbb{E}_z^\pi\left[\sum_{n=0}^{\tau-1}\delta^n R(z_n^\pi)\right]}{\mathbb{E}_z^\pi\left[\sum_{n=0}^{\tau-1}\delta^n\right]}.$$

The numerator is the expected discounted reward obtained from the family generated by z before that family is retired. The denominator is the expected discounted amount of time allocated to the same family. Hence $I(z)$ is the highest discounted average payoff attainable by first activating z, managing only the descendants that it creates, and choosing when to abandon that descendant process.[12]

In particular, outside opportunities do not enter the calculation of $I(z)$. If z is a gap arm, the safe alternatives located at its two endpoints are not included in $\mathbf{M}_{-1}^z$. Those alternatives are separate safe arms in the global portfolio. The index of the gap includes only the newly revealed safe arms and sub-gaps generated by experiments conducted inside the gap.

This construction assigns each local arm state a value independently of the rest of the global portfolio. The next theorem shows that, under the granularity constraint, these single-arm values are sufficient to solve the original search problem: at every history, it is optimal to select an available opportunity with the highest index.

Theorem 1. *Under Assumptions 1, 2, 3, and 4, $I(z)$ is finite for every reachable local arm state $z \in \mathcal{Z}$. Moreover, any admissible policy satisfying, at every history,*

$$a_t \in \arg\max_{a\in\mathcal{A}(X_{t-1})} I\big(z(a;S_{t-1})\big)$$

is optimal.

The proof, given in Appendix A, first maps the search problem into an Open Bandit Process. For bounded approximations of the expected reward function R, index optimality follows from Wu and Zhou [2013, Theorem 3.1].[13] Lemma 5 then extends index optimality to the original, unbounded rewards problem satisfying Assumption 4.

To see the reduction delivered by the theorem, let $\mathbf{m}$ be a finite multiset of available arm states and let $W(\mathbf{m})$ denote its global continuation value. A direct Bellman equation would take the form

$$W(\mathbf{m}) = \max_{z\in\mathbf{m}} \{R(z) + \delta\mathbb{E}\left[W\left((\mathbf{m}\ominus\{z\})\uplus\mathbf{D}(z)\right)\right]\},$$

[11] We suppress the dependence of I on the landscape law, the local protocols, the utility and cost functions, and the discount factor.

[12] Under Assumptions 1–4, the supremum is attained for every z; see Corollary 2 in Appendix A.

[13] The index above is a normalized version of the index used by Wu and Zhou [2013]. Indeed, for every positive stopping time τ,

$$1-\mathbb{E}[\delta^\tau] = (1-\delta)\mathbb{E}\left[\sum_{n=0}^{\tau-1}\delta^n\right].$$

Consequently, $I(z)$ equals $1-\delta$ times their index, and the two indices generate the same ranking.

where $\mathbf{D}(z)$ is the random multiset generated by activating z. Because one arm can generate several descendants, the cardinality and composition of $\mathbf{m}$ are not bounded by a fixed-dimensional state description.

Theorem 1 eliminates the need to solve this Bellman equation jointly over all finite portfolios. Each local index $I(z)$ is computed from the single-arm descendant problem, and the global decision is obtained by comparing the resulting scalar values. The local valuation problem remains dynamic—the index incorporates all future descendants of the initial arm—but it can be solved independently of the other opportunities currently available.

Indexability does not imply that the ranking of physical opportunities remains unchanged over time. An activated arm is replaced by descendants with new local states, and these descendants can have different indices. The result is that an idle opportunity retains the same index when learning occurs elsewhere, so the same local index function can be used after every history.

3 Flexible Local Search and the Failure of Indexability

The previous section establishes that fixed local protocols are sufficient for indexability. We now show that this conclusion need not survive when the DM can choose the sampling location within a gap as a function of the global state. Even in simple finite random-walk environments, flexible local search can make both the optimal internal experiment and the ranking of distinct gaps depend on outside opportunities.

The counterexamples below do not assert that every model with flexible local search is non-indexable. They establish the narrower result needed here: once the local transition rule can respond to the global state, the structural separability used in Theorem 1 is no longer guaranteed.

To formally demonstrate this breakdown, we construct counterexamples using a discrete environment. We assume the underlying stochastic process is a standard symmetric Random Walk with Bernoulli increments whose support is $\{-1, 1\}$. In the first three examples, the decision maker is risk-neutral, $U(y) = y$, in the fourth we allow for the utility used in Callander [2011], $U(y) = -y^2$. To model internal search flexibility, we introduce a relaxed cost function where search cost is zero if the step size is less than or equal to a boundary k: $|x_t - x| \leq k$ for some $x \in X_{t-1}$, and infinite otherwise. This setup, with $k > 2$, grants the decision maker the choice among multiple interior points within a gap, contrasting with the baseline model where the sampling distribution is fixed. In the examples, $\delta = 0.99$.

In the language of multi-armed bandits, relaxing the granularity constraint transforms the problem from a standard Open Bandit Process into a *Bandit Superprocess* [Nash, 1973, Glazebrook, 1982]. In a superprocess, "arms" are themselves Markov Decision Processes. As established in the literature, such problems are indexable only under restrictive conditions. We present counterexamples demonstrating that locally correlated search with strategic flexibility violates these conditions, leading to a breakdown of the separability required for index policies.[14]

3.1 Example 1

[14]The following examples are constructed with the help of Mathematica. Observe that the computations performed with the software are algebraic.

First, we construct a counterexample where the optimal internal policy for a gap is non-monotonic in the value of the outside option, ρ. ρ could be interpreted as the maximum value of the random walk found so far.

Consider a gap with 5 unknown locations, bounded by a known value of 0 on the left and 2 on the right. Let the unknown locations be $\{1, 2, 3, 4, 5\}$. We compare the optimal internal choice as ρ varies. Observe that the DM is allowed to choose the outside option $\frac{\rho}{1-\delta}$. If the problem were indexable, the DM would assign a fixed "Gittins index" to the gap, and the decision to exploit the gap versus the outside option would depend on a simple comparison. However, we find that the optimal *choice of action within the gap* depends on ρ in a complex, non-monotonic way:

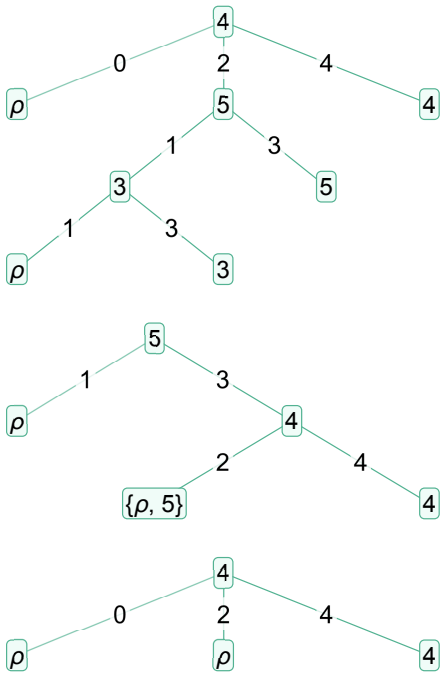

Figure 2: The optimal strategy for $\rho = 2.5, 3$, and 3.5, respectively. Nodes are the optimal points to explore, and the edges' labels from each node to down are the corresponding realization of that point's value.

Regime 1, $\rho = 2.5$: When the outside option is low, the DM prefers Action 4. The relative cost of failure is low, so the DM maximizes the option value of finding a high payoff deep in the gap.

Regime 2, $\rho = 3.0$: When the outside option increases to an intermediate level, the DM switches to Action 5. Action 5 offers the highest probability of incrementally exceeding ρ without the risk of a deep plunge.

Regime 3, $\rho = 3.5$: When the outside option is high, the DM switches back to Action 4. The conservative Action 5 cannot mathematically generate a payoff high enough to justify switching away from ρ. The DM is forced to accept higher variance ("gambling for resurrection") because it is the only way to generate a realization in the tail of the distribution that exceeds ρ.

3.2 Example 2

Consider now the problem where the DM can select a point in $\{0, \ldots, 9\}$, or the outside option ρ. The outcomes at $0, 2$, and 9 are known: $\psi(0) = \psi(2) = 0$ and $\psi(9) = -3$. As the picture shows, the right interval is chosen when the outside option is low or high, while the left interval is chosen when the outside option is intermediate. From the middle tree, we can observe where the free internal choice breaks indexability: when the DM chooses point 1 (left interval), he subsequently explores the right interval following any outcome realization; however, he chooses point 4 after a bad realization and point 3 after a good outcome.

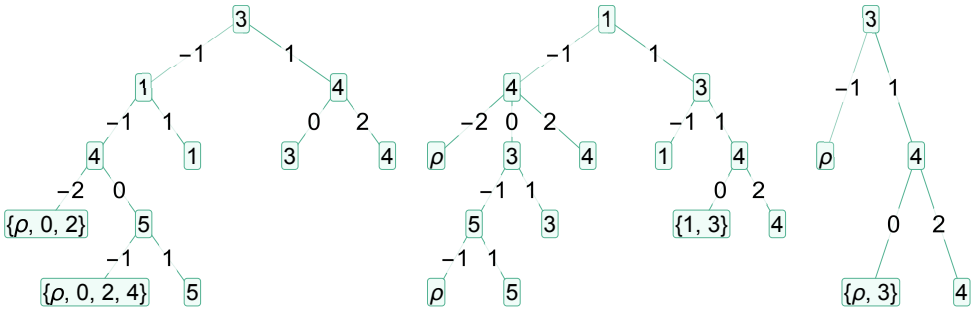

Figure 3: The optimal strategy for $\rho = 0, 0.5$, and 1, respectively. Nodes are the optimal points to explore, and the edges' labels from each node to down are the corresponding realization of that point's value.

3.3 Example 3

In this example we consider the choice between points in $\{0, \ldots, 12\}$, without an external outside option. It is as if the DM has explored the frontier and the outcomes at $13, 14, \cdots$ realized to be extremely bad. The DM knows that $\psi(0) = \psi(2) = \psi(4) = 0$, and that $\psi(12) = -4$. We compare two cases: one where $\psi(1) = -1$ and another where $\psi(1)$ is not known. In these cases, the maximum number found so far is the same, but in the first case, the DM chooses to fill the narrow gap between 4 and 12, and in the second case, the DM chooses to fill the wide gap. So, the decision between the wide gap and the narrow gap depends not only on the maximum found so far, but also on the number of available narrow gaps suggesting that there is little hope for qualitative results without indexability.

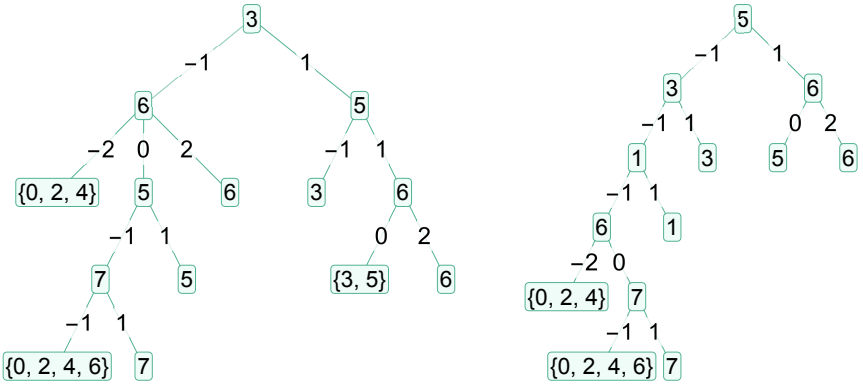

Figure 4: The optimal strategy when $\psi(1) = -1$ (left), and when $\psi(1)$ is unknown (right). Nodes are the optimal points to explore, and the edges' labels from each node to down are the corresponding realization of that point's value.

3.4 Example 4

Consider now the choice betwenn points in $\{0, \ldots, 9\}$ and the outside utility ρ when $\psi(0) = \psi(2) = 3$, $\psi(9) = 6$, and $U(y) = -y^2$. This example shows that the instability of the ranking between different intervals is not due to risk neutrality.

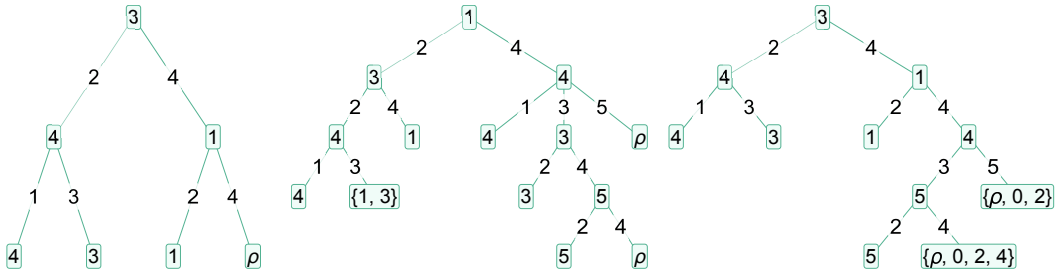

Figure 5: The optimal strategy for $\rho = -5, -7$ and -9, respectively. Nodes are the optimal points to explore, and the edges' labels from each node to down are the corresponding realization of that point's value.

4 R&D Before Market Entry

We now specialize the general framework to the R&D problem of a startup before market entry. The search domain is $\mathcal{X} = \mathbb{R}_+$, and each location $x \in \mathbb{R}_+$ represents a potential product or technology. Its flow profitability is

$$\Psi_x = \bar{y}_0 + \mu x + \sigma W_x, \qquad \sigma > 0,$$

where $W = \{W_x : x \geq 0\}$ is a standard Brownian motion. The startup knows the parameters (μ, σ) and the initial profitability $\bar{y}_0$, but not the realized profitability landscape $\psi(x) = \Psi_x(\omega)$. Developing a product reveals its profitability permanently.

We retain the informational state S_{t-1}, the set of developed products X_{t-1}, and the action set $\mathcal{A}(X_{t-1})$ defined in the general model. The three action types now have the following economic interpretations. Selecting $x \in X_{t-1}$ means putting a previously developed product on the market. Selecting an interval $(x_L, x_R) \in \mathcal{I}_{t-1}$ means funding an R&D project that combines the technologies developed at x_L and x_R. Selecting (x^r_{t-1}, ∞) means funding research beyond the most innovative technology developed so far.

This distinction separates two forms of R&D. Frontier research moves into a new technological direction and remains exposed to the drift μ of the profitability landscape. Recombination instead takes place between two realized technological anchors. Conditional on their observed profitabilities, the unresolved landscape between them is a Brownian bridge: its distribution depends on the endpoint profitabilities and their technological distance, but not on the drift accumulated before those endpoints were observed.

Every R&D experiment costs $c > 0$. The startup cannot develop a prototype and market it in the same period. In the notation of the general model, the application therefore imposes $u(y, \chi) = \chi y$. The period payoff is

$$r_t = \chi_t y_t - (1 - \chi_t)c, \qquad \chi_t = \mathbf{1}\{a_t \in X_{t-1}\},$$

and the startup solves

$$V(S_{-1}) = \sup_{\Sigma} \mathbb{E}^{\Sigma} \left[\sum_{t=0}^{\infty} \delta^t \left(\chi_t y_t - (1 - \chi_t)c \right) \middle| S_{-1} \right].$$

Here and below, expectations incorporate both the Brownian prior and the fixed gap and frontier protocols specified in the next subsection.

Market-entry interpretation. For an informational state S, define the profitability of the best developed product by

$$m(S) \equiv \max\{y : (x, y) \in S\}.$$

A known product with profitability y is represented by the safe arm $\mathsf{S}(y)$, whose normalized Gittins index is $I\big(\mathsf{S}(y)\big) = y$. Operating this arm yields y, reveals no information, and reproduces the same arm. Moreover, all unselected opportunities remain unchanged. Hence, if a known product is optimal at some history, it remains optimal after it is selected. There is therefore an optimal policy under which the first selection of a known product is absorbing: the startup puts that product on the market and operates it forever.

The repeated-action formulation is consequently equivalent to an optimal market-entry problem. If τ denotes the first period in which a known product is selected and

$$\alpha = (a_t)_{t<\tau}, \qquad a_t \in \mathcal{I}_{t-1} \cup \{(x^r_{t-1}, \infty)\},$$

denotes the pre-entry R&D policy, then

$$V(S_{-1}) = \sup_{\tau,\alpha} \mathbb{E}^{\alpha} \left[-c \sum_{t=0}^{\tau-1} \delta^t + \delta^{\tau} \frac{m(S_{\tau-1})}{1-\delta} \middle| S_{-1} \right].$$

The terminal term is understood to be zero on the event $\{\tau = \infty\}$. Thus, selecting a safe arm in the flow formulation is equivalent to stopping R&D and launching the best product developed by that date.

4.1 Research Protocols

The economic force behind the granularity constraint in this application is imperfect control. A startup can decide which broad research direction to pursue, but it cannot perfectly choose the exact prototype that will emerge from the R&D process. For example, management may choose to combine two existing technologies, but the precise technological design generated by the research team depends on scientific and engineering frictions. Similarly, the startup may choose to push the technological frontier in an attempt to develop a radical innovation. However, how much innovative the new technology will be depends on the unpredictable nature of basic research and the fundamental limits of the underlying science.

The general model allows the sampling kernel of an exploratory opportunity to depend on its local state. In the R&D application, we impose a more specific structure.

Assumption 5. *There exists a probability measure F on $[0, 1]$ satisfying $F((0, 1)) = 1$ and*

$$U \overset{d}{=} 1 - U \qquad \text{for } U \sim F$$

such that if the startup selects the interval (x_L, x_R), with $\ell = x_R - x_L$, then the new prototype is developed at

$$x_t = x_L + U_t \ell, \qquad U_t \sim F.$$

Note that assumption 5 implies that the distribution determining the location sampled inside an interval scales linearly with length, and is independent of the profitability of the technologies at

the boundaries of the interval.[15] Within the class of protocols that are not responsive to boundary profitability, the symmetry condition is the natural benchmark. A fixed asymmetric distribution would mechanically privilege one side of a bounded interval over the other, even though the labels "left" and "right" have no intrinsic economic content. Symmetry simply rules out this arbitrary directional bias.

Economically, this captures the idea that skewing the distribution of a single prototype toward a desired endpoint is costly. A research team operating over a short time interval may have limited control over the precise design that emerges from a fixed experimental routine. To move systematically toward a more profitable technology, the firm must instead iterate: develop a prototype, observe its payoff, update the set of available gaps, and decide whether to continue in the newly created region. This form of steering is costly because each iteration incurs the R&D cost c and takes time, so it discounts future market payoffs. The fixed protocol F therefore represents short-run imperfect control, while the index policy captures the firm's long-run ability to direct research by repeatedly reallocating effort toward the most valuable descendants.

Assumption 6. *There exists a probability measure G on $[0,\infty)$ satisfying*

$$G((0,\infty)) = 1 \qquad \textit{and} \qquad \int_{(0,\infty)} \Delta \, G(d\Delta) < \infty$$

such that if the startup expands the frontier from location x^r_{t-1}, then

$$x_t = x^r_{t-1} + \Delta_t, \qquad \Delta_t \sim G.$$

Moreover, arbitrarily small and arbitrarily large frontier steps occur with positive probability:

$$G((0,\varepsilon)) > 0 \quad \textit{for every } \varepsilon > 0, \qquad \textit{and} \qquad G((M,\infty)) > 0 \quad \textit{for every } M > 0.$$

The distribution G is the frontier analogue of the fixed protocol F used for bounded gaps. It captures the idea that, when the startup funds basic research, it can decide to push the frontier but cannot perfectly choose the magnitude of the technological jump. Under these research protocols, the finite first moment of G yields a bound on expected absolute profitability that grows at most linearly in the number of subsequent periods.. Appendix B.1 establishes this bound. It follows that both requirements of Assumption 4 are satisfied by Remark 1. The support conditions allow both incremental and radical discoveries to occur with positive probability.

4.2 Initialization

Before the first R&D experiment, the startup has developed only the product at the origin. The initial portfolio therefore consists of the safe arm $\mathsf{S}(\bar{y}_0)$ and the frontier arm $\mathsf{F}(\bar{y}_0)$. There are no bounded recombination opportunities at this stage.

To state the condition under which the startup begins by conducting R&D, consider the single-arm descendant process initiated by $\mathsf{F}(0)$. For an admissible policy π and stopping time τ, let

$$\chi^\pi_n \equiv \mathbf{1}\{z^\pi_n \in \mathcal{Z}_{\mathrm{s}}\}$$

[15] Assumption 5 should not be interpreted as saying that the firm ignores profitability information. Rather, it separates two margins of control. At the one-period implementation margin, the firm can choose the broad pair of technologies to combine, but it cannot finely tailor the exact prototype generated by the research process. Conditional on choosing the interval (x_L, x_R), the relative location U is drawn from a fixed protocol F. At the dynamic allocation margin, however, the firm remains fully responsive to payoff information. After each prototype is developed, the original interval is replaced by descendant intervals and a new safe product. The firm then compares the indices of these descendants and decides where to continue. Thus, profitability information tilts the realized R&D path through continuation choices rather than through the one-shot sampling kernel.

indicate that the descendant selected at local date n is a safe arm. When $\chi_n^\pi = 1$, write $z_n^\pi = \mathsf{S}(\widetilde{y}_n^\pi)$, where $\widetilde{y}_n^\pi$ is the safe payoff measured relative to the initial frontier payoff. The value of $\widetilde{y}_n^\pi$ is immaterial when $\chi_n^\pi = 0$.

Define the gross productivity of a research program initiated at the frontier by

$$\kappa^F(\mu, \sigma, F, G, \delta) \equiv \sup_{\substack{\pi \in \Pi(\mathsf{F}(0)) \\ \tau \in \mathcal{T}(\mathsf{F}(0), \pi)}} \frac{\mathbb{E}_0^\pi \left[\sum_{n=0}^{\tau-1} \delta^n \chi_n^\pi \widetilde{y}_n^\pi \right]}{\mathbb{E}_0^\pi \left[\sum_{n=0}^{\tau-1} \delta^n (1 - \chi_n^\pi) \right]}.$$

The numerator is the expected discounted incremental profitability obtained from safe products generated by the research program. The denominator is the expected discounted number of R&D activations within that program. Importantly, the program includes not only subsequent frontier experiments, but also any recombination projects generated by frontier discoveries. The quantity κ^F therefore depends on both research protocols F and G.

Define the initial-profitability cutoff

$$k^F(\mu, \sigma, F, G, c, \delta) \equiv \kappa^F(\mu, \sigma, F, G, \delta) - c.$$

Assumption 7. *The initial product satisfies*

$$\bar{y}_0 < k^F(\mu, \sigma, F, G, c, \delta).$$

Appendix B.2 shows that $0 < \kappa^F(\mu, \sigma, F, G, c, \delta) < \infty$ and that Assumption 7 is necessary and sufficient for frontier expansion to be uniquely optimal at date 0. At equality, frontier expansion and immediate market entry are both optimal; above the cutoff, immediate market entry is uniquely optimal. We impose the strict inequality because our interest is in the dynamics of the research process after it has begun.

The condition has a direct economic interpretation. Relative to commercializing the initial product, every period devoted to R&D entails the direct cost c and forgoes the current profit $\bar{y}_0$. Thus, $\bar{y}_0 + c$ is the effective cost of allocating one period to research. The quantity κ^F is the highest discounted incremental profitability that the frontier research technology can generate per discounted R&D period. Assumption 7 requires the productivity of the research program to exceed this effective cost:

$$\bar{y}_0 + c < \kappa^F(\mu, \sigma, F, G, \delta).$$

Because κ^F does not depend on the direct cost c, the initial profitability cutoff decreases one-for-one with that cost, so that:

$$\frac{\partial k^F}{\partial c} = -1.$$

A higher R&D cost therefore reduces by exactly the same amount the maximum profitability of the initial product for which experimentation is optimal.

4.3 R&D after a Frontier Discovery

Under assumption 7, it is optimal to start the R&D process and postpone market entry. The unique R&D opportunity initially available is frontier expansion. So, this section studies what follows a frontier discovery.

Consider an informational state $S = \{(x, \psi(x)) : x \in X\}$, and suppose that the startup selects the frontier opportunity (x^r, ∞). Let $y \equiv \psi(x^{\mathrm{r}})$ be the profitability of the product at the old frontier, and define

$$\rho \equiv \max_{a \in \mathcal{A}(X) \setminus \{(x^r, \infty)\}} I\big(z(a; S)\big)$$

as the highest index among all opportunities available before the experiment other than the selected frontier arm. Since the product at the old frontier remains available as a safe arm, we have $\rho \geq y$.

Frontier research produces a step $\Delta > 0$ and a new product with profitability

$$y' = y + \mu\Delta + \sigma\sqrt{\Delta}\,\varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, 1).$$

The experiment replaces the old frontier arm by

$$\big\{\mathsf{S}(y'),\, \mathsf{F}(y'),\, \mathsf{G}(\Delta, y, y')\big\}.$$

All other opportunities remain unchanged. For notational convenience, throughout the remainder of this section we write $I^{\mathrm{gap}}(\ell, y_L, y_R) \equiv I\big(\mathsf{G}(\ell, y_L, y_R)\big)$ for the index of a recombination project and $I^{\mathrm{front}}(y) \equiv I\big(\mathsf{F}(y)\big)$ for the index of frontier research. We suppress their dependence on the research protocols and the remaining model primitives.

By Theorem 1, the startup's next decision is therefore determined by the four scalar values ρ, y', $I^{\mathrm{front}}(y')$, and $I^{\mathrm{gap}}(\Delta, y, y')$. Thus, the selected frontier arm is summarized by its old profitability y, while the entire collection of other pre-existing opportunities enters the post-discovery decision only through its highest index ρ. Proposition 1 presents the dynamics that follows frontier expansion. A short economic interpretation of the proposition concludes the section.

Proposition 1 (R&D after a frontier discovery)**.** *Fix the old frontier profitability y, the highest pre-existing index ρ, and a realized frontier experiment (Δ, y').*

(i) ***Market entry after a valuable discovery.*** *There exists a finite threshold $\overline{y}^{\,F}(\Delta, y, \rho)$ such that the newly developed product is optimal if and only if*

$$y' \geq \overline{y}^{\,F}(\Delta, y, \rho).$$

If the inequality is strict, immediate market entry is uniquely optimal.

(ii) ***Return to pre-existing projects.*** *Suppose $\rho > -c$. There exists a finite threshold $\underline{y}^{\,F}(\Delta, y, \rho)$ such that an opportunity from the pre-existing portfolio is optimal if and only if*

$$y' \leq \underline{y}^{\,F}(\Delta, y, \rho).$$

If the inequality is strict, every opportunity generated by the new frontier branch has index strictly below ρ. Moreover,

$$\underline{y}^{\,F}(\Delta, y, \rho) \leq \overline{y}^{\,F}(\Delta, y, \rho).$$

(iii) ***Frontier continuation.*** *Suppose that frontier research was strictly optimal before the experiment, that is, $I^{\mathrm{front}}(y) > \rho$. Then there exist $\eta > 0$ and $\overline{\Delta} > 0$ such that the new frontier arm is uniquely optimal whenever*

$$0 < \Delta < \overline{\Delta} \qquad \text{and} \qquad |y' - y| < \eta.$$

Under Assumption 6, this region is reached with positive probability.

(iv) ***Technological-distance cutoff for recombination.*** *For every fixed (y, y', ρ), define*

$$\Delta^F(y, y', \rho) \equiv \inf\left\{\ell > 0 : I^{\text{gap}}(\ell, y, y') \geq \max\left\{\rho,\, y',\, I^{\text{front}}(y')\right\}\right\}.$$

Then

$$0 < \Delta^F(y, y', \rho) < \infty.$$

Furthermore,

$$\Delta > \Delta^F(y, y', \rho) \Longrightarrow \mathsf{G}(\Delta, y, y') \text{ is uniquely optimal.}$$

Part (i) is an opportunity-cost-of-delay result. A higher realization y' raises the value of the new product one-for-one. It also raises the values of the new frontier and the new recombination project, but those alternatives require at least one additional period of R&D. Their indices therefore rise by at most δ for each unit increase in y'. Once the new product is profitable enough to justify market entry, every still more profitable realization also induces entry.

Part (ii) describes the opposite tail. As y' becomes very low, the new product becomes unattractive, the new frontier index converges to $-c$, and the newly created gap is pulled down by its bad endpoint. Its index also converges to $-c$.[16] Hence, whenever the startup has a pre-existing opportunity with index strictly above $-c$, a sufficiently poor frontier discovery causes it to abandon the new branch and return to an older product or research project.

Between these tails, continued R&D within the new branch can be optimal. Part (iii) establishes that this region is nonempty whenever frontier research was strictly optimal before the experiment. An incremental discovery that leaves profitability close to its previous level preserves the value of continued frontier research, while creating only a short recombination project.

Part (iv) isolates the role of technological distance. Conditional on (y, y', ρ), the indices of the old portfolio, the new product, and the new frontier do not depend on Δ. Only the index of the newly created gap does. A vanishingly short gap does not justify paying the R&D cost and delaying market entry, which gives $\Delta^F(y, y', \rho) > 0$. As technological distance grows, Brownian-bridge dispersion and the associated option value grow without bound, which gives $\Delta^F(y, y', \rho) < \infty$.[17] Part (iv) also implies that whenever

$$I^{\text{front}}(y') > \max\{\rho, y'\},$$

the research direction followed after frontier exploration depends on the technological distance between the previous frontier and the technology just developed: incremental discoveries sustain frontier momentum, whereas sufficiently distant discoveries induce the startup to turn back and develop the technological region that the discovery has left unexplored.

[16] Even though the newly created gap is still anchored on the old frontier payoff y, the firm cannot costlessly choose a prototype arbitrarily close to that good endpoint. Under the fixed sampling protocol, every R&D attempt produces an interior point. When y' is very low, interior products are pulled down by the bad endpoint unless they happen to lie extremely close to the old technology. Reaching such points requires repeated experimentation, and each iteration costs c and is discounted. As y' goes to $-\infty$, the option value of this newly created gap therefore vanishes, and its index converges to the value of try once and retire, namely $-c$.

[17] This conditional comparison should not be read as assigning separate roles to profitability and technological distance. The thresholds in parts (i) and (ii) depend on Δ, because a larger step changes the value of the recombination opportunity generated by the discovery. Conversely, $\Delta^F(y, y', \rho)$ depends on y', because the realized profitability affects the new product, the new frontier, and the newly created gap. Post-discovery R&D is therefore determined jointly by the economic value and the technological location of the discovery. The proposition characterizes two conditional slices of this joint decision problem: variation in y' holding Δ fixed, and variation in Δ holding y' fixed.

4.4 R&D after a Recombination Experiment

We next characterize the evolution of R&D after the startup combines two previously developed technologies. Consider a realized informational state $S = \{(x, \psi(x)) : x \in X\}$, and suppose that the startup selects the gap $(x_L, x_R) \in \mathcal{I}(X)$. Write

$$\ell \equiv x_R - x_L, \qquad y_L \equiv \psi(x_L), \qquad y_R \equiv \psi(x_R).$$

Let

$$\rho \equiv \max_{a \in \mathcal{A}(X) \setminus \{(x_L, x_R)\}} I\big(z(a; S)\big)$$

be the highest index among all opportunities available before the experiment other than the selected gap.

The two endpoint products and the technological frontier remain available after the experiment. Hence $\rho \geq \max\{y_L, y_R\}$ and $\rho > -c$. Suppose that the recombination protocol draws the relative location $u \in (0, 1)$ and that the resulting prototype has profitability y_M. The root gap is replaced by

$$\{\mathsf{S}(y_M),\, \mathsf{G}(u\ell, y_L, y_M),\, \mathsf{G}((1-u)\ell, y_M, y_R)\}\,.$$

For a generic prototype profitability $p \in \mathbb{R}$, define the indices of the left and right descendant projects by

$$J_L(p) \equiv I^{\text{gap}}(u\ell, y_L, p) \qquad \text{and} \qquad J_R(p) \equiv I^{\text{gap}}((1-u)\ell, p, y_R).$$

At the realized state, the startup therefore compares ρ, y_M, $J_L(y_M)$, and $J_R(y_M)$.

Proposition 2 (R&D after a recombination experiment)**.** *Fix (u, y_L, y_R, ρ).*

(i) ***Profitability thresholds.*** *In addition, fix also ℓ. Then, there exist unique finite thresholds*

$$\underline{y}^{\,\text{rec}} \equiv \underline{y}^{\,\text{rec}}(\ell, u, y_L, y_R, \rho) \qquad \textit{and} \qquad \overline{y}^{\,\text{rec}} \equiv \overline{y}^{\,\text{rec}}(\ell, u, y_L, y_R, \rho)$$

satisfying

$$\underline{y}^{\,\text{rec}} \leq \overline{y}^{\,\text{rec}}, \quad \rho = \max\left\{\underline{y}^{\,\text{rec}}, J_L\big(\underline{y}^{\,\text{rec}}\big), J_R\big(\underline{y}^{\,\text{rec}}\big)\right\}, \quad \textit{and} \quad \overline{y}^{\,\text{rec}} = \max\left\{\rho, J_L\big(\overline{y}^{\,\text{rec}}\big), J_R\big(\overline{y}^{\,\text{rec}}\big)\right\}.$$

such that the post-experiment policy has the following form:

$$\begin{aligned} y_M < \underline{y}^{\,\text{rec}} \quad &\Longrightarrow \quad \textit{every newly generated opportunity has index below } \rho, \\ \underline{y}^{\,\text{rec}} < y_M < \overline{y}^{\,\text{rec}} \quad &\Longrightarrow \quad \max\{J_L(y_M), J_R(y_M)\} > \max\{\rho, y_M\}, \\ y_M > \overline{y}^{\,\text{rec}} \quad &\Longrightarrow \quad y_M > \max\{\rho, J_L(y_M), J_R(y_M)\}. \end{aligned}$$

Thus, a sufficiently poor prototype causes the startup to return to its pre-existing portfolio; an intermediate prototype induces further recombination in one of the two descendant regions; and a sufficiently profitable prototype induces market entry.

At $y_M = \underline{y}^{\,\text{rec}}$, an opportunity attaining ρ is weakly optimal. At $y_M = \overline{y}^{\,\text{rec}}$, commercialization is weakly optimal. The strict continuation region is nonempty if and only if $\underline{y}^{\,\text{rec}} < \overline{y}^{\,\text{rec}}$.

(ii) ***Technological-span cutoff.*** *In addition, fix also y_M and, for every hypothetical root length $\lambda > 0$, define*

$$K(\lambda) \equiv \max\left\{I^{\text{gap}}(u\lambda, y_L, y_M), I^{\text{gap}}((1-u)\lambda, y_M, y_R)\right\} \qquad \textit{and} \qquad M \equiv \max\{\rho, y_M\}.$$

Define

$$\ell^{\text{rec}}(u, y_M, y_L, y_R, \rho) \equiv \inf\{\lambda > 0 : K(\lambda) \geq M\}.$$

Then

$$0 < \ell^{\text{rec}}(u, y_M, y_L, y_R, \rho) < \infty.$$

Moreover,

$$\ell > \ell^{\text{rec}} \Longrightarrow \max\{J_L(y_M), J_R(y_M)\} > \max\{\rho, y_M\}.$$

Hence, holding fixed the prototype's relative location and profitability, a sufficiently wide original technological gap necessarily induces another round of recombination.

Proposition 2 shows that recombination need not be a one-shot experiment. For fixed project geometry and a fixed pre-existing portfolio, the realized profitability of the prototype orders the firm's next decision. A sufficiently poor prototype causes the firm to abandon the descendants of the selected gap and return to an older opportunity. An intermediate prototype makes one of the two descendant regions more valuable than both the outside portfolio and immediate market entry, inducing another round of recombination. A sufficiently profitable prototype instead ends the R&D process through commercialization.

The profitability thresholds depend on the length and endpoint profitabilities of the original gap and on the prototype's relative location. The realized payoff and the geometry of the project therefore remain intertwined. Part (ii) holds the prototype realization and its relative location fixed and isolates the amount of technological space left after the experiment. A narrow original gap may be largely exhausted by the first prototype, so that neither descendant justifies another costly R&D round. A sufficiently wide gap leaves enough unresolved technological variation for at least one descendant project to dominate both market entry and the pre-existing portfolio. Recombination can therefore initiate an endogenous phase of local refinement, but it need not do so after every experiment.

4.5 Empirical Interpretation and Predictions

The endogenous movement between frontier expansion and recombination provides a mechanism consistent with the phased search process documented by Randle and Pisano [2021]. They find that firms producing breakthrough inventions initially explore unfamiliar technological terrain and subsequently shift toward exploiting the knowledge accumulated during that exploration. In their terminology, exploitation means a more focused use of accumulated knowledge and therefore corresponds most closely to recombination in our model, rather than necessarily to market entry. The evidence establishes the phased pattern, but does not identify the threshold mechanism developed here. In the model, outward research creates technological anchors, and the regions between these anchors can later become more valuable than another frontier experiment.

A second empirical connection concerns the recombinant uncertainty documented by Fleming [2001]. Fleming studies inventions as combinations of technological components and shows that experimentation with unfamiliar components and new component combinations is associated with a distinctive risk–return pattern: such inventions are less useful on average, but their outcomes are also more dispersed, generating both more failures and a greater possibility of breakthroughs. In our model, technological distance and heterogeneity between the technologies being combined provide two natural counterparts to this notion of unfamiliarity.

The model generates the dispersion effect directly. Let

$$\bar{y} \equiv \frac{y_L + y_R}{2}, \qquad s \equiv y_R - y_L.$$

The profitability of the first prototype developed inside a gap can be written as

$$Y = (1-U)y_L + Uy_R + \sigma\sqrt{\ell U(1-U)}\,\varepsilon, \qquad \varepsilon \sim \mathcal{N}(0,1).$$

Under the symmetric protocol,

$$\mathbb{E}[Y \mid \ell, \bar{y}, s] = \bar{y},$$

whereas

$$\mathrm{Var}(Y \mid \ell, \bar{y}, s) = s^2\,\mathrm{Var}(U) + \sigma^2 \ell\,\mathbb{E}[U(1-U)].$$

Hence technological distance ℓ increases the dispersion of research outcomes. It raises both downside and upside risk, closely mirroring Fleming's recombinant-uncertainty mechanism.

There is, however, an important distinction. In the model, greater distance does not mechanically reduce the expected profitability of a given project: holding its endpoint profitabilities fixed,

$$\mathbb{E}[Y \mid \ell, \bar{y}, s] = \bar{y}$$

for every ℓ. The negative relationship between unfamiliarity and average usefulness documented by Fleming can instead arise endogenously through *project selection*.

To see this, fix an index $\rho > -c$ representing the value of the best alternative available to the firm, and fix an endpoint difference $s \geq 0$. Define the minimum average endpoint profitability that makes a recombination project competitive with this alternative:

$$\bar{y}^*(\ell, s; \rho) \equiv \inf\left\{\bar{y} \in \mathbb{R} : I^{\mathrm{gap}}\left(\ell, \bar{y} - \frac{s}{2}, \bar{y} + \frac{s}{2}\right) \geq \rho\right\}.$$

Because the gap index is weakly increasing in technological distance and strictly increasing in average endpoint profitability,[18]

$$\ell_2 \geq \ell_1 \quad \Longrightarrow \quad \bar{y}^*(\ell_2, s; \rho) \leq \bar{y}^*(\ell_1, s; \rho).$$

Thus, a technologically more distant project can be worth undertaking even when the average profitability of its endpoint technologies—and hence the expected profitability of its first prototype—is lower. The reason is that the additional dispersion created by distance carries option value: the firm can abandon an unfavorable realization but preserve and build on a favorable one.

The connection to Fleming is consequently twofold. First, the model directly reproduces the greater outcome dispersion associated with less familiar combinations. Second, it provides a selection mechanism through which the projects actually undertaken can display lower average usefulness as unfamiliarity rises, even though unfamiliarity does not lower the conditional mean of any fixed project.

The threshold results generate two sharper predictions. First, conditional on the old frontier payoff, the realized value of the new invention, and the firm's pre-existing portfolio, the probability of follow-on recombination should display a threshold-like increase in the technological distance of a frontier invention. Second, after a recombination experiment, conditional on the prototype's value, its relative location, and the parent technologies' values, a wider original technological span should be more likely to generate another round of recombination rather than commercialization or a return to an old project.

These predictions can be taken to patent or project-level data. Text-based similarity measures in the spirit of Kelly et al. [2021] can be used to measure a discovery's distance from the firm's existing knowledge and the proximity of subsequent inventions to the old and new technological anchors.

[18] Lemma 10 part (ii) and (i), respectively.

For public firms, the patent-value estimates of Kogan et al. [2017] provide one possible proxy for the economic value of discoveries. The model calls for within-firm tests that condition on the pre-discovery portfolio and distinguish same-firm follow-on research from immediate commercialization or project termination.

5 Career Mobility with Correlated Jobs

5.1 The Worker's Career-Search Problem

We close the paper with a simple application to career mobility. The purpose of the application is to isolate how correlation across jobs affects a worker's search policy. Whereas the R&D application uses a continuous Brownian landscape to study how discoveries redirect innovation, here we use a discrete occupational landscape: the occupational domain is $\mathcal{X} = \mathbb{N}_0$.[19] This allows us to fully solve for the worker's optimal policy and to compare that with the optimal behavior in an environment that preserves the marginal distribution of match quality at every occupation, but removes all correlation across occupations.

The worker searches across occupations that differ in their task and skill content. Because closely related jobs require similar human capital, a worker's match quality in one occupation informs her match quality in adjacent ones. We capture this informational spillover by modeling the payoff process as a symmetric random walk

$$\Psi_n = \sum_{j=1}^{n} \varepsilon_j,$$

where

$$\mathbb{P}(\varepsilon_j = 1) = \mathbb{P}(\varepsilon_j = -1) = \frac{1}{2}$$

and the increments $(\varepsilon_j)_{j\geq 1}$ are mutually independent. All occupational positions therefore have the same expected match quality. Positions differ in their marginal dispersion, while nearby positions are correlated because they share a common sequence of occupational increments.[20]

We retain the history h_{t-1}, the observed set X_{t-1}, and the informational state S_{t-1} defined in Section 2, relabeling the period-t location x_t as the occupational choice n_t. Thus,

$$S_{t-1} = \{(n, y_n) : n \in X_{t-1}\}, \qquad X_{-1} = \{0\}.$$

The occupation-indexed variable y_n is the observed match at position n; the worker's period-t payoff is y_{n_t}.

At the beginning of period t, the worker chooses one feasible occupation using only the information in h_{t-1}. If the occupation has not been tried previously, its match quality is revealed after the choice. The worker spends the current period in the selected occupation and receives its realized payoff. Hence, she cannot inspect a new job, observe a poor match, and return to a known job before the current-period payoff is received. Returning to a previously tried occupation reveals no new information. In the notation of the general model, the career application imposes

$$u(y, \chi) = y, \qquad c = 0.$$

Thus, the worker receives the match payoff of the selected occupation whether or not that payoff was known before the choice, and trying new jobs is not costly.

[19] Since locations are restricted to the natural numbers, we denote them by n rather than x throughout this section.

[20] More precisely, for every $n, k \in \mathbb{N}_0$, $\mathbb{E}[\Psi_n] = 0$, $\mathrm{Var}(\Psi_n) = n$, and $\mathrm{Cov}(\Psi_n, \Psi_k) = \min\{n, k\}$.

Career mobility is also local. Experience in a job makes nearby jobs accessible, but does not allow the worker to move immediately to arbitrarily distant parts of the occupational space. Jobs therefore have a dual role: they generate a current match payoff and act as stepping stones that expand the set of occupations the worker can reach in the future. For a finite set of previously tried occupations $X \subseteq \mathbb{N}_0$, the set of feasible occupations is

$$\mathcal{C}(X) \equiv \left\{ n \in \mathbb{N}_0 : \min_{j \in X} |n - j| \leq 2 \right\}.$$

An untried occupation becomes feasible only when it lies within two positions of a job the worker has already held. The worker may remain at or return to any previously tried occupation.

This constraint is a reduced-form stepping-stone friction. Experience in a job both reveals the worker's match at that position and provides access to nearby jobs. Initially,

$$\mathcal{C}(X_{-1}) = \{0, 1, 2\}.$$

If the worker moves directly from position 0 to position 2, then positions 3 and 4 become feasible in the following period, while position 1 remains available as an untried job behind the career frontier. This interaction between learning and access creates a dynamic career problem. A worker may move outward into jobs that use increasingly different skills, leaving some nearby occupations untried while the career frontier remains attractive. If match quality later deteriorates, those skipped jobs can become valuable lateral options because their expected quality is informed by the worker's earlier experience.

A pure career policy is a sequence

$$\alpha = (\alpha_t)_{t \geq 0}$$

of measurable decision rules satisfying

$$n_t = \alpha_t(h_{t-1}) \in \mathcal{C}(X_{t-1})$$

for every $t \geq 0$. Let $\mathfrak{A}(S)$ denote the set of admissible continuation policies from a realized informational state S. The worker's continuation value is

$$V^C(S) = \sup_{\alpha \in \mathfrak{A}(S)} \mathbb{E}^\alpha \left[\sum_{t=0}^{\infty} \delta^t y_{n_t} \mid S \right].$$

The objective is well defined. Starting from the initial state, the mobility constraint gives $n_t \leq 2(t+1)$, while the support of the random walk gives $|y_{n_t}| \leq n_t$. Consequently, under every admissible policy,

$$\mathbb{E}^\alpha \left[\sum_{t=0}^{\infty} \delta^t |y_{n_t}| \right] \leq 2 \sum_{t=0}^{\infty} \delta^t (t+1) = \frac{2}{(1-\delta)^2} < \infty.$$

The same argument, with the current frontier added to the bound, applies after every reachable finite history. Assumption 4 is therefore satisfied.

5.2 Endogenous Granularity and Local Career Indices

From now on, unless it creates ambiguity, we suppress the time index and the dependence on the informational state S. Before proceeding with the analysis, we establish the following notation.

Define the career frontier and its match quality, respectively, by

$$r \equiv \max X, \qquad y \equiv y_r.$$

Define the best match observed so far and the set of occupations attaining that payoff by

$$m \equiv \max_{n \in X} y_n, \qquad \mathcal{B} \equiv \{n \in X : y_n = m\}.$$

Finally, let the frontier drawdown be denoted by

$$d \equiv y - m \leq 0.$$

The set of untried one-position gaps behind the frontier is

$$\mathcal{H} \equiv \left\{ n \in \mathbb{N} : \begin{array}{l} n \notin X, \\ n-1 \in X, \\ n+1 \in X \end{array} \right\}.$$

Among these gaps, define

$$\mathcal{G} \equiv \{n \in \mathcal{H} : y_{n-1} = y_{n+1} = m\}, \qquad g \equiv |\mathcal{G}|.$$

A position in $\mathcal{G}$ is an untried occupation between two jobs that both attain the worker's best observed match. Its payoff is $m+1$ or $m-1$ with equal probability. Such a position is therefore a latent lateral option: success improves the worker's fallback, while failure leaves the best previously observed job available.

Endogenous Granularity Constraint The local mobility constraint permits two ways to expand the career frontier: both $r+1$ and $r+2$ are feasible and untried. For $k \in \{1, 2\}$, define the maximum value achievable by first expanding the frontier by k steps:

$$Q_k(S) \equiv \sup_{\substack{\alpha \in \mathfrak{A}(S) \\ n_0 = r+k}} \mathbb{E}^{\alpha} \left[\sum_{t=0}^{\infty} \delta^t y_{n_t} \mid S \right].$$

The two moves have the same expected current payoff. However, the two-position realization is a mean-preserving spread of the one-position realization. So, it has a larger option value. It also advances the feasible frontier farther and leaves the intermediate occupation available for future search. It is then easy to see that the worker weakly prefers to expand the frontier as much as possible. The following lemma formalizes this statement.

Lemma 1. *At every reachable informational state S,*

$$Q_2(S) \geq Q_1(S).$$

Consequently, there exists an optimal policy under which every frontier expansion moves exactly two positions to the right. We select the two-position move whenever the worker is indifferent.

The proof couples the one- and two-position experiments as a martingale and uses convexity of the continuation value in the newly observed frontier match. The worker can ignore the additional midpoint opportunity created by the two-position move, so the larger opportunity set cannot lower continuation value.

Lemma 1 endogenously fixes the frontier protocol. Starting from position zero, frontier expansion visits the even positions

$$0, 2, 4, \ldots.$$

Each frontier experiment leaves an interval of length two behind it, whose only interior position is its midpoint. Thus, on the path of the selected optimal policy, both local protocols are deterministic:

$$\text{frontier step} = 2, \qquad \text{gap displacement} = 1.$$

There is no remaining choice over where to search within a bounded gap or how far to move when the frontier is selected. The career problem therefore satisfies the granularity constraint of the general model.

Career Indices. For every reachable match value $v \in \mathbb{Z}$, introduce the shorthand

$$\mathsf{S}_v \equiv \mathsf{S}(v), \qquad \mathsf{G}_v \equiv \mathsf{G}(2, v, v), \qquad \mathsf{F}_v \equiv \mathsf{F}(v).$$

A safe arm S_v represents a previously identified job with match quality v. It yields current payoff v and reproduces itself:

$$R(\mathsf{S}_v) = v, \qquad \mathsf{S}_v \longrightarrow \{\mathsf{S}_v\}.$$

A gap arm G_v represents the unique untried midpoint between two occupations, two positions apart, both with match quality v. Its expected current reward is v, and activation reveals $v+1$ or $v-1$ with equal probability:

$$R(\mathsf{G}_v) = v, \qquad \mathsf{G}_v \longrightarrow \begin{cases} \{\mathsf{S}_{v+1}\}, & \text{with probability } 1/2, \\ \{\mathsf{S}_{v-1}\}, & \text{with probability } 1/2. \end{cases}$$

No descendant gap remains because each resulting subinterval has length one.

A frontier arm F_v represents a two-position expansion from a frontier whose current match is v. Its expected current reward is also v. The new frontier match equals $v+2$, v, or $v-2$ with probabilities $1/4$, $1/2$, and $1/4$, respectively. The corresponding descendant laws are

$$R(\mathsf{F}_v) = v, \qquad \mathsf{F}_v \longrightarrow \begin{cases} \{\mathsf{S}_{v+2}, \mathsf{S}_{v+1}, \mathsf{F}_{v+2}\}, & \text{with probability } 1/4, \\ \{\mathsf{S}_v, \mathsf{G}_v, \mathsf{F}_v\}, & \text{with probability } 1/2, \\ \{\mathsf{S}_{v-2}, \mathsf{S}_{v-1}, \mathsf{F}_{v-2}\}, & \text{with probability } 1/4. \end{cases}$$

In the first and third branches, the two random-walk increments are identified from their sum. The midpoint payoff is therefore known with certainty even though the worker has not previously held that job, and it is represented as a safe opportunity. In the middle branch, the two endpoints have the same payoff and the midpoint remains uncertain, producing the gap arm G_v. Opportunities that were available before the frontier experiment remain unchanged and are omitted from the display.

For the three career-arm indices, write

$$I^{\mathrm{S}}(v) \equiv I(\mathsf{S}_v), \qquad I^{\mathrm{G}}(v) \equiv I(\mathsf{G}_v), \qquad I^{\mathrm{F}}(v) \equiv I(\mathsf{F}_v).$$

The following lemma describes the index of each career opportunity.

Lemma 2. *There exist finite constants*

$$\iota_G = \iota_G(\delta) \qquad \textit{and} \qquad \iota_F = \iota_F(\delta)$$

such that, for every reachable $v \in \mathbb{Z}$,

$$I^{\mathrm{S}}(v) = v, \quad I^{\mathrm{G}}(v) = v + \iota_G, \quad \iota_G = \frac{\delta}{2-\delta}, \quad \textit{and} \quad I^{\mathrm{F}}(v) = v + \iota_F.$$

Moreover,

$$0 < \iota_G < 1 \qquad \textit{and} \qquad \iota_F > \iota_G.$$

The intuition behind the lemma is the following. Vertical translation increases the payoff from every activation, including the initial exploratory activation, by the same amount. Hence, each career index shifts one-for-one with v. This differs from the R&D application, where an exploratory activation yields the fixed payoff $-c$ and a payoff translation can affect rewards only through later safe-arm activations.

The gap constant is available in closed form because a gap creates only one safe descendant. Following the realization $v+1$, it is optimal in the single-gap problem to retain that safe arm; following $v-1$, it is optimal to retire the gap family. The frontier has a strictly larger index at the same current match: it offers a larger upper tail, creates additional frontier opportunities, and may generate a latent midpoint option.

The index formulas also identify which parts of the complete career history remain relevant.

Corollary 1. *At every state reachable under the policy selected in Lemma 1, the worker may restrict attention, without loss of optimality, to the multiset*

$$\left\{ \mathsf{S}_m, \mathsf{F}_y, \underbrace{\mathsf{G}_m, \ldots, \mathsf{G}_m}_{g \text{ copies}} \right\},$$

where the gap arms are omitted when $g = 0$. Consequently, the optimal action is determined by comparing

$$m, \qquad y + \iota_F, \qquad m + \iota_G \quad \text{when } g > 0.$$

To see the reduction, first note that every previously tried occupation with payoff below m is dominated by a safe occupation attaining m. Next consider an untried midpoint behind the frontier. If its endpoint payoffs differ by two, its payoff is known and lies below the larger endpoint. If its endpoints both equal some $v < m$, then

$$I^{\mathrm{G}}(v) = v + \iota_G < v + 1 \leq m,$$

where the strict inequality uses $\iota_G < 1$. The only uncertain midpoint that can outrank the best known occupation is therefore one whose two endpoints both attain the current maximum.

The number of such gaps affects the worker's continuation value, because failed lateral experiments can be followed by exploration of additional gaps. It does not affect the current action comparison since every available maximal gap has the same local index $m + \iota_G$.

Notice that at the initial state, $m = y = 0$, and $g = 0$. Since

$$I^{\mathrm{F}}(0) = \iota_F > 0 = I^{\mathrm{S}}(0),$$

frontier expansion is uniquely optimal, and the tie-breaking rule in Lemma 1 sends the worker to position 2.

The next subsection uses the three indices in Corollary 1 to characterize the correlated career policy as a threshold rule in the frontier drawdown d. We then compare this policy with an independent-marginal benchmark that preserves the distribution of Ψ_n at every position while removing the information that jobs provide about one another.

5.3 The Optimal Career Policy

Corollary 1 reduces the worker's decision to a comparison among three indices: the best known occupation has index m, the career frontier has index $y + \iota_F$, and, whenever $g > 0$, each maximal

gap has index $m + \iota_G$. Because all three indices shift one-for-one with match quality, their ranking depends on the frontier only through the drawdown d.

We adopt the following tie-breaking convention. An equality between the best safe occupation and the frontier is resolved in favor of settlement. An equality between a maximal gap and the frontier is resolved in favor of the lateral experiment. Define

$$\overline{d} \equiv \max\{d \in \mathbb{Z} : d + \iota_F \leq 0\} = \lfloor -\iota_F \rfloor$$

and

$$d^* \equiv \max\{d \in \mathbb{Z} : d + \iota_F \leq \iota_G\} = \lfloor \iota_G - \iota_F \rfloor .$$

The first threshold is the largest frontier drawdown at which the best known occupation weakly dominates further outward search. The second is the largest drawdown at which a maximal gap weakly dominates the frontier. The next proposition characterizes the optimal career policy.

Proposition 3 (Optimal career policy under correlation)**.** *An optimal occupation choice satisfies*

$$n^* \in \begin{cases} \mathcal{B}, & g = 0 \text{ and } d \leq \overline{d}, \\ \mathcal{G}, & g > 0 \text{ and } d \leq d^*, \\ \{r+2\}, & \text{otherwise.} \end{cases}$$

Thus, the worker settles at a previously observed best match when the frontier is sufficiently poor and no maximal gap remains; she makes a lateral move when the frontier has deteriorated sufficiently and a maximal gap is available; and she otherwise expands the career frontier by two positions.

The thresholds are finite and satisfy

$$\overline{d} < 0, \qquad d^* < 0, \qquad \overline{d} \leq d^* \leq \overline{d} + 1.$$

They depend only on the discount factor δ.

Proposition 3 shows how indexability reduces a potentially long career history to a simple comparison of local opportunities. The absolute level of the worker's best match does not affect the policy. What matters is how far the current frontier has fallen below that match and whether the worker has accumulated at least one informative lateral option.

The number of maximal gaps also does not affect the current action once there is at least one. Every maximal gap has the same local state and hence the same index. Additional gaps can raise the worker's continuation value because they provide further opportunities after a failed lateral experiment, but they do not change the comparison between the frontier and the next gap to be explored.

What follows a lateral experiment. Suppose that the worker selects a maximal gap. The untried midpoint yields $m + 1$ or $m - 1$ with equal probability.

If the lateral experiment succeeds and produces $m+1$, the new safe occupation has index $m+1$. Because $\iota_G < 1$, we have

$$m + 1 > m + \iota_G.$$

Moreover, the gap was selected only if

$$m + \iota_G \geq y + \iota_F.$$

It follows that

$$m + 1 > \max\{m + \iota_G,\, y + \iota_F\}.$$

The newly discovered occupation therefore strictly dominates the frontier, all remaining gaps, and every previously observed occupation. The worker settles there permanently.

If the lateral experiment fails and produces $m - 1$, the best observed match m and the frontier match y remain unchanged. The failed gap disappears, so the number of maximal gaps falls from g to $g - 1$. If another maximal gap remains, it has the same index $m + \iota_G$ as the gap just selected. Since the frontier drawdown is unchanged, the worker continues testing maximal gaps sequentially.

If all maximal gaps are exhausted after failures, the worker again compares the best safe occupation with the frontier. She settles if $d \leq \overline{d}$ and resumes frontier expansion if $d > \overline{d}$.

The optimal career path consequently moves among three regimes:

$$\text{frontier expansion} \quad \longrightarrow \quad \text{lateral experimentation} \quad \longrightarrow \quad \text{settlement or renewed expansion.}$$

When the frontier match remains close to the best match previously observed, the worker continues moving outward. When the frontier deteriorates and a maximal gap exists, she pauses outward search and tries a previously skipped occupation. A successful lateral move produces a better fallback and ends search. Failed lateral moves are followed by additional gaps, if available, and eventually by either settlement or a return to the frontier.

The lateral move is not a return to an old occupation. It is the first attempt at a job that was previously feasible but left untried while the outward path remained attractive. Its value comes from the information provided by the two successful occupations on its sides. The next subsection removes this informational link while preserving the marginal distribution of match quality at every occupational position.

5.4 Independent-Marginal Jobs

We now remove correlation across occupations while preserving the marginal distribution of match quality at every position. This benchmark changes one feature of the environment: observing the match at one occupation no longer provides information about any other occupation. The mobility constraint, timing, preferences, and discount factor remain unchanged.

For every position $n \geq 1$, define

$$\Psi_n^I \equiv \sum_{j=1}^{n} \varepsilon_{n,j},$$

where the random variables

$$\{\varepsilon_{n,j} : n \geq 1,\ 1 \leq j \leq n\}$$

are mutually independent and

$$\mathbb{P}(\varepsilon_{n,j} = 1) = \mathbb{P}(\varepsilon_{n,j} = -1) = \frac{1}{2}.$$

The variables $\{\Psi_n^I\}_{n\geq 1}$ are therefore independent across positions, while $\Psi_n^I \overset{d}{=} \Psi_n$ for every n. Thus, the correlated and independent environments assign exactly the same mean and variance to each individual occupation. They differ only in what an observed job reveals about other jobs.[21] In the correlated environment, the frontier match predicts outward occupations, and the values at

[21] In particular, $\mathbb{E}[Y_n^I] = 0$, $\mathrm{Var}(Y_n^I) = n$, and $\mathrm{Cov}(Y_n^I, Y_k^I) = 0$ for $n \neq k$.

the endpoints of a skipped interval predict its midpoint. Under independence, neither inference is available.

The worker faces the same feasible set $\mathcal{C}(X)$ and maximizes the same discounted match component $V^I(s)$. The linear growth argument used in the correlated environment applies unchanged.

Bold expansion under independence. Let r denote the career frontier. Both $r+1$ and $r+2$ are feasible. For $k \in \{1,2\}$, let $Q_k^I(S)$ denote the expected discounted value obtained by selecting $r+k$ in the current period and behaving optimally thereafter. Once again, the match distribution at $r+2$ is a mean-preserving spread of the distribution at $r+1$. Moreover, selecting $r+2$ creates a weakly larger future opportunity set: it advances the career frontier farther and leaves $r+1$ available as an untried occupation. The worker can ignore this additional opportunity, so it cannot reduce value. Thus, the following lemma trivially holds.

Lemma 3. *At every reachable state,*

$$Q_2^I(S) \geq Q_1^I(S).$$

Consequently, there exists an optimal policy that chooses $r+2$ whenever it expands the career frontier. We select the two-position move whenever the worker is indifferent.

Under the policy selected in Lemma 3, frontier positions are again even and every frontier experiment leaves one untried midpoint behind. Unlike in the correlated economy, however, the value of this midpoint is unrelated to the matches observed on its two sides.

Independent career arms. An untried occupation n behind the career frontier is represented by the one-shot arm H_n^I. Activating it yields the current payoff Ψ_n^I, and generates the safe arm $\mathsf{S}(\Psi_n^I)$:

$$R(\mathsf{H}_n^I) = 0, \qquad \mathsf{H}_n^I \longrightarrow \{\mathsf{S}(\Psi_n^I)\},$$

A career-frontier arm F_r^I represents a two-position expansion from absolute frontier position r. Activating it yields the current payoff Ψ_{r+2}^I, and generates a safe arm, an untried occupation, and a new frontier. Thus,

$$R(\mathsf{F}_r^I) = 0, \qquad \mathsf{F}_r^I \longrightarrow \left\{\mathsf{S}(\Psi_{r+2}^I),\, \mathsf{H}_{r+1}^I,\, \mathsf{F}_{r+2}^I\right\}.$$

The newly generated hole remains independent of both the observed frontier payoff and all other occupations.

Once the two-position protocol has been selected, these arms form an Open Bandit Process. The relevant local state of an independent frontier is its absolute position r, because the marginal distribution of the next draw depends on $r+2$. The current realized payoff at the old frontier is not part of this local state.

Define

$$\gamma_n \equiv I\big(\mathsf{H}_n^I\big) \qquad \text{and} \qquad b_r \equiv I\big(\mathsf{F}_r^I\big).$$

The following lemma characterizes these indices.

Lemma 4. *For every $n \geq 1$, γ_n is the unique positive solution to*

$$\delta \mathbb{E}\left[\left(\Psi_n^I - \gamma_n\right)^+\right] = (1-\delta)\gamma_n.$$

Moreover,

$$n \leq k \quad \Longrightarrow \quad \gamma_n \leq \gamma_k.$$

For every reachable frontier position r, the frontier index b_r is finite and satisfies

$$b_r \geq \gamma_{r+2} > 0.$$

The inequality $b_r \geq \gamma_{r+2}$ has a simple interpretation. A frontier experiment contains at least as much value as a one-shot draw from the same occupational distribution, because the worker may ignore all additional descendants. In addition, it creates access to still more distant occupations. We are now ready to provide the optimal career policy when match quality is independent across jobs.

Proposition 4 (Optimal career policy under independent marginals)**.** *Under the tie-breaking rule that resolves indifference in favor of settlement, an optimal policy is*

$$n^{I,*} \in \begin{cases} \mathcal{B}^I, & m \geq b_r, \\ \{r+2\}, & m < b_r. \end{cases}$$

In particular, the worker never selects an untried occupation behind the career frontier and never expands the frontier by one position. Once the worker selects an occupation attaining m, settlement is permanent.

The independent policy depends on the best match m and the absolute frontier position r. It does not depend on the realized match at the current frontier. That realization contains no information about the next untried occupation. The frontier position matters because it determines the marginal dispersion of the next draw and the distributions of the future occupations that the worker can access.

Starting from the origin, the worker therefore experiments successively at $2, 4, 6, \ldots$ until the best match observed along this path becomes high enough to dominate the frontier index. The worker then settles permanently at an occupation attaining that match. The independent career has only two regimes:

$$\text{outward experimentation} \quad \longrightarrow \quad \text{permanent settlement.}$$

The correlated and independent environments have the same marginal distribution of match quality at every occupational position. Nevertheless, their optimal career policies differ. Under correlation, whenever $g > 0$ and $d \leq d^*$, the worker selects an untried occupation behind the frontier. Under independent marginals, an untried occupation behind the frontier is never selected at any history.

Thus, the delayed lateral-experimentation regime in Proposition 3 is generated by the information embedded in neighboring occupations.

The comparison separates two forces. In both environments, the payoff distribution becomes more dispersed with occupational distance. The option to retain a good match consequently supports bold outward experimentation even under independence. Correlation adds a different force. Successful jobs on both sides of a skipped occupation raise the value of trying that occupation later. A poor frontier realization can then make the latent lateral option more attractive than another outward move.

This distinction yields a testable prediction. Following adverse news in an outward career move, transitions should be especially likely toward an *untried* occupation that uses skills related to successful jobs held earlier in the worker's career. The prediction is not merely that workers return to familiar jobs or move to nearby jobs. The value of the lateral move should depend on the

quality of the worker's experiences on both sides of the new occupation. This mechanism provides one interpretation of evidence that adverse career shocks induce retraining in fields related to prior experience and that early occupational assignments have persistent effects on employment in the same or related occupations [Humlum et al., 2025, Bruhn et al., 2025]. These findings need not be generated exclusively by correlated learning; the model gives one mechanism and a conditional prediction.

Appendix A Indexability

Notation for random objects. The main text is written in terms of realized histories and states. In the appendices, whenever the induced stochastic process must be made explicit, A_t, $\hat{x}_t$, Y_t, and R_t denote the random action, sampled location, observed payoff, and per-period payoff, respectively. The random observed set and informational state are denoted by $\widehat{X}_t$ and $\widehat{S}_t$, while X_t and S_t denote their realizations, as in the main text. A fixed local arm state is denoted by z, whereas Z_n^π denotes the random arm selected at local date n under a descendant policy π.

The initial history, the policy, the payoff process, and the auxiliary randomizations used by the sampling protocols jointly induce a probability law over these objects.

Extending an Open Bandit result. We first record an extension of the bounded-reward index theorem of Wu and Zhou [2013]. The extension will be used to prove indexability of the problem introduced in section 2. Consider a discrete-time Open Bandit Process with arm-state space $\mathcal{Z}$, reward function R, and descendant kernel P: activating an arm z yields $R(z)$, replaces it by a finite random multiset with law $P(\cdot \mid z)$, and leaves all other arms unchanged. For a nonempty finite portfolio m, let $\Pi(m)$ denote the set of admissible history-dependent policies, and let Z_n^π denote the arm selected at date n under $\pi \in \Pi(m)$. We impose the usual measurability conditions, with $\mathcal{Z}$ standard Borel. We are interested in policies attaining

$$W(m_0) := \sup_{\pi \in \Pi(m_0)} \mathbb{E}_{m_0}^\pi \left[\sum_{t=0}^{\infty} \delta^t R(Z_t^\pi) \right]$$

for an initial nonempty finite portfolio m_0.

Lemma 5. *Consider the Open Bandit Process described above. Suppose that, for every nonempty finite portfolio m,*

$$C(m) := \sup_{\pi \in \Pi(m)} \mathbb{E}_m^\pi \left[\sum_{t=0}^{\infty} \delta^t |R(Z_t^\pi)| \right] < \infty, \tag{1}$$

$$\eta_T(m) := \sup_{\pi \in \Pi(m)} \mathbb{E}_m^\pi \left[\sum_{t=T}^{\infty} \delta^t |R(Z_t^\pi)| \right] \longrightarrow 0 \quad \text{as } T \to \infty. \tag{2}$$

Then the Gittins index $I(z)$ is finite and measurable for every $z \in \mathcal{Z}$. Moreover, every admissible policy satisfying

$$Z_t^\pi \in \arg\max_{z \in m_t^\pi} I(z)$$

at every date t is optimal.

Proof. The index denominator lies in $[1, (1-\delta)^{-1}]$, so (1) implies $|I(z)| \le C(\{z\}) < \infty$.

Consider approximations of the problem where reward functions are bounded: for $K \in \mathbb{N}$, define

$$R_K(z) := \begin{cases} -K, & R(z) < -K, \\ R(z), & -K \le R(z) < K, \\ K, & K \le R(z), \end{cases}$$

and

$$d_K(z) := |R(z) - R_K(z)|.$$

Let W_K and I_K be the corresponding value and index. The Open Bandit problems with reward functions $R_K(z)$ are indexable by Wu and Zhou [2013, Theorem 3.1].

We first establish

$$e_K(m) := \sup_{\pi \in \Pi(m)} \mathbb{E}_m^\pi \left[\sum_{t=0}^{\infty} \delta^t d_K(Z_t^\pi) \right] \longrightarrow 0 \quad \text{as } K \to \infty. \tag{3}$$

Define the finite horizon versions of C and e_K as

$$C_T(m) := \sup_{\pi \in \Pi(m)} \mathbb{E}_m^\pi \left[\sum_{t=0}^{T-1} \delta^t |R(Z_t^\pi)| \right],$$
$$e_{T,K}(m) := \sup_{\pi \in \Pi(m)} \mathbb{E}_m^\pi \left[\sum_{t=0}^{T-1} \delta^t d_K(Z_t^\pi) \right].$$

Both are zero at $T = 0$. For $z \in m$, write

$$m'(z) := (m \ominus \{z\}) \uplus D(z), \qquad D(z) \sim P(\cdot \mid z).$$

Backward induction gives

$$C_{T+1}(m) = \max_{z \in m} \big\{ |R(z)| + \delta \mathbb{E}[C_T(m'(z))] \big\},$$
$$e_{T+1,K}(m) = \max_{z \in m} \big\{ d_K(z) + \delta \mathbb{E}[e_{T,K}(m'(z))] \big\},$$

where maximizers exist because m is finite. In particular, $d_K(z) \le |R(z)|$ implies

$$0 \le e_{T,K} \le C_T \le C.$$

We claim by induction on T that, for every finite portfolio m,

$$e_{T,K}(m) \longrightarrow 0 \qquad \text{as } K \to \infty.$$

The claim is immediate for $T = 0$. Suppose it holds at horizon T, and fix a finite portfolio m and an arm $z \in m$. Conditional on activating z, the next portfolio $m'(z)$ is random. By the induction hypothesis,

$$e_{T,K}(m'(z)) \longrightarrow 0 \qquad \text{almost surely.}$$

Moreover,

$$0 \le e_{T,K}(m'(z)) \le C_T(m'(z)).$$

Notice that

$$|R(z)| + \delta \mathbb{E}[C_T(m'(z))] \le C_{T+1}(m) \le C(m) < \infty,$$

and hence

$$\mathbb{E}[C_T(m'(z))] < \infty.$$

Dominated convergence therefore yields

$$\mathbb{E}[e_{T,K}(m'(z))] \longrightarrow 0 \qquad \text{as } K \to \infty.$$

Since also $d_K(z) \to 0$, for every fixed $z \in m$,

$$d_K(z) + \delta \mathbb{E}[e_{T,K}(m'(z))] \longrightarrow 0.$$

Finally, m is finite, so taking the maximum over $z \in m$ preserves the convergence:

$$e_{T+1,K}(m) = \max_{z \in m} \left\{ d_K(z) + \delta \mathbb{E}[e_{T,K}(m'(z))] \right\} \longrightarrow 0.$$

This completes the induction.

As $d_K \leq |R|$,

$$e_K(m) \leq e_{T,K}(m) + \eta_T(m).$$

First let $K \to \infty$ and then $T \to \infty$ to obtain (3).

Let $Q(m, z)$ be the optimal value subject to activating z first, and define $Q_K(m, z)$ analogously. Observe that

$$\begin{aligned} |W_K(m) - W(m)| &\leq e_K(m), \\ |Q_K(m, z) - Q(m, z)| &\leq e_K(m), \\ |I_K(z) - I(z)| &\leq e_K(\{z\}). \end{aligned} \tag{4}$$

The last inequality again uses the lower bound of the index denominator. W_k and I_k are Borel measurable by standard arguments. The inequalities in (4) imply that W and I are pointwise limits of W_K and I_K, respectively. Thus, W and I are measurable.

Let us now show that, starting from any finite portfolio, it is optimal to choose an arm with maximal index. Fix m and $z \in \arg\max_{w \in m} I(w)$. Define

$$\Delta_K := \max_{w \in m} I_K(w) - I_K(z), \qquad a_K := \frac{\Delta_K + K^{-1}}{1 - \delta}.$$

By (4), $I_K(w) \to I(w)$ for every $w \in m$. Since m is finite,

$$\max_{w \in m} |I_K(w) - I(w)| \longrightarrow 0.$$

Observe that

$$|\max_{w \in m} I_K(w) - \max_{w \in m} I(w)| \leq \max_{w \in m} |I_K(w) - I(w)|.$$

Hence

$$\max_{w \in m} I_K(w) \longrightarrow \max_{w \in m} I(w) = I(z).$$

Together with $I_K(z) \to I(z)$, this implies $\Delta_K \to 0$ and $a_K \to 0$.

In the bounded problem, add a bonus a_K to the reward of the fixed initial arm z the first time it is activated. This is another bounded-reward open bandit, and the bonus is paid at most once. Every admissible policy and stopping time in the maximization problem associated with the index of z has its numerator increased by a_K and its denominator bounded above by $(1 - \delta)^{-1}$. Consequently, the index satisfies

$$I_K^+(z) \geq I_K(z) + (1 - \delta) a_K = \max_{w \in m} I_K(w) + K^{-1}.$$

The indices of all other initial arms are unchanged. Thus, z is the unique maximal-index initial arm, so the index theorem in Wu and Zhou [2013] guarantees that it is an optimal first action. Writing W_K^+ for the perturbed value, we obtain

$$W_K(m) \leq W_K^+(m) = Q_K(m, z) + a_K.$$

It follows that

$$0 \leq W(m) - Q(m, z) \leq 2e_K(m) + a_K \longrightarrow 0,$$

where the last inequality follows from adding and subtracting W_K and Q_k, the above inequality, and the inequalities in (4). Hence every original maximal-index arm is an optimal first action.

We first record the one-step decomposition of the original value. For every finite portfolio m and $z \in m$,

$$Q(m, z) = R(z) + \delta \mathbb{E}[W(m'(z))].$$

Indeed, the corresponding identity holds in every bounded problem, and the convergence established in as $K \to \infty$, together with $|W_K(m'(z))| \leq C(m'(z))$ and $\mathbb{E}[C(m'(z))] < \infty$, allows us to pass to the limit by dominated convergence. Hence, by the previous step, every maximal-index arm $z \in m$ satisfies

$$W(m) = R(z) + \delta \mathbb{E}[W(m'(z))].$$

Let π^I be any admissible maximal-index policy and let m_n denote its portfolio immediately before date n. Iterating the preceding equality gives, for every T,

$$W(m_0) = \mathbb{E}^{\pi^I}\left[\sum_{n=0}^{T-1} \delta^n R(Z_n^{\pi^I}) + \delta^T W(m_T)\right].$$

Moreover,

$$\delta^T \mathbb{E}^{\pi^I}[|W(m_T)|] \leq \eta_T(m_0) \longrightarrow 0.$$

Letting $T \to \infty$ and using discounted absolute integrability yields

$$W(m_0) = \mathbb{E}^{\pi^I}\left[\sum_{n=0}^{\infty} \delta^n R(Z_n^{\pi^I})\right].$$

Thus π^I is optimal. □

Corollary 2. *Under the hypotheses of Lemma 5, fix a relevant arm state z. Let $\pi^I \in \Pi(z)$ select an arm of maximal index in its single-arm descendant process, with $\mathcal{M}_{-1}^z = \{z\}$, and define*

$$\tau_z := \inf\left\{n \geq 1 : \max_{w \in \mathcal{M}_{n-1}^z} I(w) < I(z)\right\}, \qquad \inf \varnothing := \infty.$$

Then (π^I, τ_z) attains $I(z)$, that is,

$$\mathbb{E}_z^{\pi^I}\left[\sum_{n=0}^{\tau_z - 1} \delta^n \big(R(Z_n^{\pi^I}) - I(z)\big)\right] = 0. \tag{5}$$

Proof. Fix $q \in \mathbb{R}$ and add to the original problem an independent arm z_q which gives reward q and reproduces itself. Its index is q, and the indices of the other arms are unchanged.

We first verify that the augmented problem satisfies the lemma's hypotheses. Activations of z_q from any policy reveal no information. Total absolute rewards are therefore bounded by

$$\widetilde{C}(m \uplus \{z_q\}) \leq C(m) + \frac{|q|}{1 - \delta}.$$

Observe that $\eta_T(m)$ is the largest expected discounted absolute payoff that can be generated from date T onward by a policy that starts at date 0 from portfolio m. In the augmented problem, however, the original family need not have been activated T times by calendar date T, because some periods may have been spent on z_q. We therefore need to control separately the cases in which

the original family has advanced many local periods and in which it has instead been delayed for many calendar periods.

Let $L = \lfloor T/2 \rfloor$. If the original family has been activated at least L times, its remaining contribution is at most $\eta_L(m)$. If it has been activated less than L times, then at least $T-L$ calendar periods have been spent delaying those activations, which contributes an additional discount factor of at most δ^{T-L}. Moreover, the expected discounted absolute reward generated from time T is bounded above by $C(m)$. Consequently,

$$\widetilde{\eta}_T(m \uplus \{z_q\}) \leq \delta^{T-L} C(m) + \eta_L(m) + \frac{|q|\delta^T}{1-\delta} \longrightarrow 0.$$

It follows that Lemma 5 applies to the augmented problem.

Set $q = I(z)$ and start from $\{z, z_q\}$. Operating z_q forever is a maximal-index policy, with optimal value $q/(1-\delta)$. Another maximal-index policy activates z first, follows π^I until τ_z, and then operates z_q forever. Hence

$$\frac{q}{1-\delta} = \mathbb{E}_z^{\pi^I}\left[\sum_{n=0}^{\tau_z - 1} \delta^n R(Z_n^{\pi^I}) + \delta^{\tau_z} \frac{q}{1-\delta}\right].$$

Rearranging and using $1 - \delta^{\tau_z} = (1-\delta)\sum_{n=0}^{\tau_z - 1} \delta^n$ gives (5). Since $\tau_z \geq 1$, the index denominator is strictly positive, and the same equality establishes attainment of the normalized index. □

Proof of Theorem 1. We show that the resulting process is an Open Bandit Process satisfying the hypotheses of Lemma 5.

It is sufficient to establish four properties: the local arm-state space is a standard Borel space and each global portfolio contains finitely many arms; each local arm state has a measurable expected reward and a measurable law over finitely many descendants; the reward and descendant law of an activated arm depend only on its local state, while idle arms remain unchanged; and the discounting condition of Wu and Zhou [2013] is satisfied.

Step 1: measurable arm space and finite portfolios.

The local arm-state space $\mathcal{Z}$ is defined in Section 2. Since

$$\mathcal{L}_{\mathbb{R}_+} = (0, \infty), \qquad \mathcal{L}_{\mathbb{N}_0} = \{2, 3, \ldots\},$$

the feasible-length set $\mathcal{L}_{\mathcal{X}}$ is a Borel subset of $\mathbb{R}$. Hence

$$\mathcal{Z}_{\mathrm{s}} \simeq \mathbb{R}, \qquad \mathcal{Z}_{\mathrm{g}} \simeq \mathcal{L}_{\mathcal{X}} \times \mathbb{R}^2, \qquad \mathcal{Z}_{\mathrm{f}} \simeq \mathbb{R}$$

are standard Borel spaces. Their finite disjoint union

$$\mathcal{Z} = \mathcal{Z}_{\mathrm{s}} \sqcup \mathcal{Z}_{\mathrm{g}} \sqcup \mathcal{Z}_{\mathrm{f}}$$

is therefore also a standard Borel space.

For every finite informational state S, the portfolio of arms $\mathbf{M}(S)$ is a multiset. Moreover, $\mathbf{M}(S_{-1}) = \{\mathsf{S}(\bar{y}_0), \mathsf{F}(\bar{y}_0)\}$. A safe activation generates one descendant, and a gap or frontier activation generates at most three. Starting from a finite portfolio, the portfolio therefore remains finite after every finite number of activations. We denote $\mathsf{M}_f(\mathcal{Z})$ the measurable space of finite multisets on $\mathcal{Z}$.

Step 2: expected rewards and descendant kernels.

We first describe the local transition laws. For $d > 0$ and $(a, b) \in \mathbb{R}^2$, define the multiset

$$\mathbf{g}(d, a, b) \equiv \begin{cases} \{\mathsf{G}(d, a, b)\}, & d \in \mathcal{L}_{\mathcal{X}}, \\ \varnothing, & d \notin \mathcal{L}_{\mathcal{X}}, \end{cases}$$

where $\varnothing$ denotes the empty multiset. This convention omits subintervals that contain no feasible unobserved location. It matters only when $\mathcal{X} = \mathbb{N}_0$, where a subinterval of length one is empty.

For a safe arm,

$$R\big(\mathsf{S}(y)\big) = u(y, 1),$$

and its descendant law is degenerate:

$$P\big(\cdot \;\big|\; \mathsf{S}(y)\big) = \delta_{\{\mathsf{S}(y)\}},$$

where the right-hand side is the Dirac measure concentrated on the one-element multiset $\{\mathsf{S}(y)\}$.

Now fix a gap arm $z = \mathsf{G}(\ell, y_L, y_R)$. Because $\mathcal{X}$ and $\mathbb{R}$ are standard Borel spaces, regular conditional distributions admit measurable versions. Fix a Borel kernel

$$Q^{\mathrm{g}}\big(d\Delta, dv \;\big|\; z\big)$$

giving the joint conditional distribution of the relative displacement $\Delta \in (0, \ell) \cap \mathcal{X}$ and the payoff v observed at the sampled location. Assumptions 1, 2, and 3 imply that this kernel depends on the selected physical gap only through $z = \mathsf{G}(\ell, y_L, y_R)$.

A realization (Δ, v) generates the descendant multiset

$$\mathbf{d}^{\mathrm{g}}(z; \Delta, v) \equiv \{\mathsf{S}(v)\} \uplus \mathbf{g}(\Delta, y_L, v) \uplus \mathbf{g}(\ell - \Delta, v, y_R).$$

The expected one-period reward is

$$R(z) = \int \big(u(v, 0) - c\big) Q^{\mathrm{g}}\big(d\Delta, dv \;\big|\; z\big),$$

and the descendant kernel is the pushforward of $Q^{\mathrm{g}}(\cdot \mid z)$ under $\mathbf{d}^{\mathrm{g}}$:

$$P(B \mid z) = \int \mathbf{1}\left\{\mathbf{d}^{\mathrm{g}}(z; \Delta, v) \in B\right\} Q^{\mathrm{g}}\big(d\Delta, dv \;\big|\; z\big)$$

for every Borel set $B \subseteq \mathsf{M}_f(\mathcal{Z})$.

Next fix a frontier arm $z = \mathsf{F}(y)$. Let

$$Q^{\mathrm{f}}\big(d\Delta, dv \;\big|\; z\big)$$

be a measurable version of the joint conditional distribution of the positive frontier step Δ and the payoff v observed at the new frontier.[22] Assumptions 1, 2, and 3 imply that this kernel depends only on the frontier state $\mathsf{F}(y)$.

A realization (Δ, v) generates

$$\mathbf{d}^{\mathrm{f}}(z; \Delta, v) \equiv \{\mathsf{S}(v), \mathsf{F}(v)\} \uplus \mathbf{g}(\Delta, y, v).$$

The expected one-period reward is

$$R(z) = \int \big(u(v, 0) - c\big) Q^{\mathrm{f}}\big(d\Delta, dv \;\big|\; z\big),$$

[22] As above, its existence is guaranteed by the fact that $\mathcal{X}$ and $\mathbb{R}$ are standard Borel spaces.

and

$$P(B \mid z) = \int \mathbf{1}\left\{\mathbf{d}^{\mathrm{f}}(z;\Delta,v) \in B\right\} Q^{\mathrm{f}}\big(d\Delta, dv \mid z\big).$$

Because u is Borel measurable, the local kernels are measurable in their conditioning arguments, and the descendant maps are Borel measurable, the function and the kernel

$$z \longmapsto R(z) \qquad z \longmapsto P(\,\cdot \mid z)$$

are measurable. Assumption 4 guarantees that $R(z)$ is finite at every reachable local arm state.

We next relate the realized period payoff to the expected reward function R. Consider the single-arm descendant process initiated by a reachable state z. Under a policy $\pi \in \Pi(z)$, let

$$\mathbb{G}^{z,\pi} = \{\mathcal{G}_n^{z,\pi}\}_{n\geq 0}$$

be its natural filtration, where $\mathcal{G}_n^{z,\pi}$ is the information available immediately before the nth activation. Let Z_n^π denote the selected arm and let r_n^π be the realized payoff from that activation.

The selected state Z_n^π is $\mathcal{G}_n^{z,\pi}$-measurable, while its local realization occurs only after selection. Therefore,

$$\mathbb{E}_z^\pi\left[r_n^\pi \mid \mathcal{G}_n^{z,\pi}\right] = R(Z_n^\pi).$$

For every $\tau \in \mathcal{T}(z,\pi)$, the event $\{n < \tau\}$ belongs to $\mathcal{G}_n^{z,\pi}$. Absolute integrability[23] and iterated expectations therefore give

$$\mathbb{E}_z^\pi\left[\sum_{n=0}^{\tau-1}\delta^n r_n^\pi\right] = \sum_{n=0}^{\infty}\delta^n\mathbb{E}_z^\pi\left[\mathbf{1}_{\{n<\tau\}}r_n^\pi\right] = \sum_{n=0}^{\infty}\delta^n\mathbb{E}_z^\pi\left[\mathbf{1}_{\{n<\tau\}}R(Z_n^\pi)\right] = \mathbb{E}_z^\pi\left[\sum_{n=0}^{\tau-1}\delta^n R(Z_n^\pi)\right].$$

The same argument applies to the global search problem. Thus, replacing the realized current payoff by its conditional expectation preserves the expected value of every admissible policy. The realized observation is not removed from the information structure: it remains encoded in the descendant arm states and can affect all subsequent choices.

Step 3: equivalence with an Open Bandit Process. For an action A_t selected from the portfolio induced by $\widehat{S}_{t-1}$, write $Z_t \equiv z(A_t;\widehat{S}_{t-1})$ for its random local arm state. Let $\mathbf{D}_t$ be the multiset of descendants generated by its local realization. The informational state update implies the pathwise portfolio update

$$\mathbf{M}(\widehat{S}_t) = \Big(\mathbf{M}(\widehat{S}_{t-1}) \ominus \{Z_t\}\Big) \uplus \mathbf{D}_t.$$

To verify its transition structure, notice that at every history, the bounded gaps have disjoint interiors, and the frontier tail is disjoint from all bounded gaps. Assumption 1 implies that, conditional on the observed boundary values, the payoffs associated with locations in these components are mutually independent. Assumption 2 implies that the conditional distribution inside a selected gap depends on its endpoint locations only through their distance, while the distribution beyond the frontier does not depend on its absolute location. Assumption 3 then implies that, conditional on Z_t, the distribution of $\mathbf{D}_t$ is $P(\,\cdot \mid Z_t)$ and the expected current reward is $R(Z_t)$, independently of the states of all other available arms. The randomizations used by the local protocols do not introduce dependence on the remaining portfolio. Every unselected physical opportunity is unchanged, and therefore every unselected arm remains frozen.

Consequently, the portfolio process

$$\{\mathbf{M}(\widehat{S}_t)\}_{t\geq -1}$$

[23]The finiteness of the integral is guaranteed by assumption 4.

has exactly the transition structure of an Open Bandit Process: one available arm is activated, it is removed and replaced by a finite random multiset of descendants whose law depends only on its local state, and all idle arms are unchanged.

Multiplicity preserves distinct physical opportunities that happen to share the same local state. Choosing either copy produces the same local reward and descendant law. Thus every admissible physical search policy induces an admissible arm-selection policy for the portfolio process, and an arm policy can be implemented in the search problem by selecting a physical opportunity corresponding to the chosen copy. The two formulations therefore generate the same attainable expected discounted payoffs.

Step 4: discounting, finiteness, and the index theorem.

Every activation lasts one period. In the notation of Wu and Zhou [2013], the duration associated with every arm type is therefore $V_{\mathrm{WZ}}(z) \equiv 1$. Their Assumption 2.1 is satisfied because

$$\sup_{z \in \mathcal{Z}} \mathbb{E}\left[\delta^{V_{\mathrm{WZ}}(z)}\right] = \delta < 1.$$

We next verify finiteness of the normalized index.[24] Fix a reachable local arm state z. There exists a reachable finite informational state S whose portfolio contains a copy of z. Define

$$C(S) \equiv \sup_{\Sigma} \mathbb{E}_S^{\Sigma,\boldsymbol{\sigma}}\left[\sum_{t=0}^{\infty} \delta^t |r_t|\right] < \infty$$

where finiteness follows from Assumption 4.

Fix $\pi \in \Pi(z)$, and $\tau \in \mathcal{T}(z,\pi)$. The descendant policy can be embedded in the global problem by selecting the copy of z and subsequently following only the arms generated from it until τ, while leaving all outside opportunities idle. Extend the policy arbitrarily after τ. By the equality of realized and expected reward representations established in Step 2,

$$\left|\mathbb{E}_z^{\pi}\left[\sum_{n=0}^{\tau-1} \delta^n R(Z_n^{\pi})\right]\right| = \left|\mathbb{E}_z^{\pi}\left[\sum_{n=0}^{\tau-1} \delta^n r_n^{\pi}\right]\right| \leq \mathbb{E}_z^{\pi}\left[\sum_{n=0}^{\tau-1} \delta^n |r_n^{\pi}|\right] \leq C(S).$$

Moreover, since τ is strictly positive,

$$1 \leq \mathbb{E}_z^{\pi}\left[\sum_{n=0}^{\tau-1} \delta^n\right] \leq \frac{1}{1-\delta}.$$

Every discounted-average payoff entering the definition of $I(z)$ is therefore finite and uniformly bounded in absolute value by $C(S)$. Hence $I(z) \in \mathbb{R}$ for every reachable local arm state z.

For any reachable finite state S and any nonempty subportfolio m of $M(S)$, a policy on m can be implemented in the search problem while leaving all outside opportunities idle. Conditional Jensen's inequality and the expected-reward representation in Step 2 give, for every $T \geq 0$,

$$\sup_{\pi \in \Pi(m)} \mathbb{E}_m^{\pi}\left[\sum_{t=T}^{\infty} \delta^t |R(Z_t^{\pi})|\right] \leq \sup_{\Sigma} \mathbb{E}_S^{\Sigma,\sigma}\left[\sum_{t=T}^{\infty} \delta^t |R_t|\right].$$

Assumption 4 therefore verifies both conditions (1) and (2) in Lemma 5 for the constructed portfolios and their single-arm descendant problems. Lemma 5 implies that every admissible policy selecting an available arm of maximal normalized index I is optimal.

□

[24] Finiteness is implicitly assumed in Wu and Zhou [2013], but it is not automatic in our environment.

A.1 Fair-Charge Representation

The single-arm descendant process, the admissible policy set $\Pi(z)$, the stopping-time set $\mathcal{T}(z,\pi)$, and the normalized Gittins index $I(z)$ are defined in Section 2. We now record an equivalent representation of the index that will be used in the proofs of the R&D results.

For a reachable arm state $z \in \mathcal{Z}$ and a candidate charge $q \in \mathbb{R}$, define the discounted excess value

$$\mathcal{V}(z,q) \equiv \sup_{\substack{\pi\in\Pi(z)\\ \tau\in\mathcal{T}(z,\pi)}} \mathbb{E}_z^{\pi}\left[\sum_{n=0}^{\tau-1}\delta^n\left(R(Z_n^{\pi})-q\right)\right].$$

Because every stopping time in $\mathcal{T}(z,\pi)$ is strictly positive, the initial arm is activated before the descendant process can be retired.

Definition 1. *A number $q^* \in \mathbb{R}$ is a* fair charge *for the arm state z if*

$$V(z,q^*) = 0.$$

Lemma 6. *Suppose that Assumption 4 holds, and fix a reachable local arm state $z \in \mathcal{Z}$. Then, for every $q \in \mathbb{R}$,*

$$\mathcal{V}(z,q)\begin{cases} > 0, & q < I(z),\\ = 0, & q = I(z),\\ < 0, & q > I(z).\end{cases}$$

Consequently, the unique fair charge for z is its normalized Gittins index: $q^* = I(z)$

Proof. Fix $\pi \in \Pi(z)$, $\tau \in \mathcal{T}(z,\pi)$, and define

$$N_{\pi,\tau} \equiv \mathbb{E}_z^{\pi}\left[\sum_{n=0}^{\tau-1}\delta^n R(Z_n^{\pi})\right], \qquad D_{\pi,\tau} \equiv \mathbb{E}_z^{\pi}\left[\sum_{n=0}^{\tau-1}\delta^n\right].$$

Since $\tau \geq 1$,

$$1 \leq D_{\pi,\tau} \leq \frac{1}{1-\delta}.$$

The excess value generated by this policy-stopping pair is

$$\mathbb{E}_z^{\pi}\left[\sum_{n=0}^{\tau-1}\delta^n\left(R(Z_n^{\pi})-q\right)\right] = N_{\pi,\tau} - qD_{\pi,\tau} = D_{\pi,\tau}\left(\frac{N_{\pi,\tau}}{D_{\pi,\tau}} - q\right). \tag{6}$$

Suppose first that $q > I(z)$. By the definition of the normalized index,

$$\frac{N_{\pi,\tau}}{D_{\pi,\tau}} \leq I(z)$$

for every admissible pair (π,τ). Hence

$$N_{\pi,\tau} - qD_{\pi,\tau} \leq D_{\pi,\tau}\big(I(z)-q\big) \leq I(z) - q < 0,$$

where the second inequality follows from $D_{\pi,\tau} \geq 1$ and $I(z) - q < 0$. Taking the supremum over (π,τ) gives

$$\mathcal{V}(z,q) \leq I(z) - q < 0.$$

Now suppose that $q < I(z)$. By the definition of the supremum, there exists an admissible pair (π, τ) such that

$$\frac{N_{\pi,\tau}}{D_{\pi,\tau}} > q.$$

Equation (6) then implies

$$N_{\pi,\tau} - qD_{\pi,\tau} > 0,$$

and therefore $\mathcal{V}(z, q) > 0$.

Finally, let $q = I(z)$. Equation (6) and the definition of Gittins index imply that for every admissible pair,

$$N_{\pi,\tau} - I(z)D_{\pi,\tau} \leq 0.$$

So $\mathcal{V}\big(z, I(z)\big) \leq 0$. Conversely, for every $\varepsilon > 0$, there exists an admissible pair $(\pi_\varepsilon, \tau_\varepsilon)$ satisfying

$$\frac{N_{\pi_\varepsilon,\tau_\varepsilon}}{D_{\pi_\varepsilon,\tau_\varepsilon}} \geq I(z) - \varepsilon.$$

Consequently,

$$N_{\pi_\varepsilon,\tau_\varepsilon} - I(z)D_{\pi_\varepsilon,\tau_\varepsilon} \geq -\varepsilon D_{\pi_\varepsilon,\tau_\varepsilon} \geq -\frac{\varepsilon}{1-\delta}.$$

Taking the supremum over admissible pairs and then letting $\varepsilon \downarrow 0$ gives $\mathcal{V}\big(z, I(z)\big) \geq 0$. Thus

$$\mathcal{V}\big(z, I(z)\big) = 0.$$

The three cases establish the result. □

Appendix B Brownian R&D

B.1 Well-Posedness of the Brownian R&D Problem

We first verify both requirements of Assumption 4 in the Brownian R&D application. The argument is uniform over continuation policies and therefore establishes finiteness of the global value function, not only of a particular single-arm problem.

For a realized informational state S, write

$$\mathbb{E}_S^\Sigma[\cdot] \equiv \mathbb{E}^\Sigma\left[\,\cdot\,|S\right].$$

Throughout Appendix B, we will suppose that the profitability landscape is

$$\Psi_x = \bar{y}_0 + \mu x + \sigma W_x, \qquad \sigma > 0,$$

and that Assumptions 5, 6, and 7 hold.

Lemma 7. *For every reachable finite informational state S, there exists a constant $K_S < \infty$ such that*

$$\sup_\Sigma \mathbb{E}_S^\Sigma\left[|Y_t|\right] \leq K_S(1+t) \qquad \textit{for every } t \geq 0.$$

Consequently, for every $T \geq 0$,

$$\sup_\Sigma \mathbb{E}_S^\Sigma\left[\sum_{t=T}^\infty \delta^t |\chi_t Y_t - (1-\chi_t)c|\right] \leq \sum_{t=T}^\infty \delta^t\big[K_S(1+t)+c\big].$$

The bound is finite at $T = 0$ and converges to zero as $T \to \infty$. Thus, Assumption 4 is satisfied in the Brownian R&D application, and $V(S) \in \mathbb{R}$ for every reachable finite state S.

Proof. Fix a reachable finite informational state

$$S = \{(x_i, y_i) : i = 0, \dots, m\}, \qquad 0 = x_0 < x_1 < \cdots < x_m.$$

We first control the unresolved profitability landscape on the compact interval already spanned by the startup. For

$$\ell_i \equiv x_i - x_{i-1}, \qquad i = 1, \dots, m,$$

the conditional process on $[x_{i-1}, x_i]$ can be represented as

$$\Psi_{x_{i-1}+u\ell_i} = (1-u)y_{i-1} + uy_i + \sigma\sqrt{\ell_i}\,\mathbb{B}_i(u), \qquad u \in [0,1],$$

where $\mathbb{B}_i$ is a standard Brownian bridge. Hence, defining $B_s \equiv \sup_{0\le x\le r_s} |\Psi_x|$,[25] we get

$$B_s \le \max_{0\le i\le m} |y_i| + \sigma \sum_{i=1}^{m} \sqrt{\ell_i}\,\|\mathbb{B}_i\|_\infty,$$

where $\|\mathbb{B}_i\|_\infty \equiv \sup_{u\in[0,1]} |\mathbb{B}_i(u)|$. Since $\mathbb{E}[\|\mathbb{B}_i\|_\infty] < \infty$, we obtain

$$b_s \equiv \mathbb{E}_s[B_s] < \infty.$$

We next control the part of the landscape generated by future frontier expansions. Let $(\Delta_j)_{j\ge 0}$ be an i.i.d. sequence with law G, from which the successive frontier steps are drawn whenever the frontier is activated. Define

$$\overline{L}_t \equiv \sum_{j=0}^{t} \Delta_j.$$

Since the startup can expand the frontier at most once in each period, the total distance added to the frontier by the end of period t is bounded pathwise by $\overline{L}_t$, under every admissible policy.

Conditional on state S, the profitability landscape beyond the frontier x_m can be written as

$$\Psi_{x_m+u} = y_m + \mu u + \sigma\widetilde{W}_u, \qquad u \ge 0,$$

where $\widetilde{W}$ is a standard Brownian motion independent of the previously unresolved bridges and of the frontier-step sequence. Every product observed by period t is therefore located in $[0, x_m + \overline{L}_t]$, and its profitability satisfies the pathwise bound

$$|Y_t| \le B_s + |\mu|\overline{L}_t + \sigma \sup_{0\le u\le \overline{L}_t} |\widetilde{W}_u|.$$

By assumption 6, $m_G \equiv \mathbb{E}_G[\Delta] < \infty$. Then

$$\mathbb{E}[\overline{L}_t] = (t+1)m_G < \infty.$$

By the Brownian maximal inequality, there exists a finite constant C_W such that

$$\mathbb{E}\left[\sup_{0\le u\le \overline{L}_t} |\widetilde{W}_u| \,\middle|\, \overline{L}_t\right] \le C_W\sqrt{\overline{L}_t}.$$

[25] When $m = 0$, we use the convention $B_s = |y_0|$.

Taking expectations and applying Jensen's inequality to the concave function $f(x) = \sqrt{x}$ gives

$$\mathbb{E}\left[\sup_{0\le u\le \overline{L}_t} |\widetilde{W}_u|\right] \le C_W \mathbb{E}[\sqrt{\overline{L}_t}] \le C_W \sqrt{(t+1)m_G}.$$

The bound is independent of the continuation policy. It follows that

$$\sup_{\Sigma} \mathbb{E}_s^{\Sigma}[|Y_t|] \le b_s + |\mu|(t+1)m_G + \sigma C_W \sqrt{(t+1)m_G} \le K_s(1+t)$$

for some finite constant K_s.

In the R&D application, the per-period payoff is

$$R_t = \chi_t Y_t - (1-\chi_t)c.$$

Therefore,

$$|R_t| \le |Y_t| + c.$$

By Tonelli's theorem and the preceding policy-uniform bound, for every $T \ge 0$,

$$\sup_{\Sigma} \mathbb{E}_S^{\Sigma}\left[\sum_{t=T}^{\infty} \delta^t |R_t|\right] \le \sum_{t=T}^{\infty} \delta^t \big[K_S(1+t)+c\big] = \delta^T \left[K_S\left(\frac{T+1}{1-\delta} + \frac{\delta}{(1-\delta)^2}\right) + \frac{c}{1-\delta}\right].$$

At $T = 0$ this gives a finite uniform bound on total absolute discounted rewards; as $T \to \infty$ it gives uniformly vanishing discounted tails. This verifies both requirements of Assumption 4. Every admissible policy has expected discounted payoff bounded in absolute value by the expression at $T = 0$, so $V(S) \in \mathbb{R}$. □

Corollary 3. *For every $y, y_L, y_R \in \mathbb{R}$ and every $\ell > 0$,*

$$I^{\text{gap}}(\ell, y_L, y_R) \in \mathbb{R}, \qquad I^{\text{front}}(y) \in \mathbb{R}.$$

Proof. Consider first the descendant process initiated by the gap $\mathsf{G}(\ell, y_L, y_R)$. Every product generated by this process lies inside the root interval. Under the conditional Brownian-bridge representation, define

$$B^{\text{g}}_{\ell,y_L,y_R} \equiv \sup_{u\in[0,1]} \Big|(1-u)y_L + uy_R + \sigma\sqrt{\ell}\,\mathbb{B}(u)\Big|,$$

where $\mathbb{B}$ is a standard Brownian bridge. Then, the proof of lemma 7 implies

$$\mathbb{E}\left[B^{\text{g}}_{\ell,y_L,y_R}\right] < \infty.$$

Every safe descendant has payoff bounded in absolute value by $B^{\text{g}}_{\ell,y_L,y_R}$, while every gap activation yields the payoff $-c$. Hence, for every admissible policy-stopping pair

$$\pi \in \Pi\big(\mathsf{G}(\ell, y_L, y_R)\big), \qquad \tau \in \mathcal{T}\big(\mathsf{G}(\ell, y_L, y_R), \pi\big),$$

the numerator of the normalized index satisfies

$$\left|\mathbb{E}^{\pi}\left[\sum_{n=0}^{\tau-1} \delta^n R(Z_n^{\pi})\right]\right| \le \frac{\mathbb{E}[B^{\text{g}}_{\ell,y_L,y_R}] + c}{1-\delta}.$$

Since

$$\mathbb{E}^{\pi}\left[\sum_{n=0}^{\tau-1}\delta^{n}\right]\geq 1,$$

all discounted-average payoffs entering the definition of $I^{\mathrm{gap}}(\ell, y_L, y_R)$ are uniformly bounded. Moreover, retiring the arm immediately after its first activation yields the finite value $-c$. Therefore, $I^{\mathrm{gap}}(\ell, y_L, y_R) \in \mathbb{R}$.

For a frontier arm $\mathsf{F}(y)$, the argument in Lemma 7 applies to its single-arm descendant process, with initial frontier payoff y. Thus there exists a constant $K_y^{\mathrm{f}} < \infty$ such that

$$\sup_{\pi}\mathbb{E}_{y}^{\pi}\left[|Y_n^{\pi}|\right]\leq K_y^{\mathrm{f}}(1+n).$$

Consequently, uniformly over all admissible policy-stopping pairs,

$$\left|\mathbb{E}_{y}^{\pi}\left[\sum_{n=0}^{\tau-1}\delta^{n}R(Z_n^{\pi})\right]\right|\leq\frac{K_y^{\mathrm{f}}}{(1-\delta)^2}+\frac{c}{1-\delta}<\infty.$$

Again, the denominator of the normalized index is at least one, and the policy that retires after the initial frontier activation yields $-c$. Therefore, $I^{\mathrm{front}}(y) \in \mathbb{R}$. □

B.2 Initialization of the R&D Process

We establish the exact condition under which the startup expands the frontier at date 0. Recall that the initial portfolio contains only the safe arm $\mathsf{S}(\bar{y}_0)$ and the frontier arm $\mathsf{F}(\bar{y}_0)$.

Proposition 5. *Under Assumptions 5 and 6,*

$$0<\kappa^{F}(\mu,\sigma,F,G,\delta)<\infty.$$

Moreover,

$$I^{\mathrm{front}}(\bar{y}_0)\begin{cases}>\bar{y}_0, & \bar{y}_0<k^{F}(\mu,\sigma,F,G,c,\delta),\\ =\bar{y}_0, & \bar{y}_0=k^{F}(\mu,\sigma,F,G,c,\delta),\\ <\bar{y}_0, & \bar{y}_0>k^{F}(\mu,\sigma,F,G,c,\delta).\end{cases}$$

Consequently, at date 0*, frontier expansion is uniquely optimal if and only if Assumption 7 holds.*

Proof. We first establish that $\kappa^F < \infty$ is well defined. Consider an admissible policy-stopping pair (π, τ) for the descendant process initiated by $\mathsf{F}(0)$, and define

$$P_{\pi,\tau}\equiv\mathbb{E}_0^{\pi}\left[\sum_{n=0}^{\tau-1}\delta^{n}\chi_n^{\pi}\widetilde{Y}_n^{\pi}\right],\qquad Q_{\pi,\tau}\equiv\mathbb{E}_0^{\pi}\left[\sum_{n=0}^{\tau-1}\delta^{n}(1-\chi_n^{\pi})\right],\qquad\text{and}\qquad D_{\pi,\tau}\equiv\mathbb{E}_0^{\pi}\left[\sum_{n=0}^{\tau-1}\delta^{n}\right].$$

Since the initial arm is a frontier arm, $\chi_0^{\pi} = 0$. Because $\tau \geq 1$, it follows that $Q_{\pi,\tau} \geq 1$. Moreover,

$$1\leq D_{\pi,\tau}\leq\frac{1}{1-\delta},\qquad Q_{\pi,\tau}\leq D_{\pi,\tau}.$$

Discounted integrability implies that

$$\sup_{\pi,\tau}\mathbb{E}_0^{\pi}\left[\sum_{n=0}^{\tau-1}\delta^{n}\left(\chi_n^{\pi}\left|\widetilde{Y}_n^{\pi}\right|+(1-\chi_n^{\pi})c\right)\right]<\infty.$$

Hence,

$$\sup_{\pi,\tau} P_{\pi,\tau} < \infty \qquad \text{and} \qquad \kappa^F = \sup_{\pi,\tau} \frac{P_{\pi,\tau}}{Q_{\pi,\tau}} < \infty,$$

since $Q_{\pi,\tau} \geq 1$,

To see that $\kappa^F > 0$, let

$$Z \equiv \mu\Delta + \sigma\sqrt{\Delta}\,\varepsilon,$$

where $\Delta \sim G$, and $\varepsilon \sim \mathcal{N}(0,1)$, independently. Consider the policy that expands the frontier once and, if $Z > 0$, operates the resulting safe arm for one period before retiring the descendant process. If $Z \leq 0$, the process is retired immediately after the frontier experiment. This policy has $Q_{\pi,\tau} = 1$ and $P_{\pi,\tau} = \delta\mathbb{E}[Z^+]$. Because $\sigma > 0$ and $\Delta > 0$ almost surely, $\mathbb{E}[Z^+] > 0$. Therefore $P_{\pi,\tau} > 0$ for every $N \geq 1$, and hence $\kappa^F > 0$.

We now characterize the initial decision. Fix an admissible pair (π, τ) for the descendant process initiated by $\mathsf{F}(0)$. Vertical translation of every observed payoff by $\bar{y}_0$ produces an admissible policy-stopping pair for the process initiated by $\mathsf{F}(\bar{y}_0)$. Conversely, subtracting $\bar{y}_0$ from every payoff maps every admissible pair starting from $\mathsf{F}(\bar{y}_0)$ into one starting from $\mathsf{F}(0)$. This correspondence preserves the sequence of arm types, the stopping time, and the locations generated by the research protocols. Every safe payoff is increased by $\bar{y}_0$, whereas every R&D activation continues to yield the payoff $-c$. Define

$$S_{\pi,\tau} \equiv \mathbb{E}_0^{\pi}\left[\sum_{n=0}^{\tau-1} \delta^n \chi_n^{\pi}\right].$$

Then

$$D_{\pi,\tau} = S_{\pi,\tau} + Q_{\pi,\tau}.$$

The discounted-average payoff generated by the translated pair starting from $\mathsf{F}(\bar{y}_0)$ is

$$A_{\pi,\tau}(\bar{y}_0; c) = \frac{\bar{y}_0 S_{\pi,\tau} + P_{\pi,\tau} - cQ_{\pi,\tau}}{D_{\pi,\tau}}.$$

Subtracting the index of the initial safe product gives

$$A_{\pi,\tau}(\bar{y}_0; c) - y_0 = \frac{P_{\pi,\tau} - (c + \bar{y}_0)Q_{\pi,\tau}}{D_{\pi,\tau}} = \frac{Q_{\pi,\tau}}{D_{\pi,\tau}}\left(\frac{P_{\pi,\tau}}{Q_{\pi,\tau}} - (c + \bar{y}_0)\right). \tag{7}$$

Notice that

$$1 - \delta \leq \frac{Q_{\pi,\tau}}{D_{\pi,\tau}} \leq 1,$$

where the lower bound follows from $Q_{\pi,\tau} \geq 1$ and $D_{\pi,\tau} \leq (1-\delta)^{-1}$. Suppose first that $\bar{y}_0 + c < \kappa^F$. By the definition of the supremum, there exists an admissible pair satisfying

$$\frac{P_{\pi,\tau}}{Q_{\pi,\tau}} > \bar{y}_0 + c.$$

Equation (7) then implies $A_{\pi,\tau}(\bar{y}_0; c) > \bar{y}_0$. Consequently, $I^{\text{front}}(\bar{y}_0) > \bar{y}_0$.

Next suppose that $\bar{y}_0 + c > \kappa^F$. For every admissible pair,

$$\frac{P_{\pi,\tau}}{Q_{\pi,\tau}} - (\bar{y}_0 + c) \leq \kappa^F - (\bar{y}_0 + c) < 0.$$

Using the lower bound on $Q_{\pi,\tau}/D_{\pi,\tau}$ in (7), we obtain

$$A_{\pi,\tau}(\bar{y}_0; c) - \bar{y}_0 \leq (1-\delta)\left[\kappa^F - (\bar{y}_0 + c)\right] < 0.$$

This inequality holds uniformly over all admissible pairs. Therefore, $I^{\text{front}}(\bar{y}_0) < \bar{y}_0$.

Finally, suppose that $\bar{y}_0 + c = \kappa^F$. Equation (7) implies $A_{\pi,\tau}(\bar{y}_0; c) \leq \bar{y}_0$ for every admissible pair, and hence $I^{\text{front}}(\bar{y}_0) \leq \bar{y}_0$. For every $\varepsilon > 0$, however, there exists an admissible pair such that

$$\frac{P_{\pi,\tau}}{Q_{\pi,\tau}} > \kappa^F - \varepsilon.$$

Since $Q_{\pi,\tau}/D_{\pi,\tau} \leq 1$,

$$A_{\pi,\tau}(\bar{y}_0; c) - \bar{y}_0 > -\varepsilon.$$

Letting $\varepsilon \downarrow 0$ gives $I^{\text{front}}(\bar{y}_0) \geq \bar{y}_0$. Thus, $I^{\text{front}}(\bar{y}_0) = \bar{y}_0$.

At date 0, the only competing arm is the initial safe arm $\mathsf{S}(\bar{y}_0)$, whose index is $\bar{y}_0$. The policy statements therefore follow from Theorem 1. □

B.3 Brownian Geometry inside a Bounded Gap

We now derive a normalized representation of the descendant process generated by a bounded gap. Fix a physical realization of a root gap with endpoints (x_L, y_L) and (x_R, y_R), and define

$$\ell \equiv x_R - x_L > 0, \qquad s \equiv y_R - y_L, \qquad \bar{y} \equiv \frac{y_L + y_R}{2}.$$

The endpoint products are boundary conditions for the descendant process; as explained in Section 2, they are not safe descendants of the root gap.

Enumerate the new products generated by activations of the root gap and its descendant gaps in their order of discovery. Write

$$(\hat{x}_k, Y_k), \qquad k = 1, 2, \ldots,$$

for their physical locations and profitabilities. We use the labels L and R for the two root endpoints:

$$(\hat{x}_L, Y_L) = (x_L, y_L), \qquad (\hat{x}_R, Y_R) = (x_R, y_R).$$

Lemma 8. *Fix an admissible policy for the single-arm descendant process initiated by $\mathsf{G}(\ell, y_L, y_R)$. For every label $i \in \{L, R, 1, 2, \ldots\}$ that is generated before retirement, there exist unique normalized variables (W_i, ζ_i) satisfying*

$$\hat{x}_i = x_L + \ell W_i \qquad \text{and} \qquad Y_i = y_L + W_i s + \sigma\sqrt{\ell}\,\zeta_i.$$

Initialize

$$W_L = 0, \qquad W_R = 1, \qquad \zeta_L = \zeta_R = 0.$$

Whenever the kth gap activation occurs, let $A_k, B_k \in \{L, R, 1, \ldots, k-1\}$ denote the labels of the two endpoints of the selected descendant gap, with $\hat{x}_{A_k} < \hat{x}_{B_k}$. Then

$$W_k = (1 - U_k)W_{A_k} + U_k W_{B_k}, \qquad U_k \sim F,$$

and

$$\zeta_k = (1 - U_k)\zeta_{A_k} + U_k\zeta_{B_k} + \sqrt{W_{B_k} - W_{A_k}}\,\xi_k, \qquad \xi_k \mid U_k \sim \mathcal{N}\big(0, U_k(1 - U_k)\big),$$

where U_k and $\xi_k \mid U_k$ are independent of the history preceding the kth gap activation.

For fixed root parameters (x_L, ℓ, y_L, s), the transformation between the physical descendant history and its normalized counterpart is deterministic and invertible. Consequently, the two histories generate the same filtration. Moreover, under any fixed policy written in normalized coordinates, the law of the normalized process $\{(W_k, \zeta_k)\}_{k \geq 1}$ does not depend on (x_L, ℓ, y_L, s).

Proof. We proceed by induction over the sequence of gap activations. The representation holds at the two root endpoints. Indeed,

$$\hat{x}_L = x_L = x_L + \ell W_L \qquad \text{and} \qquad Y_L = y_L = y_L + W_L s + \sigma\sqrt{\ell}\,\zeta_L,$$

because $W_L = \zeta_L = 0$. Similarly,

$$\hat{x}_R = x_R = x_L + \ell = x_L + \ell W_R, \qquad \text{and} \qquad Y_R = y_R = y_L + s = y_L + W_R s + \sigma\sqrt{\ell}\,\zeta_R,$$

because $W_R = 1$ and $\zeta_R = 0$.

Suppose that the representation holds for every product observed before the kth gap activation. At that activation, the policy selects the descendant gap whose endpoint products are $(\hat{x}_{A_k}, Y_{A_k})$ and $(\hat{x}_{B_k}, Y_{B_k})$, where $\hat{x}_{A_k} < \hat{x}_{B_k}$. By Assumption 5, the new product is developed at

$$\hat{x}_k = (1 - U_k)\hat{x}_{A_k} + U_k \hat{x}_{B_k}, \qquad U_k \sim F.$$

Using the induction hypothesis,

$$\hat{x}_{A_k} = x_L + \ell W_{A_k}, \qquad \hat{x}_{B_k} = x_L + \ell W_{B_k},$$

we obtain

$$\hat{x}_k = (1 - U_k)\,(x_L + \ell W_{A_k}) + U_k\,(x_L + \ell W_{B_k}) = x_L + \ell\,[(1 - U_k)W_{A_k} + U_k W_{B_k}]\,.$$

Therefore,

$$W_k = (1 - U_k)W_{A_k} + U_k W_{B_k}.$$

Conditional on the pre-activation history, the two endpoint values, and U_k, the unresolved profitability process inside $[\hat{x}_{A_k}, \hat{x}_{B_k}]$ is a Brownian bridge. Hence

$$Y_k = (1 - U_k)Y_{A_k} + U_k Y_{B_k} + \sigma\sqrt{\hat{x}_{B_k} - \hat{x}_{A_k}}\,\xi_k, \qquad \xi_k \mid U_k \sim \mathcal{N}\big(0, U_k(1 - U_k)\big),$$

and the bridge innovation is independent of the preceding history. Notice that the drift μ does not enter this conditional representation: after conditioning on both endpoints, it is absorbed by the linear interpolation between their realized values.

By the induction hypothesis,

$$Y_{A_k} = y_L + W_{A_k} s + \sigma\sqrt{\ell}\,\zeta_{A_k} \qquad \text{and} \qquad Y_{B_k} = y_L + W_{B_k} s + \sigma\sqrt{\ell}\,\zeta_{B_k}.$$

Moreover,

$$\hat{x}_{B_k} - \hat{x}_{A_k} = \ell\,(W_{B_k} - W_{A_k})\,.$$

Substituting these expressions gives

$$\begin{aligned} Y_k &= (1 - U_k)\left[y_L + W_{A_k} s + \sigma\sqrt{\ell}\,\zeta_{A_k}\right] + U_k\left[y_L + W_{B_k} s + \sigma\sqrt{\ell}\,\zeta_{B_k}\right] + \sigma\sqrt{\ell\,(W_{B_k} - W_{A_k})}\,\xi_k \\ &= y_L + [(1 - U_k)W_{A_k} + U_k W_{B_k}]\,s + \sigma\sqrt{\ell}\left[(1 - U_k)\zeta_{A_k} + U_k\zeta_{B_k} + \sqrt{W_{B_k} - W_{A_k}}\,\xi_k\right]. \end{aligned}$$

Using the expression for W_k, this becomes

$$Y_k = y_L + W_k s + \sigma\sqrt{\ell}\,\zeta_k,$$

where

$$\zeta_k = (1 - U_k)\zeta_{A_k} + U_k\zeta_{B_k} + \sqrt{W_{B_k} - W_{A_k}}\,\xi_k.$$

This completes the induction.

For fixed (x_L, ℓ, y_L, s), the inverse transformation is

$$W_i = \frac{\hat{x}_i - x_L}{\ell} \qquad \text{and} \qquad \zeta_i = \frac{Y_i - y_L - W_i s}{\sigma\sqrt{\ell}}.$$

Thus, keeping the sequence of selected arm labels and arm types unchanged, the physical and normalized descendant histories are deterministically equivalent. In particular, they generate the same filtration.

Finally, the recursion for (W_k, ζ_k) depends only on the normalized history, the draws $U_k \sim F$, and the Brownian-bridge innovations ξ_k. None of these objects depends on the root-specific parameters (x_L, ℓ, y_L, s). Therefore, under any fixed policy written as a measurable function of normalized histories, the law of the normalized process is independent of those parameters. □

Lemma 8 allows us to identify the policy spaces of all single-gap descendant problems with a common normalized policy space. Let Π^{g} denote the resulting class of admissible policies on normalized descendant histories, and, for $\pi \in \Pi^{\mathrm{g}}$, let $\mathcal{T}^{\mathrm{g}}(\pi)$ denote the corresponding set of strictly positive stopping times. Although the optimizing policy may differ across root-gap parameters, the class of measurable normalized decision rules is the same for every root gap.

For $x > 0$, and $\ell = x^2$, let

$$\mathcal{V}^{\mathrm{gap}}(x^2, \bar{y}, s; q) \equiv \mathcal{V}\left(\mathsf{G}\left(x^2, \bar{y} - \frac{s}{2}, \bar{y} + \frac{s}{2}\right), q\right)$$

be the gap discounted excess value relative to q, as defined in Appendix A.1. Fix $\pi \in \Pi^{\mathrm{g}}$, and $\tau \in \mathcal{T}^{\mathrm{g}}(\pi)$. At local date n, let

$$\chi_n^\pi \equiv \mathbf{1}_{\{Z_n^\pi \in \mathcal{Z}_{\mathrm{s}}\}}.$$

On the event $\{\chi_n^\pi = 1\}$, let (W_n^π, ζ_n^π) denote the normalized coordinates of the product represented by the selected safe arm. The payoff of the selected safe arm can then be written as

$$Y_n^\pi = \bar{y} + \left(W_n^\pi - \frac{1}{2}\right) s + x\sigma\zeta_n^\pi \qquad \text{on } \{\chi_n^\pi = 1\}.$$

Define the excess value delivered by the fixed pair (π, τ) by

$$\mathcal{V}_{\pi,\tau}^{\mathrm{gap}}(x^2, \bar{y}, s; q) \equiv \mathbb{E}^\pi\left[\sum_{n=0}^{\tau-1} \delta^n \left[\chi_n^\pi \left(Y_n^\pi - q\right) - (1 - \chi_n^\pi)(c + q)\right]\right].$$

Lemma 9. *For every normalized policy-stopping pair $\pi \in \Pi^{\mathrm{g}}$, $\tau \in \mathcal{T}^{\mathrm{g}}(\pi)$, there exist finite coefficients $C_{\pi,\tau}(\bar{y}, q)$, $H_{\pi,\tau}$, and $B_{\pi,\tau}$, such that*

$$\mathcal{V}_{\pi,\tau}^{\mathrm{gap}}(x^2, \bar{y}, s; q) = C_{\pi,\tau}(\bar{y}, q) + sH_{\pi,\tau} + xB_{\pi,\tau}.$$

The coefficients are

$$C_{\pi,\tau}(\bar{y}, q) \equiv \mathbb{E}^\pi\left[\sum_{n=0}^{\tau-1} \delta^n \left[\chi_n^\pi(\bar{y} - q) - (1 - \chi_n^\pi)(c + q)\right]\right],$$

$$H_{\pi,\tau} \equiv \mathbb{E}^\pi\left[\sum_{n=0}^{\tau-1} \delta^n \chi_n^\pi \left(W_n^\pi - \frac{1}{2}\right)\right],$$

and

$$B_{\pi,\tau} \equiv \mathbb{E}^{\pi}\left[\sum_{n=0}^{\tau-1} \delta^n \chi_n^{\pi} \sigma \zeta_n^{\pi}\right].$$

For a fixed normalized pair (π,τ), the coefficients $H_{\pi,\tau}$ and $B_{\pi,\tau}$ do not depend on $(x,\bar{y},s,q)$, while $C_{\pi,\tau}(\bar{y},q)$ does not depend on (x,s).

Consequently,

$$\mathcal{V}^{\mathrm{gap}}(x^2,\bar{y},s;q) = \sup_{\substack{\pi\in\Pi^{\mathrm{g}} \\ \tau\in\mathcal{T}^{\mathrm{g}}(\pi)}} \left\{C_{\pi,\tau}(\bar{y},q) + sH_{\pi,\tau} + xB_{\pi,\tau}\right\}.$$

Proof. The excess payoff generated at local date n is

$$\chi_n^{\pi}\left(Y_n^{\pi} - \lambda\right) - (1-\chi_n^{\pi})(c+\lambda).$$

By Lemma 8, the payoff of a selected safe descendant satisfies

$$Y_n^{\pi} = \bar{y} + \left(W_n^{\pi} - \frac{1}{2}\right)s + x\sigma\zeta_n^{\pi}.$$

Substituting this representation gives

$$\mathcal{V}_{\pi,\tau}^{\mathrm{gap}}(x^2,\bar{y},s;q) = \mathbb{E}^{\pi}\left[\sum_{n=0}^{\tau-1}\delta^n\left\{\chi_n^{\pi}\left[\bar{y} + \left(W_n^{\pi}-\frac{1}{2}\right)s + x\sigma\zeta_n^{\pi} - q\right] - (1-\chi_n^{\pi})(c+q)\right\}\right]$$

$$= \mathbb{E}^{\pi}\left[\sum_{n=0}^{\tau-1}\delta^n\left[\chi_n^{\pi}(\bar{y}-q) - (1-\chi_n^{\pi})(c+q)\right]\right] + s\mathbb{E}^{\pi}\left[\sum_{n=0}^{\tau-1}\delta^n\chi_n^{\pi}\left(W_n^{\pi}-\frac{1}{2}\right)\right] + x\mathbb{E}^{\pi}\left[\sum_{n=0}^{\tau-1}\delta^n\chi_n^{\pi}\sigma\zeta_n^{\pi}\right].$$

The three terms on the right-hand side are precisely

$$C_{\pi,\tau}(\bar{y},q), \qquad sH_{\pi,\tau}, \qquad xB_{\pi,\tau}.$$

Their finiteness follows from Lemma 7.

For a fixed normalized policy-stopping pair, the law of $(\chi_n^{\pi}, W_n^{\pi}, \zeta_n^{\pi})_{n\geq 0}$ and of τ is independent of the root-gap parameters by Lemma 8. Hence $H_{\pi,\tau}$ and $B_{\pi,\tau}$ are independent of $(x,\bar{y},s,q)$, while $C_{\pi,\tau}(\bar{y},q)$ depends on the root parameters only through the displayed deterministic terms $\bar{y}$ and q.

Finally, Lemma 8 shows that the correspondence between physical and normalized histories is bijective. Therefore, taking the supremum over all admissible physical policy-stopping pairs is equivalent to taking the supremum over $\pi\in\Pi^{\mathrm{g}},\ \tau\in\mathcal{T}^{\mathrm{g}}(\pi)$. This yields

$$\mathcal{V}^{\mathrm{gap}}(x^2,\bar{y},s;q) = \sup_{\pi,\tau}\left\{C_{\pi,\tau}(\bar{y},q) + sH_{\pi,\tau} + xB_{\pi,\tau}\right\}.$$

□

By Lemma 6, the gap index is the unique charge q satisfying

$$\mathcal{V}^{\mathrm{gap}}\left(x^2,\bar{y},s;q\right) = 0.$$

Equivalently,

$$\mathcal{V}^{\mathrm{gap}}\left(x^2,\bar{y},s;I^{\mathrm{gap}}\left(x^2,\bar{y}-\frac{s}{2},\bar{y}+\frac{s}{2}\right)\right) = 0.$$

The affine decomposition is the common input for the length and steepness comparative statics proved below. In particular, for fixed $(\bar{y},s,q)$, the excess value is a supremum of affine functions of $x=\sqrt{\ell}$; for fixed $(x,\bar{y},q)$, it is a supremum of affine functions of s.

B.4 Properties of the Gap Index

For convenience, we use both endpoint and midpoint–steepness notation:

$$\bar{y} \equiv \frac{y_L + y_R}{2}, \qquad s \equiv y_R - y_L.$$

Thus,

$$I^{\text{gap}}(\ell, y_L, y_R) = I^{\text{gap}}\left(\ell, \bar{y} - \frac{s}{2}, \bar{y} + \frac{s}{2}\right).$$

We use the same convention for the pair-specific and optimized excess-value functions:

$$\mathcal{V}^{\text{gap}}_{\pi,\tau}(\ell, y_L, y_R, q) \equiv \mathcal{V}^{\text{gap}}_{\pi,\tau}(\ell, \bar{y}, s, q), \qquad \text{and} \qquad \mathcal{V}^{\text{gap}}(\ell, y_L, y_R, q) \equiv \mathcal{V}^{\text{gap}}(\ell, \bar{y}, s, q).$$

Lemma 10 (Properties of the gap index)**.** *The gap index satisfies the following properties:*

(i) ***Endpoint profitability.*** *For every $\ell > 0$, $(y_L, y_R) \in \mathbb{R}^2$,*

$$I^{\text{gap}}(\ell, y_L, y_R) > -c.$$

Moreover, for every $a > 0$,

$$0 < I^{\text{gap}}(\ell, y_L{+}a, y_R) - I^{\text{gap}}(\ell, y_L, y_R) \le \delta a \quad \textit{and} \quad 0 < I^{\text{gap}}(\ell, y_L, y_R{+}a) - I^{\text{gap}}(\ell, y_L, y_R) \le \delta a.$$

In particular, the gap index is δ-Lipschitz continuous and strictly increasing in each endpoint profitability. Hence, holding s fixed, it is strictly increasing in average profitability $\bar{y}$.

(ii) ***Technological distance.*** *For every fixed $(y_L, y_R) \in \mathbb{R}^2$, the map*

$$x \longmapsto I^{\text{gap}}(x^2, y_L, y_R)$$

is finite, continuous, convex, and weakly increasing on $(0, \infty)$. Moreover,

$$\lim_{\ell \to \infty} I^{\text{gap}}(\ell, y_L, y_R) = +\infty.$$

(iii) ***Steepness.*** *For fixed $\ell > 0$ and $\bar{y} \in \mathbb{R}$, the map*

$$s \longmapsto I^{\text{gap}}\left(\ell, \bar{y} - \frac{s}{2}, \bar{y} + \frac{s}{2}\right)$$

is finite, continuous, even, and convex. It is therefore weakly increasing in $|s|$.

(iv) ***Bad-endpoint limits.*** *For every fixed $\ell > 0$ and $y \in \mathbb{R}$,*

$$\lim_{b \to -\infty} I^{\text{gap}}(\ell, y, b) = -c \qquad \textit{and} \qquad \lim_{b \to -\infty} I^{\text{gap}}(\ell, b, y) = -c.$$

(v) ***Small-gap bound.*** *Let $m \equiv \max\{y_L, y_R\}$. Then*

$$\limsup_{\ell \downarrow 0} I^{\text{gap}}(\ell, y_L, y_R) \le \max\left\{-c,\, \delta m - (1-\delta)c\right\}.$$

More generally, if

$$\ell_k \downarrow 0, \qquad y_{L,k} \longrightarrow y_L, \qquad y_{R,k} \longrightarrow y_R,$$

then

$$\limsup_{k \to \infty} I^{\text{gap}}(\ell_k, y_{L,k}, y_{R,k}) \le \max\left\{-c,\, \delta m - (1-\delta)c\right\}.$$

Proof. We proceed in five steps.

Step 1: endpoint profitability and continuity.

We first establish the strict lower bound. Let Y denote the payoff of the first prototype generated from the root gap:

$$Y = (1-U)y_L + Uy_R + \sigma\sqrt{\ell U(1-U)}\,\varepsilon, \qquad U \sim F, \quad \varepsilon \sim N(0,1),$$

with U and ε independent. Since $F((0,1)) = 1$ and $\sigma > 0$, the conditional variance of Y is strictly positive almost surely. Hence $\mathbb{P}(Y > K) > 0$ for every finite K. Fix $K > -c$ and consider the policy that activates the root gap once and then operates the newly generated safe arm for one period if $Y > K$, retiring immediately otherwise. Its discounted-average payoff is

$$\frac{-c + \delta\mathbb{E}\big[Y\mathbf{1}_{\{Y>K\}}\big]}{1+\delta\mathbb{P}(Y>K)}.$$

Its difference from $-c$ is

$$\frac{\delta\mathbb{E}\big[(Y+c)\mathbf{1}_{\{Y>K\}}\big]}{1+\delta\mathbb{P}(Y>K)} > 0.$$

Therefore,

$$I^{\mathrm{gap}}(\ell, y_L, y_R) > -c.$$

Fix $\ell > 0$, $(y_L, y_R) \in \mathbb{R}^2$, and let

$$q^* \equiv I^{\mathrm{gap}}(\ell, y_L, y_R).$$

Consider a normalized policy-stopping pair (π, τ). Let

$$D_{\pi,\tau} \equiv \mathbb{E}^\pi\left[\sum_{n=0}^{\tau-1}\delta^n\right], \qquad L_{\pi,\tau} \equiv \mathbb{E}^\pi\left[\sum_{n=0}^{\tau-1}\delta^n\chi_n^\pi(1-W_n^\pi)\right], \qquad \text{and} \qquad R_{\pi,\tau} \equiv \mathbb{E}^\pi\left[\sum_{n=0}^{\tau-1}\delta^n\chi_n^\pi W_n^\pi\right].$$

By Lemma 8, increasing the left endpoint from y_L to $y_L + a$ increases the payoff of every safe descendant at normalized location W_n^π by $a(1-W_n^\pi)$. Therefore,

$$\mathcal{V}^{\mathrm{gap}}_{\pi,\tau}(\ell, y_L + a, y_R, q^*) - \mathcal{V}^{\mathrm{gap}}_{\pi,\tau}(\ell, y_L, y_R, q^*) = aL_{\pi,\tau}. \tag{8}$$

Similarly,

$$\mathcal{V}^{\mathrm{gap}}_{\pi,\tau}(\ell, y_L, y_R + a, q^*) - \mathcal{V}^{\mathrm{gap}}_{\pi,\tau}(\ell, y_L, y_R, q^*) = aR_{\pi,\tau}. \tag{9}$$

Since the stopping time is strictly positive, $\chi_0^\pi = 0$. Moreover, $0 \le W_n^\pi \le 1$. Hence

$$0 \le L_{\pi,\tau} \le \mathbb{E}^\pi\left[\sum_{n=1}^{\tau-1}\delta^n\right] \le \delta D_{\pi,\tau},$$

and likewise

$$0 \le R_{\pi,\tau} \le \delta D_{\pi,\tau}.$$

Using (8) and the definition of excess value relative to $q^* + \delta a$, for every admissible pair,

$$\mathcal{V}^{\mathrm{gap}}_{\pi,\tau}(\ell, y_L + a, y_R, q^* + \delta a) = \mathcal{V}^{\mathrm{gap}}_{\pi,\tau}(\ell, y_L, y_R, q^*) + aL_{\pi,\tau} - \delta a D_{\pi,\tau} \le \mathcal{V}^{\mathrm{gap}}_{\pi,\tau}(\ell, y_L, y_R, q^*).$$

Taking the supremum over (π, τ) and using the fair-charge representation gives

$$\mathcal{V}^{\mathrm{gap}}(\ell, y_L + a, y_R, q^* + \delta a) \le \mathcal{V}^{\mathrm{gap}}(\ell, y_L, y_R, q^*) = 0.$$

Lemma 6 implies that

$$I^{\text{gap}}(\ell, y_L + a, y_R) \leq q^* + \delta a.$$

At the original charge q^*, equation (8) gives

$$\mathcal{V}^{\text{gap}}(\ell, y_L + a, y_R, q^*) \geq \mathcal{V}^{\text{gap}}(\ell, y_L, y_R, q^*) = 0.$$

So, by Lemma 6,

$$I^{\text{gap}}(\ell, y_L + a, y_R) \geq q^*.$$

We now establish strictness. By Corollary 2, there exists an admissible pair (π^*, τ^*) satisfying

$$\mathcal{V}^{\text{gap}}_{\pi^*,\tau^*}(\ell, y_L, y_R, q^*) = 0. \tag{10}$$

By the strict lower bound established above, the attaining pair must operate a safe descendant with positive probability. Otherwise every activation before retirement would yield the payoff $-c$, and its discounted-average payoff would equal $-c$, contradicting $I^{\text{gap}}(\ell, y_L, y_R) > -c$.

Under Assumption 5, every product developed inside the root gap is strictly interior. Thus, $W_n^{\pi^*} \in (0, 1)$ almost surely, whenever $\chi_n^{\pi^*} = 1$. It follows that

$$L_{\pi^*,\tau^*} > 0 \qquad \text{and} \qquad R_{\pi^*,\tau^*} > 0.$$

Consequently, equations (8) and (10) imply

$$\mathcal{V}^{\text{gap}}_{\pi^*,\tau^*}(\ell, y_L + a, y_R, q^*) = aL_{\pi^*,\tau^*} > 0.$$

By Lemma 6, we then have

$$I^{\text{gap}}(\ell, y_L + a, y_R) > q^*.$$

The proof for the right endpoint is identical, using (9). We have therefore proved

$$0 < I^{\text{gap}}(\ell, y_L + a, y_R) - I^{\text{gap}}(\ell, y_L, y_R) \leq \delta a$$

and

$$0 < I^{\text{gap}}(\ell, y_L, y_R + a) - I^{\text{gap}}(\ell, y_L, y_R) \leq \delta a.$$

This also establishes strict monotonicity in average endpoint profitability.

Step 2: convexity and monotonicity in technological distance.

Let $x \equiv \sqrt{\ell}$. For a normalized policy-stopping pair (π, τ), the expected discounted reward is

$$V^{\text{gap}}_{\pi,\tau}(x^2, \bar{y}, s, 0) = C_{\pi,\tau}(\bar{y}, 0) + sH_{\pi,\tau} + xB_{\pi,\tau},$$

by Lemma 9. Since $D_{\pi,\tau}$ is independent of the root-gap parameters, we can define

$$\widehat{I}(x; \bar{y}, s) \equiv \sup_{\pi,\tau} \frac{C_{\pi,\tau}(\bar{y}, 0) + sH_{\pi,\tau} + xB_{\pi,\tau}}{D_{\pi,\tau}}, \qquad x \in \mathbb{R}.$$

Notice that, for $x > 0$,

$$\widehat{I}(x; \bar{y}, s) = I^{\text{gap}}\left(x^2, \bar{y} - \frac{s}{2}, \bar{y} + \frac{s}{2}\right).$$

As the supremum of affine functions of x, the map

$$x \longmapsto \widehat{I}(x; \bar{y}, s)$$

is convex.

We next establish symmetry around zero. Fix a normalized pair (π, τ). Construct its noise-mirror pair (π^-, τ^-) as follows. After a normalized history

$$h_n = \{(W_i, \zeta_i)\}_{i \le n},$$

the mirrored pair behaves as (π, τ) would behave after

$$h_n^- = \{(W_i, -\zeta_i)\}_{i \le n}.$$

Because the Brownian-bridge innovations are symmetric, the mirrored process has the same distribution as

$$(W_n, -\zeta_n)_{n \ge 1}$$

under the original pair. The safe indicators, normalized locations, and stopping decisions have the same distribution. Therefore, recalling that $B_{\pi,\tau}$ and B_{π^-,τ^-} are the only coefficients that depend on $(\zeta_n)_{n \ge 1}$, we get

$$C_{\pi^-,\tau^-}(\bar{y}, 0) = C_{\pi,\tau}(\bar{y}, 0), \quad H_{\pi^-,\tau^-} = H_{\pi,\tau}, \quad B_{\pi^-,\tau^-} = -B_{\pi,\tau}, \quad \text{and} \quad D_{\pi^-,\tau^-} = D_{\pi,\tau}.$$

Thus, for every x and every policy-stopping pair, we have constructed another policy-stopping pair that achieves the same expected discounted average reward when x is replaced by $-x$. It follows that

$$\widehat{I}(x; \bar{y}, s) = \widehat{I}(-x; \bar{y}, s).$$

A finite convex function that is even is weakly increasing on $[0, \infty)$. Hence

$$x \longmapsto I^{\text{gap}}\left(x^2, \bar{y} - \frac{s}{2}, \bar{y} + \frac{s}{2}\right)$$

is weakly increasing. Finiteness follows from Corollary 3, and a finite convex function is continuous on the interior of its domain.

It remains to prove divergence. Let Y_ℓ denote the first prototype generated from the root gap when the gap length is ℓ, and consider the policy that explores the root gap once, operates the newly generated safe arm for one period if $Y_\ell > 0$, and retires otherwise. As discussed earlier, its discounted-average payoff is

$$\frac{-c + \delta \mathbb{E}[Y_\ell^+]}{1 + \delta \mathbb{P}(Y_\ell > 0)}.$$

Let

$$M \equiv \max\{|y_L|, |y_R|\}.$$

Using

$$(a + b)^+ \ge a^+ - |b|,$$

we obtain

$$\mathbb{E}[Y_\ell^+] \ge \sigma \sqrt{\ell}\, \mathbb{E}\left[\sqrt{U(1-U)}\right] \mathbb{E}[\varepsilon^+] - M.$$

Since $F((0,1)) = 1$, we have

$$\mathbb{E}\left[\sqrt{U(1-U)}\right] > 0.$$

Therefore,

$$\mathbb{E}[Y_\ell^+] \longrightarrow +\infty \qquad \text{as } \ell \to \infty.$$

The denominator of the displayed policy value is bounded above by $1+\delta$. Hence that feasible policy has value tending to $+\infty$, and so

$$I^{\text{gap}}(\ell, y_L, y_R) \longrightarrow +\infty.$$

Step 3: steepness.

Fix $x > 0$, $\bar{y}$, and q. By Lemma 9,

$$s \longmapsto V^{\text{gap}}(x^2, \bar{y}, s, q)$$

is the supremum of affine functions of s and is therefore convex. The same argument applied to discounted-average payoffs shows that

$$s \longmapsto I^{\text{gap}}\left(x^2, \bar{y} - \frac{s}{2}, \bar{y} + \frac{s}{2}\right)$$

is convex.

We next establish spatial symmetry. Fix a normalized pair (π, τ). After a normalized history

$$h_n = \{(W_i, \zeta_i)\}_{i \le n},$$

suppose that π selects a descendant gap with normalized endpoints (W_A, W_B), where $W_A < W_B$. Define the spatial-mirror pair (π^m, τ^m) so that, after the reflected history

$$h_n^m = \{(1 - W_i, \zeta_i)\}_{i \le n},$$

it selects the reflected gap

$$(1 - W_B, 1 - W_A)$$

and makes the corresponding reflected stopping decision.

Because

$$U \overset{d}{=} 1 - U,$$

the normalized location generated under the mirrored pair has the same law as $1 - W$ under the original pair. The Brownian-bridge innovation has the same distribution under spatial reflection. Consequently, recalling that the only coefficients depending on $W_i's$ are $H_{\pi,\tau}$ and H_{π^m,τ^m}, it holds

$$C_{\pi^m,\tau^m}(\bar{y}, q) = C_{\pi,\tau}(\bar{y}, q), \quad H_{\pi^m,\tau^m} = -H_{\pi,\tau}, \quad B_{\pi^m,\tau^m} = B_{\pi,\tau}, \quad \text{and} \quad D_{\pi^m,\tau^m} = D_{\pi,\tau}.$$

It follows that

$$I^{\text{gap}}\left(x^2, \bar{y} - \frac{s}{2}, \bar{y} + \frac{s}{2}\right) = I^{\text{gap}}\left(x^2, \bar{y} + \frac{s}{2}, \bar{y} - \frac{s}{2}\right).$$

Thus the index is even in s. Since it is also finite and convex, it is continuous and weakly increasing in $|s|$.

Step 4: the bad-endpoint limit.

Fix $\ell > 0$ and $y \in \mathbb{R}$. The lower bound is immediate: activating the root gap once and retiring yields $I^{\text{gap}}(\ell, y, b) \ge -c$ for every $b \in \mathbb{R}$.

We prove the reverse bound. Fix a charge $q > -c$. It is sufficient, by the fair-charge representation, to show that

$$\mathcal{V}^{\text{gap}}(\ell, y, b, q) < 0$$

for all sufficiently negative b. Let $\mathbb{B}$ denote a standard Brownian bridge. The payoff at normalized location $u \in [0,1]$ is

$$Y_b(u) = (1-u)y + ub + \sigma\sqrt{\ell}\,\mathbb{B}(u).$$

Fix $b_0 \in \mathbb{R}$ and restrict attention to $b \leq b_0$. Then

$$Y_b(u) \leq Y_{b_0}(u) \qquad \text{for every } u \in [0,1].$$

Define

$$\overline{Y}_{b_0} \equiv \sup_{u\in[0,1]} Y_{b_0}(u).$$

The random variable $\overline{Y}_{b_0}$ has a finite first moment.

Fix a finite horizon T. Consider the complete finite tree of all products that can be generated within the first T gap activations, allowing every possible sequence of choices among descendant gaps. Let $\mathcal{N}_T$ denote the finite collection of safe nodes in this tree, and let W_j be the normalized coordinate of node j. Since all protocol draws are strictly interior, $W_j \in (0,1)$ almost surely, for every $j \in \mathcal{N}_T$. Define

$$M_T(b) \equiv \max_{j\in\mathcal{N}_T} Y_b(W_j).$$

Because $\mathcal{N}_T$ is finite and every $W_j > 0$ almost surely,

$$M_T(b) \longrightarrow -\infty \qquad \text{almost surely as } b \to -\infty.$$

Moreover,

$$M_T(b) \leq \overline{Y}_{b_0} \qquad \text{for every } b \leq b_0.$$

Dominated convergence therefore gives

$$\mathbb{E}\left[(M_T(b)-q)^+\right] \longrightarrow 0 \qquad \text{as } b \to -\infty.$$

For any admissible pair (π,τ), the initial activation produces the excess payoff $-(c+q) < 0$. All later exploratory activations also produce negative excess payoff. Therefore, retaining only the possible positive excess payoff from safe descendants,

$$\mathcal{V}^{\text{gap}}_{\pi,\tau}(\ell,y,b,q) \leq -(c+q) + \frac{1}{1-\delta}\mathbb{E}\left[(M_T(b)-q)^+\right] + \frac{\delta^T}{1-\delta}\mathbb{E}\left[(\overline{Y}_{b_0}-q)^+\right].$$

The bound is uniform over admissible pairs. Choose T sufficiently large that

$$\frac{\delta^T}{1-\delta}\mathbb{E}\left[(\overline{Y}_{b_0}-q)^+\right] < \frac{c+q}{3}.$$

Then choose b sufficiently negative that

$$\frac{1}{1-\delta}\mathbb{E}\left[(M_T(b)-q)^+\right] < \frac{c+q}{3}.$$

For such b,

$$\mathcal{V}^{\text{gap}}(\ell,y,b,q) \leq -\frac{c+q}{3} < 0.$$

Hence

$$I^{\text{gap}}(\ell,y,b) < q$$

for all sufficiently negative b. Since this holds for every $q > -c$ and the index is always at least $-c$,

$$\lim_{b \to -\infty} I^{\mathrm{gap}}(\ell, y, b) = -c.$$

The limit with the left endpoint tending to $-\infty$ follows from the the fact the the index is even in steepness, as proved in Step 3.

Step 5: the small-gap bound.

Fix (y_L, y_R) and write $m \equiv \max\{y_L, y_R\}$. Let $\mathbb{B}$ be the standard Brownian bridge associated with the root gap, and define

$$K_\ell \equiv \sigma\sqrt{\ell}\,\|\mathbb{B}\|_\infty, \qquad \|\mathbb{B}\|_\infty \equiv \sup_{u \in [0,1]} |\mathbb{B}(u)|.$$

Every safe descendant generated inside the root gap has payoff bounded above by

$$m + K_\ell.$$

Fix a charge $q > -c$. For any admissible pair, the first activation yields the excess payoff $-(c+q)$. Subsequent exploratory activations also yield negative excess payoff. Thus,

$$\mathcal{V}^{\mathrm{gap}}_{\pi,\tau}(\ell, y_L, y_R, q) \le -(c+q) + \frac{\delta}{1-\delta}\mathbb{E}\left[(m + K_\ell - q)^+\right]. \tag{11}$$

Taking the supremum over (π, τ) preserves the same bound. Since

$$K_\ell \longrightarrow 0 \quad \text{in } L^1 \qquad \text{as } \ell \downarrow 0,$$

the right-hand side of (11) converges to

$$-(c+q) + \frac{\delta}{1-\delta}(m-q)^+.$$

Define

$$q_0 \equiv \max\left\{-c,\, \delta m - (1-\delta)c\right\}.$$

We claim that the preceding limit is strictly negative whenever $q > q_0$.

If $q \ge m$, then

$$(m-q)^+ = 0,$$

so the limit equals

$$-(c+q) < 0.$$

If $q < m$, then

$$-(c+q) + \frac{\delta}{1-\delta}(m-q) = \frac{\delta m - (1-\delta)c - q}{1-\delta} < 0,$$

because

$$q > \delta m - (1-\delta)c.$$

Therefore, for every $q > q_0$, $V^{\mathrm{gap}}(\ell, y_L, y_R, q) < 0$ for all sufficiently small ℓ. The fair-charge representation implies $I^{\mathrm{gap}}(\ell, y_L, y_R) < q$ for all sufficiently small ℓ. Letting $q \downarrow q_0$ gives

$$\limsup_{\ell \downarrow 0} I^{\mathrm{gap}}(\ell, y_L, y_R) \le q_0.$$

The sequential version follows from the same argument. If

$$y_{L,k} \to y_L, \qquad y_{R,k} \to y_R,$$

then

$$m_k \equiv \max\{y_{L,k}, y_{R,k}\} \longrightarrow m.$$

Applying (11) with (ℓ_k, m_k) and taking limits yields the stated bound.

□

B.5 Properties of the Frontier Index

We now collect the properties of the frontier index used in the proofs of the initialization and post-discovery results. Recall that

$$I^{\text{front}}(y) \equiv I\big(\mathsf{F}(y)\big).$$

Lemma 11 (Properties of the frontier index)**.** *For every $y \in \mathbb{R}$ and every $a > 0$,*

$$0 < I^{\text{front}}(y + a) - I^{\text{front}}(y) \leq \delta a.$$

Therefore, the map

$$y \longmapsto I^{\text{front}}(y)$$

is strictly increasing and δ-Lipschitz continuous. The map is also convex. Moreover,

$$I^{\text{front}}(y) > -c \qquad \text{for every } y \in \mathbb{R},$$

and

$$\lim_{y \to -\infty} I^{\text{front}}(y) = -c, \qquad \lim_{y \to +\infty} I^{\text{front}}(y) = +\infty.$$

Proof. We proceed in three steps.

Step 1: continuity, convexity, and monotonicity.

Fix an admissible pair $\pi \in \Pi(\mathsf{F}(0))$, $\tau \in \mathcal{T}(\mathsf{F}(0), \pi)$, and define

$$D_{\pi,\tau} \equiv \mathbb{E}_0^{\pi}\left[\sum_{n=0}^{\tau-1} \delta^n\right], \qquad S_{\pi,\tau} \equiv \mathbb{E}_0^{\pi}\left[\sum_{n=0}^{\tau-1} \delta^n \chi_n^{\pi}\right],$$

where $\chi_n^{\pi} \equiv \mathbf{1}\{Z_n^{\pi} \in \mathcal{Z}_{\text{s}}\}$. Let

$$A_{\pi,\tau}^{\text{front}}(0) \equiv \frac{\mathbb{E}_0^{\pi}\left[\sum_{n=0}^{\tau-1} \delta^n R(Z_n^{\pi})\right]}{D_{\pi,\tau}}.$$

Adding y to every observed profitability gives a bijection between admissible policy-stopping pairs starting from $\mathsf{F}(0)$ and those starting from $\mathsf{F}(y)$. Under this correspondence, every safe payoff increases by y, while every exploratory activation continues to yield $-c$. Therefore,

$$I^{\text{front}}(y) = \sup_{\substack{\pi \in \Pi(\mathsf{F}(0)) \\ \tau \in \mathcal{T}(\mathsf{F}(0), \pi)}} \left\{ A_{\pi,\tau}^{\text{front}}(0) + y \frac{S_{\pi,\tau}}{D_{\pi,\tau}} \right\}. \tag{A}$$

Since the initial arm is exploratory, $\chi_0^\pi = 0$, and hence

$$0 \le \frac{S_{\pi,\tau}}{D_{\pi,\tau}} \le \delta.$$

Equation (A) expresses I^{front} as the supremum of affine functions of y whose slopes belong to $[0, \delta]$. It follows that I^{front} is convex and, for every $a > 0$,

$$0 \le I^{\text{front}}(y + a) - I^{\text{front}}(y) \le \delta a.$$

We next establish that $I^{\text{front}}(y) > -c$. Fix $K > -c$. The first frontier activation generates

$$Y' = y + \mu\Delta + \sigma\sqrt{\Delta}\,\varepsilon, \qquad \Delta \sim G, \quad \varepsilon \sim N(0,1),$$

independently. Since $\Delta > 0$ almost surely and $\sigma > 0$, $\mathbb{P}(Y' > K) > 0$. Consider the policy that activates the frontier once and then operates the newly generated safe arm for one period if $Y' > K$, retiring immediately otherwise. Its discounted-average payoff exceeds $-c$ by

$$\frac{\delta\mathbb{E}\left[(Y' + c)\mathbf{1}_{\{Y'>K\}}\right]}{1 + \delta\mathbb{P}(Y' > K)} > 0.$$

Hence, $I^{\text{front}}(y) > -c$.

Finally, Corollary 2 gives an admissible pair (π^*, τ^*) attaining $I^{\text{front}}(y)$. It must satisfy $S_{\pi^*,\tau^*} > 0$; otherwise every activation before retirement would yield $-c$, contradicting $I^{\text{front}}(y) > -c$. Applying the same translated pair at $y + a$ therefore gives

$$I^{\text{front}}(y + a) \ge I^{\text{front}}(y) + a\frac{S_{\pi^*,\tau^*}}{D_{\pi^*,\tau^*}} > I^{\text{front}}(y).$$

Consequently,

$$0 < I^{\text{front}}(y + a) - I^{\text{front}}(y) \le \delta a.$$

Thus I^{front} is strictly increasing and δ-Lipschitz, and therefore continuous.

Step 2: the limit as $y \to +\infty$.

Consider the feasible policy that activates the frontier once, operates the newly generated safe arm for one period regardless of its realization, and then retires the descendant process. This policy has deterministic stopping time $\tau = 2$ and discounted-average payoff

$$\frac{-c + \delta\mathbb{E}[Y']}{1 + \delta}.$$

Since

$$\mathbb{E}[Y'] = y + \mu\mathbb{E}_G[\Delta],$$

we obtain

$$I^{\text{front}}(y) \ge \frac{-c + \delta y + \delta\mu\mathbb{E}_G[\Delta]}{1 + \delta}.$$

The right-hand side converges to $+\infty$ as $y \to +\infty$. Therefore,

$$\lim_{y\to+\infty} I^{\text{front}}(y) = +\infty.$$

Step 3: the limit as $y \to -\infty$.

The lower bound $I^{\text{front}}(y) \geq -c$ follows from activating the frontier once and retiring. For the reverse bound, fix $\pi \in \Pi(\mathsf{F}(y)),\ \tau \in \mathcal{T}(\mathsf{F}(y), \pi)$. At local date n, let $\overline{Y}_n^{\pi}$ denote the highest payoff among the safe descendants available immediately before the nth activation. Set $\overline{Y}_0^{\pi} = -\infty$. Then

$$R(Z_n^{\pi}) \leq -c + \left(\overline{Y}_n^{\pi} + c\right)^+.$$

We use the same policy-independent frontier upper-bound as in the proof of Lemma 7. Let $(\Delta_j)_{j\geq 1}$ be i.i.d. with law G, and

$$L_n \equiv \sum_{j=1}^{n} \Delta_j, \qquad B_n \equiv |\mu| L_n + \sigma \sup_{0\leq u\leq L_n} |W_u^f|,$$

where W^f is a standard Brownian motion independent of the frontier-step sequence. The argument in the proof of Lemma 7 implies, uniformly over admissible descendant policies,

$$\overline{Y}_n^{\pi} \leq y + B_n,$$

and there exists $K < \infty$, depending only on (μ, σ, G), such that

$$\mathbb{E}[B_n] \leq K(1+n) \qquad \text{for every } n \geq 0.$$

Define

$$N_{\pi,\tau}(y) \equiv \mathbb{E}_y^{\pi}\left[\sum_{n=0}^{\tau-1} \delta^n R(Z_n^{\pi})\right], \qquad D_{\pi,\tau} \equiv \mathbb{E}_y^{\pi}\left[\sum_{n=0}^{\tau-1} \delta^n\right].$$

Using the preceding bound and $D_{\pi,\tau} \geq 1$,

$$\frac{N_{\pi,\tau}(y)}{D_{\pi,\tau}} \leq -c + \sum_{n=0}^{\infty} \delta^n \mathbb{E}\left[(y + c + B_n)^+\right].$$

The bound is uniform over (π, τ), so

$$I^{\text{front}}(y) \leq -c + \sum_{n=0}^{\infty} \delta^n \mathbb{E}\left[(y + c + B_n)^+\right]. \tag{12}$$

Fix $y^\circ \in \mathbb{R}$. For every $y \leq y^\circ$,

$$0 \leq (y + c + B_n)^+ \leq |y^\circ + c| + B_n.$$

Moreover,

$$\sum_{n=0}^{\infty} \delta^n \mathbb{E}\left[|y^\circ + c| + B_n\right] < \infty,$$

because $\mathbb{E}[B_n] \leq K(1+n)$. Since, for every fixed n,

$$(y + c + B_n)^+ \longrightarrow 0 \qquad \text{a.s. as } y \to -\infty,$$

dominated convergence gives

$$\sum_{n=0}^{\infty} \delta^n \mathbb{E}\left[(y + c + B_n)^+\right] \longrightarrow 0.$$

Combining this with (12) and $I^{\text{front}}(y) \geq -c$, yields

$$\lim_{y\to-\infty} I^{\text{front}}(y) = -c.$$

□

B.6 Proof of Proposition 1

Proof. After the frontier experiment, the highest index among the pre-existing opportunities remains ρ. The newly generated safe product has index y', the new frontier has index $I^{\text{front}}(y')$, and the newly created recombination project has index $I^{\text{gap}}(\Delta, y, y')$. By Theorem 1, the startup chooses an opportunity attaining the maximum of these four values.

We prove the four parts separately.

Part (i): market entry after a valuable discovery.

Fix (Δ, y, ρ). For a frontier realization $v \in \mathbb{R}$, define the highest index among the alternatives to immediate commercialization by

$$C(v) \equiv \max\left\{\rho,\, I^{\text{front}}(v),\, I^{\text{gap}}(\Delta, y, v)\right\}.$$

By Lemmas 10 and 11, C is finite, continuous, and weakly increasing. Moreover, for every $a > 0$,

$$0 \le C(v+a) - C(v) \le \delta a.$$

Indeed, ρ is constant, while both exploratory indices increase by at most δa when their relevant payoff coordinate increases by a. Define

$$D^{\text{entry}}(v) \equiv v - C(v).$$

For every $a > 0$,

$$D^{\text{entry}}(v+a) - D^{\text{entry}}(v) = a - \big(C(v+a) - C(v)\big) \ge (1-\delta)a > 0.$$

Thus, D^{entry} is continuous and strictly increasing. Since $C(v) \ge \rho$,

$$D^{\text{entry}}(v) \le v - \rho \longrightarrow -\infty \qquad \text{as } v \to -\infty.$$

For $v \ge 0$, the δ-Lipschitz bounds established by Lemmas 10 and 11 imply

$$I^{\text{front}}(v) \le I^{\text{front}}(0) + \delta v \qquad \text{and} \qquad I^{\text{gap}}(\Delta, y, v) \le I^{\text{gap}}(\Delta, y, 0) + \delta v.$$

Consequently,

$$C(v) \le C(0) + \delta v,$$

and therefore

$$D^{\text{entry}}(v) \ge (1-\delta)v - C(0) \longrightarrow +\infty \qquad \text{as } v \to +\infty.$$

There is therefore a unique finite value

$$\overline{y}^{\,F}(\Delta, y, \rho)$$

satisfying

$$\overline{y}^{\,F}(\Delta, y, \rho) = C\big(\overline{y}^{\,F}(\Delta, y, \rho)\big).$$

Equivalently,

$$\overline{y}^{\,F}(\Delta, y, \rho) = \max\left\{\rho,\, I^{\text{front}}\big(\overline{y}^{\,F}(\Delta, y, \rho)\big),\, I^{\text{gap}}\big(\Delta, y, \overline{y}^{\,F}(\Delta, y, \rho)\big)\right\}.$$

Since D^{entry} is strictly increasing,

$$v \ge \overline{y}^{\,F}(\Delta, y, \rho) \quad \Longleftrightarrow \quad v \ge C(v).$$

Thus the newly developed safe product is weakly optimal if and only if

$$y' \geq \overline{y}^{\,F}(\Delta, y, \rho).$$

If the inequality is strict, then

$$y' > \max\left\{\rho,\, I^{\text{front}}(y'),\, I^{\text{gap}}(\Delta, y, y')\right\},$$

so immediate market entry is uniquely optimal.

Part (ii): return to pre-existing projects.

Continue to hold (Δ, y, ρ) fixed, and define the highest index among the opportunities generated by the frontier experiment by

$$B(v) \equiv \max\left\{v,\, I^{\text{front}}(v),\, I^{\text{gap}}(\Delta, y, v)\right\}.$$

The three functions entering this maximum are finite, continuous, and strictly increasing. Hence B is itself finite, continuous, and strictly increasing. By Lemmas 10 and 11 ,

$$\lim_{v\to-\infty} I^{\text{front}}(v) = -c \qquad \text{and} \qquad \lim_{v\to-\infty} I^{\text{gap}}(\Delta, y, v) = -c.$$

Since $v \to -\infty$, it follows that

$$\lim_{v\to-\infty} B(v) = -c.$$

Moreover, $B(v) \geq v$, and hence

$$\lim_{v\to+\infty} B(v) = +\infty.$$

Under the maintained condition $\rho > -c$, there is therefore a unique finite threshold

$$\underline{y}^{\,F}(\Delta, y, \rho)$$

satisfying

$$B\big(\underline{y}^{\,F}(\Delta, y, \rho)\big) = \rho.$$

That is,

$$\rho = \max\left\{\underline{y}^{\,F}(\Delta, y, \rho),\, I^{\text{front}}\big(\underline{y}^{\,F}(\Delta, y, \rho)\big),\, I^{\text{gap}}\big(\Delta, y, \underline{y}^{\,F}(\Delta, y, \rho)\big)\right\}.$$

Since B is strictly increasing,

$$v \leq \underline{y}^{\,F}(\Delta, y, \rho) \quad \Longleftrightarrow \quad B(v) \leq \rho.$$

Therefore, an opportunity from the pre-existing portfolio is weakly optimal if and only if

$$y' \leq \underline{y}^{\,F}(\Delta, y, \rho).$$

If the inequality is strict, then

$$\max\left\{y',\, I^{\text{front}}(y'),\, I^{\text{gap}}(\Delta, y, y')\right\} < \rho,$$

so every opportunity generated by the new frontier branch has index strictly below the highest pre-existing index.

It remains to compare the two profitability thresholds. By the definition of the upper threshold,

$$\overline{y}^{\,F}(\Delta, y, \rho) = C\big(\overline{y}^{\,F}(\Delta, y, \rho)\big) \geq \rho.$$

Hence

$$B\big(\overline{y}^{\,F}(\Delta, y, \rho)\big) \geq \overline{y}^{\,F}(\Delta, y, \rho) \geq \rho.$$

Since

$$B\big(\underline{y}^{\,F}(\Delta, y, \rho)\big) = \rho$$

and B is strictly increasing, we obtain

$$\underline{y}^{\,F}(\Delta, y, \rho) \leq \overline{y}^{\,F}(\Delta, y, \rho).$$

Part (iii): frontier continuation.

Suppose $I^{\text{front}}(y) > \rho$. Because the safe product at the old frontier belongs to the pre-existing portfolio, $\rho \geq y$. Moreover, $I^{\text{front}}(y) > -c$, by Lemma 11. Define

$$b(y) \equiv \max\left\{-c,\, \delta y - (1-\delta)c\right\}.$$

We have

$$b(y) \leq \max\{y, -c\}.$$

Indeed, if $y > -c$, then

$$\delta y - (1-\delta)c = y - (1-\delta)(y+c) < y;$$

if $y \leq -c$, then

$$\delta y - (1-\delta)c \leq -c.$$

Consequently,

$$I^{\text{front}}(y) > \max\{\rho, y, b(y)\}.$$

Let

$$\gamma \equiv I^{\text{front}}(y) - \max\{\rho, y, b(y)\} > 0.$$

By continuity of the frontier index, there exists $\eta_0 > 0$ such that

$$|v - y| < \eta_0 \quad \Longrightarrow \quad I^{\text{front}}(v) > I^{\text{front}}(y) - \frac{\gamma}{4}.$$

Reducing η_0 if necessary, assume also that

$$\eta_0 < \frac{\gamma}{4}.$$

Then, whenever $|v - y| < \eta_0$,

$$I^{\text{front}}(v) > \rho \qquad \text{and} \qquad I^{\text{front}}(v) > v.$$

The sequential small-gap bound in Lemma 10 implies that there exist $\eta_1 > 0$ and $\overline{\Delta} > 0$ such that

$$I^{\text{gap}}(\Delta, y, v) < b(y) + \frac{\gamma}{4}$$

whenever

$$0 < \Delta < \overline{\Delta} \qquad \text{and} \qquad |v - y| < \eta_1.$$

To see this, otherwise one could construct sequences $\Delta_k \downarrow 0,\ v_k \to y$, satisfying

$$I^{\text{gap}}(\Delta_k, y, v_k) \geq b(y) + \frac{\gamma}{4},$$

contradicting the sequential version of the small-gap bound. Set

$$\eta \equiv \min\{\eta_0, \eta_1\}.$$

For $0 < \Delta < \overline{\Delta},\ |y' - y| < \eta$, we have

$$I^{\text{front}}(y') > I^{\text{front}}(y) - \frac{\gamma}{4}, \qquad \text{and} \qquad I^{\text{gap}}(\Delta, y, y') < b(y) + \frac{\gamma}{4}.$$

Since

$$I^{\text{front}}(y) \geq b(y) + \gamma,$$

it follows that

$$I^{\text{front}}(y') > I^{\text{gap}}(\Delta, y, y').$$

Together with

$$I^{\text{front}}(y') > \rho \qquad \text{and} \qquad I^{\text{front}}(y') > y',$$

this shows that the new frontier arm is uniquely optimal.

Finally, Assumption 6 gives $G((0, \overline{\Delta})) > 0$. Conditional on every realized step $d \in (0, \overline{\Delta})$,

$$y' - y = \mu d + \sigma\sqrt{d}\,\varepsilon$$

has a nondegenerate normal distribution. Therefore,

$$\mathbb{P}\left(|y' - y| < \eta \middle| \Delta = d\right) > 0.$$

It follows that

$$\mathbb{P}\left(0 < \Delta < \overline{\Delta},\, |y' - y| < \eta\right) = \int_{(0,\overline{\Delta})} \mathbb{P}\left(|\mu d + \sigma\sqrt{d}\,\varepsilon| < \eta \middle| \Delta = d\right) G(dd) > 0.$$

Part (iv): the technological-distance cutoff.

Fix (y, y', ρ), and define the highest non-gap index after the discovery by

$$M \equiv \max\left\{\rho,\, y',\, I^{\text{front}}(y')\right\}.$$

This number is finite. Moreover, $M > -c$ because $I^{\text{front}}(y') > -c$. Let

$$m \equiv \max\{y, y'\}.$$

Since $\rho \geq y$ and $M \geq y'$, $M \geq m$. The small-gap bound in Lemma 10 gives

$$\limsup_{\ell \downarrow 0} I^{\text{gap}}(\ell, y, y') \leq \max\left\{-c,\, \delta m - (1-\delta)c\right\}.$$

The right-hand side is strictly below M. Indeed, if $m > -c$, then

$$\max\left\{-c,\, \delta m - (1-\delta)c\right\} < m \leq M.$$

If $m \leq -c$, then the right-hand side equals $-c$, while

$$-c < I^{\text{front}}(y') \leq M.$$

Thus

$$\limsup_{\ell \downarrow 0} I^{\text{gap}}(\ell, y, y') < M.$$

It follows that

$$I^{\text{gap}}(\ell, y, y') < M$$

for all sufficiently small $\ell > 0$.

On the other hand, the large-gap result in Lemma 10 gives

$$I^{\text{gap}}(\ell, y, y') \longrightarrow +\infty \qquad \text{as } \ell \to \infty.$$

Therefore, the set

$$\left\{\ell > 0 : I^{\text{gap}}(\ell, y, y') \geq M\right\}$$

is nonempty and is bounded away from zero. Hence

$$0 < \Delta^F(y, y', \rho) < \infty.$$

We now establish the threshold structure. Define

$$h(x) \equiv I^{\text{gap}}(x^2, y, y'), \qquad x > 0,$$

and set

$$x^F \equiv \sqrt{\Delta^F(y, y', \rho)}.$$

By Lemma 10, h is finite, continuous, convex, and weakly increasing. For every $x < x^F$, the definition of the cutoff implies $h(x) < M$. Continuity then gives $h(x^F) = M$. Indeed, if $h(x^F) < M$, continuity would imply that $h(x) < M$ on a right-neighborhood of x^F, contradicting the definition of the infimum. If $h(x^F) > M$, continuity would imply that $h(x) \geq M$ for some $x < x^F$, producing the opposite contradiction.

Strict dominance above the cutoff follows from recalling that we have proved above that $h(x) < M$ for x low enough. So, continuity and convexity imply that there exists a left neighborhood of x^F on which h is strictly increasing. Convexity then implies that h is strictly increasing also at the right of x^F. □

B.7 Proof of Proposition 2

Proof. After the recombination experiment, the highest index among the pre-existing opportunities remains ρ. The newly developed safe product has index y_M, while the two descendant recombination projects have indices $J_L(y_M)$ and $J_R(y_M)$. Theorem 1 therefore reduces the post-experiment decision to a comparison of these four values.

We proceed in two steps.

Step 1: thresholds in prototype profitability.

By Lemma 10, the functions

$$J_L, J_R : \mathbb{R} \longrightarrow \mathbb{R}$$

are finite, continuous, and strictly increasing. Moreover, for every $v \in \mathbb{R}$ and every $a > 0$,

$$0 < J_L(v + a) - J_L(v) \leq \delta a \qquad \text{and} \qquad 0 < J_R(v + a) - J_R(v) \leq \delta a.$$

We first characterize market entry. Define

$$C(v) \equiv \max\{\rho, J_L(v), J_R(v)\}.$$

The preceding Lipschitz bounds imply

$$0 \le C(v+a) - C(v) \le \delta a.$$

Hence the function

$$D^{\text{entry}}(v) \equiv v - C(v)$$

is continuous and strictly increasing. Indeed,

$$D^{\text{entry}}(v+a) - D^{\text{entry}}(v) \ge (1-\delta)a > 0.$$

Since $C(v) \ge \rho$,

$$D^{\text{entry}}(v) \le v - \rho \longrightarrow -\infty \qquad \text{as } v \to -\infty.$$

For $v \ge 0$, the endpoint Lipschitz bounds give

$$J_L(v) \le J_L(0) + \delta v \qquad \text{and} \qquad J_R(v) \le J_R(0) + \delta v.$$

Consequently,

$$C(v) \le \max\{\rho, J_L(0), J_R(0)\} + \delta v,$$

and therefore

$$D^{\text{entry}}(v) \longrightarrow +\infty \qquad \text{as } v \to +\infty.$$

There is thus a unique finite threshold

$$\overline{y}^{\,\text{rec}} = \overline{y}^{\,\text{rec}}(\ell, u, y_L, y_R, \rho)$$

satisfying

$$\overline{y}^{\,\text{rec}} = \max\left\{\rho, J_L\left(\overline{y}^{\,\text{rec}}\right), J_R\left(\overline{y}^{\,\text{rec}}\right)\right\}.$$

Since D^{entry} is strictly increasing,

$$v \ge \overline{y}^{\,\text{rec}} \quad \Longleftrightarrow \quad v \ge \max\{\rho, J_L(v), J_R(v)\}.$$

Thus the newly developed product is weakly optimal if and only if $y_M \ge \overline{y}^{\,\text{rec}}$, and it is uniquely optimal whenever the inequality is strict.

We next characterize return to the pre-existing portfolio. Define

$$B(v) \equiv \max\{v, J_L(v), J_R(v)\}.$$

Because all three functions entering the maximum are continuous and strictly increasing, B is continuous and strictly increasing. The bad-endpoint limits in Lemma 10 imply

$$\lim_{v \to -\infty} J_L(v) = \lim_{v \to -\infty} J_R(v) = -c.$$

It follows that

$$\lim_{v \to -\infty} B(v) = -c.$$

Moreover, $B(v) \ge v$, and hence

$$\lim_{v \to +\infty} B(v) = +\infty.$$

Since $\rho > -c$, there is a unique finite threshold

$$\underline{y}^{\,\mathrm{rec}} = \underline{y}^{\,\mathrm{rec}}(\ell, u, y_L, y_R, \rho)$$

satisfying

$$\rho = \max\left\{\underline{y}^{\,\mathrm{rec}}, J_L\big(\underline{y}^{\,\mathrm{rec}}\big), J_R\big(\underline{y}^{\,\mathrm{rec}}\big)\right\}.$$

Strict monotonicity of B gives

$$v \le \underline{y}^{\,\mathrm{rec}} \quad \Longleftrightarrow \quad B(v) \le \rho.$$

Thus an opportunity attaining the pre-existing index ρ is weakly optimal if and only if $y_M \le \underline{y}^{\,\mathrm{rec}}$. If the inequality is strict, then

$$\max\{y_M, J_L(y_M), J_R(y_M)\} < \rho,$$

so every opportunity generated by the recombination experiment has index strictly below the highest pre-existing index.

We now compare the two thresholds. At the upper threshold,

$$\overline{y}^{\,\mathrm{rec}} = C\big(\overline{y}^{\,\mathrm{rec}}\big) \ge \rho.$$

Therefore,

$$B\big(\overline{y}^{\,\mathrm{rec}}\big) \ge \overline{y}^{\,\mathrm{rec}} \ge \rho.$$

Since

$$B\big(\underline{y}^{\,\mathrm{rec}}\big) = \rho$$

and B is strictly increasing,

$$\underline{y}^{\,\mathrm{rec}} \le \overline{y}^{\,\mathrm{rec}}.$$

Finally, suppose that

$$\underline{y}^{\,\mathrm{rec}} < v < \overline{y}^{\,\mathrm{rec}}.$$

Write

$$J(v) \equiv \max\{J_L(v), J_R(v)\}.$$

The first inequality implies

$$\max\{v, J(v)\} > \rho,$$

while the second implies

$$v < \max\{\rho, J(v)\}.$$

These two inequalities jointly imply

$$J(v) > \max\{\rho, v\}.$$

Indeed, if $J(v) \le \rho$, the first inequality would require $v > \rho$, whereas the second would require $v < \rho$, a contradiction. Similarly, if $J(v) \le v$, the second inequality would require $\rho > v$, whereas the first would require $v > \rho$. Consequently,

$$\underline{y}^{\,\mathrm{rec}} < y_M < \overline{y}^{\,\mathrm{rec}}$$

implies

$$\max\{J_L(y_M), J_R(y_M)\} > \max\{\rho, y_M\}.$$

One of the two descendant recombination projects is therefore uniquely preferred to both the pre-existing portfolio and immediate market entry. At $y_M = \underline{y}^{\,\mathrm{rec}}$, an opportunity attaining ρ is weakly optimal. At $y_M = \overline{y}^{\,\mathrm{rec}}$, the newly developed safe product is weakly optimal. The strict continuation region is nonempty precisely when $\underline{y}^{\,\mathrm{rec}} < \overline{y}^{\,\mathrm{rec}}$.

Step 2: the technological-span cutoff.

Fix (u, y_M, y_L, y_R, ρ) and write

$$M \equiv \max\{\rho, y_M\}.$$

Since both endpoint products belong to the pre-existing portfolio,

$$y_L \le \rho, \qquad y_R \le \rho.$$

It follows that

$$\max\{y_L, y_M, y_R\} \le M.$$

Moreover,

$$M \ge \rho > -c.$$

Recall that

$$K(\lambda) = \max\left\{I^{\mathrm{gap}}(u\lambda, y_L, y_M), I^{\mathrm{gap}}((1-u)\lambda, y_M, y_R)\right\}.$$

It is convenient to set

$$H(x) \equiv K(x^2), \qquad x > 0.$$

By Lemma 10, each of the functions

$$x \longmapsto I^{\mathrm{gap}}(ux^2, y_L, y_M) \qquad \text{and} \qquad x \longmapsto I^{\mathrm{gap}}((1-u)x^2, y_M, y_R)$$

is finite, continuous, convex, and weakly increasing. Their maximum H therefore has the same properties. Let

$$b(M) \equiv \max\left\{-c,\, \delta M - (1-\delta)c\right\}.$$

Since $M > -c$,

$$b(M) < M.$$

The small-gap bound in Lemma 10 gives

$$\limsup_{x \downarrow 0} I^{\mathrm{gap}}(ux^2, y_L, y_M) \le b(M) \qquad \text{and} \qquad \limsup_{x \downarrow 0} I^{\mathrm{gap}}((1-u)x^2, y_M, y_R) \le b(M).$$

Hence

$$\limsup_{x \downarrow 0} H(x) \le b(M) < M.$$

Thus, for sufficiently small root lengths,

$$K(\lambda) < M,$$

and neither descendant gap is optimal.

Since $u \in (0,1)$, both

$$ux^2 \quad \text{and} \quad (1-u)x^2$$

diverge as $x \to \infty$. The large-gap property in Lemma 10 therefore implies

$$H(x) \longrightarrow +\infty \qquad \text{as } x \to \infty.$$

It follows that

$$0 < \ell^{\text{rec}}(u, y_M, y_L, y_R, \rho) < \infty.$$

Write

$$x^{\text{rec}} \equiv \sqrt{\ell^{\text{rec}}(u, y_M, y_L, y_R, \rho)}.$$

By the definition of the cutoff,

$$H(x) < M \qquad \text{for every } x < x^{\text{rec}}.$$

Continuity gives

$$H(x^{\text{rec}}) = M.$$

Strict dominance above the cutoff follows from $H(x) < M$ for x low enough. This is so because continuity and convexity then imply that there exists a left neighborhood of x^{rec} on which H is strictly increasing. Convexity then implies that H is strictly increasing also at the right of x^{rec}.

□

Appendix C Career Mobility

Proof of Lemma 1. Fix a reachable informational state S, and write $r \equiv r(S)$ and $y \equiv y(S)$. We first consider a finite continuation horizon. For $T \geq 1$ and $k \in \{1, 2\}$, let $Q_{k,T}(S)$ denote the expected discounted payoff over the next T periods when the worker selects $r + k$ in the current period and behaves optimally thereafter. Temporarily suppress any midpoint opportunity created between the old and new frontiers. Conditional on the opportunities already available at S, let

$$\Phi_T(x; S)$$

denote the optimal expected payoff over the next T periods when a new right-hand frontier branch has been created with frontier payoff x. Set $\Phi_0(\cdot; S) \equiv 0$.

The absolute location of the new branch is immaterial for this restricted continuation problem. By stationary independent increments, a continuation plan for a branch beginning at $r + 1$ can be translated one position to the right and followed from a branch beginning at $r+2$. All opportunities that were available before the frontier experiment remain available in either problem.

We claim that

$$x \longmapsto \Phi_T(x; S)$$

is convex. Represent future uncertainty by a canonical collection of random-walk increments, independent of x. A contingent plan specifies, after every canonical innovation history, whether the worker selects an opportunity that was already available at S or one descending from the new frontier branch. The set of such contingent plans is independent of x.

For any fixed contingent plan, every payoff obtained from a pre-existing opportunity is independent of x, whereas every payoff obtained from the new branch is equal to x plus a random term independent of x. The expected discounted payoff generated by that plan therefore has the form

$$A + xB$$

for constants A and B that depend on the plan and on s, but not on x. Since $\Phi_T(\cdot; S)$ is the supremum of these affine functions, it is convex.

Let

$$Y^{(1)} \equiv y + \varepsilon_1, \qquad Y^{(2)} \equiv y + \varepsilon_1 + \varepsilon_2,$$

where ε_1 and ε_2 are independent Rademacher variables. A one-position frontier move creates no midpoint opportunity. Hence

$$Q_{1,T}(S) = \mathbb{E}\left[Y^{(1)} + \delta\Phi_{T-1}\big(Y^{(1)}; S\big)\right] = y + \delta\mathbb{E}\left[\Phi_{T-1}\big(Y^{(1)}; S\big)\right].$$

After a two-position move, the worker can ignore the newly created midpoint and follow a translated version of any continuation plan available after a one-position move. Therefore,

$$Q_{2,T}(S) \geq \mathbb{E}\left[Y^{(2)} + \delta\Phi_{T-1}\big(Y^{(2)}; S\big)\right] = y + \delta\mathbb{E}\left[\Phi_{T-1}\big(Y^{(2)}; S\big)\right].$$

Moreover,

$$\mathbb{E}\left[Y^{(2)} \mid Y^{(1)}\right] = Y^{(1)}.$$

Conditional Jensen's inequality and the convexity of $\Phi_{T-1}(\cdot; S)$ imply

$$\mathbb{E}\left[\Phi_{T-1}\big(Y^{(2)}; S\big)\right] \geq \mathbb{E}\left[\Phi_{T-1}\big(Y^{(1)}; S\big)\right].$$

Consequently,

$$Q_{2,T}(s) \geq Q_{1,T}(S)$$

for every finite T.

Since Rademacher variables have support $\{-1, 1\}$, it is immediate to see that finite-horizon values converge to their infinite-horizon counterparts and that the discounted truncation error converges to zero uniformly over admissible policies. Letting $T \to \infty$ therefore gives

$$Q_2(S) \geq Q_1(S).$$

Thus, whenever a one-position frontier expansion is optimal, the two-position expansion is also optimal. Resolving indifference in favor of the latter yields an optimal policy under which every frontier expansion moves exactly two positions to the right. □

Proof of Lemma 2. A safe arm S_v yields the deterministic reward v in every period and reproduces itself. Hence $I^{\mathrm{S}}(v) = v$.

We next establish vertical translation for the exploratory career arms. For $a \in \mathbb{Z}$, define

$$T_a\mathsf{G}_v = \mathsf{G}_{v+a}, \qquad T_a\mathsf{F}_v = \mathsf{F}_{v+a}.$$

The transition laws of the career arms imply that activating $T_a z$ generates the translation by a of every descendant generated by z. Thus, translating every arm in a descendant history by a gives a bijection between admissible policy-stopping pairs starting from z and those starting from $T_a z$. Fix an admissible pair (π, τ) starting from z, and let (π^a, τ^a) denote its translated counterpart. Since the career application imposes

$$u(y, \chi) = y, \qquad c = 0,$$

every reward obtained under the translated pair is larger by exactly a. Therefore,

$$\mathbb{E}^{\pi^a}_{T_a z}\left[\sum_{n=0}^{\tau^a - 1} \delta^n R(Z_n^{\pi^a})\right] = \mathbb{E}^{\pi}_{z}\left[\sum_{n=0}^{\tau - 1} \delta^n R(Z_n^{\pi})\right] + a\,\mathbb{E}^{\pi}_{z}\left[\sum_{n=0}^{\tau - 1} \delta^n\right].$$

The discounted duration is unchanged. Hence the discounted-average payoff of the translated pair is the original discounted-average payoff plus a. Taking suprema and using the inverse translation gives

$$I(T_a z) = I(z) + a.$$

It follows that

$$I^{\mathrm{G}}(v) = v + I^{\mathrm{G}}(0), \qquad I^{\mathrm{F}}(v) = v + I^{\mathrm{F}}(0).$$

Define

$$\iota_G \equiv I^{\mathrm{G}}(0), \qquad \iota_F \equiv I^{\mathrm{F}}(0).$$

We now compute ι_G. Starting from G_0, the first activation has expected reward zero and generates

$$\mathsf{S}_1 \quad \text{or} \quad \mathsf{S}_{-1}$$

with equal probability. These are the only possible descendants.

Following the realization -1, operating the resulting safe arm can only lower the discounted-average payoff, so it is optimal to retire the family immediately. Following the realization 1, suppose that the worker operates S_1 for N periods and then retires. The corresponding discounted-average payoff is

$$\frac{\frac{1}{2}\sum_{n=1}^{N} \delta^n}{1 + \frac{1}{2}\sum_{n=1}^{N} \delta^n}.$$

This expression is increasing in N. Therefore, the supremum is obtained by retaining S_1 forever after the positive realization. The resulting numerator and denominator are

$$N_G = \frac{1}{2}\sum_{n=1}^{\infty} \delta^n = \frac{\delta}{2(1-\delta)}$$

and

$$D_G = 1 + \frac{1}{2}\sum_{n=1}^{\infty} \delta^n = 1 + \frac{\delta}{2(1-\delta)}.$$

Consequently,

$$\iota_G = \frac{N_G}{D_G} = \frac{\delta}{2-\delta}.$$

Since $\delta \in (0,1)$,

$$0 < \iota_G < 1.$$

Finiteness of ι_F follows from the discounted-integrability bound and Theorem 1.

To compare the two exploratory indices, consider the following feasible policy in the single-arm descendant process initiated by F_0. Activate the frontier once. If the new frontier match is 2, operate the resulting safe arm S_2 forever. If the new frontier match is 0 or -2, retire the descendant family immediately.

The event in which the new frontier match equals 2 has probability 1/4. The discounted-average payoff generated by this policy is therefore

$$\frac{\frac{1}{4}\,2\sum_{n=1}^{\infty}\delta^n}{1+\frac{1}{4}\sum_{n=1}^{\infty}\delta^n}=\frac{2\delta}{4-3\delta}.$$

It follows that

$$\iota_F \geq \frac{2\delta}{4-3\delta}.$$

Moreover,

$$\frac{2\delta}{4-3\delta}-\frac{\delta}{2-\delta}=\frac{\delta^2}{(4-3\delta)(2-\delta)}>0.$$

Hence

$$\iota_F > \iota_G > 0.$$

The normalized transition laws of G_0 and F_0 contain no primitive other than the fixed Rademacher probabilities. Thus ι_G and ι_F depend only on the discount factor δ. Combining this observation with vertical translation proves

$$I^{\mathrm{S}}(v)=v, \qquad I^{\mathrm{G}}(v)=v+\iota_G, \qquad I^{\mathrm{F}}(v)=v+\iota_F$$

for every reachable $v \in \mathbb{Z}$. □

Proof of Proposition 3. Suppose first that $g(s)=0$. By Corollary 1, the worker compares the best safe occupation with the career frontier. The safe occupation is weakly preferred if and only if

$$m(s) \geq y(s)+\iota_F,$$

or equivalently,

$$d(s)+\iota_F \leq 0.$$

Since $d(s)$ is integer valued, this condition is equivalent to

$$d(s) \leq \overline{d}.$$

Under the stated tie-breaking rule, the worker settles whenever this inequality holds and expands the frontier otherwise.

Now suppose that $g(s)>0$. A maximal gap has index

$$m(s)+\iota_G > m(s),$$

because $\iota_G > 0$. Hence the best safe occupation is strictly dominated, and the worker compares a maximal gap with the frontier. The gap is weakly preferred if and only if

$$m(s)+\iota_G \geq y(s)+\iota_F,$$

or equivalently,

$$d(s)+\iota_F \leq \iota_G.$$

Since $d(s)$ is integer valued, this condition is equivalent to

$$d(s) \leq d^*.$$

The tie-breaking rule selects a gap at equality.

Finally,

$$\iota_F > \iota_G > 0$$

implies

$$\overline{d} < 0 \qquad \text{and} \qquad d^* < 0.$$

Since

$$0 < \iota_G < 1,$$

we also have

$$\lfloor -\iota_F \rfloor \leq \lfloor -\iota_F + \iota_G \rfloor \leq \lfloor -\iota_F \rfloor + 1,$$

which proves

$$\overline{d} \leq d^* \leq \overline{d} + 1.$$

□

Proof of Lemma 3. Fix a reachable informational state S with career frontier r. As in the proof of Lemma 1, we first consider a finite continuation horizon. For $T \geq 0$, let

$$\Phi_T^I(x, p; S)$$

denote the optimal continuation value over the next T periods in the auxiliary problem in which the worker retains all opportunities available at S, has a new career frontier at position p with match x, and any untried position initially left between the old and new frontiers is suppressed. Set $\Phi_0^I \equiv 0$.

The same finite-horizon argument used in the proof of Lemma 1 implies that

$$x \longmapsto \Phi_T^I(x, p; S)$$

is convex. In addition,

$$\Phi_T^I(x, p+1; S) \geq \Phi_T^I(x, p; S). \tag{C.1}$$

Indeed, this follows by backward induction on T. Any continuation plan from frontier position p can be shifted one position to the right from frontier position $p+1$. Pre-existing opportunities remain unchanged, while each new occupation sampled under the shifted plan has a marginal distribution that is a mean-preserving spread of the corresponding distribution under the original plan. Conditional Jensen's inequality and the convexity of the continuation value therefore imply that shifting the plan to the right cannot reduce its expected value. This proves (C.1).

If $Q_{k,T}^I(S)$ denotes the finite-horizon value of first selecting $r+k$, then

$$Q_{1,T}^I(S) = \mathbb{E}\left[\Psi_{r+1}^I + \delta \Phi_{T-1}^I(\Psi_{r+1}^I, r+1; S)\right].$$

After selecting $r+2$, the worker can ignore the additional untried occupation at $r+1$. Hence

$$Q_{2,T}^I(S) \geq \mathbb{E}\left[\Psi_{r+2}^I + \delta \Phi_{T-1}^I(\Psi_{r+1}^I, r+2; S)\right].$$

Using (C.1),

$$Q_{2,T}^I(S) \geq \mathbb{E}\left[\Psi_{r+2}^I + \delta \Phi_{T-1}^I(\Psi_{r+2}^I, r+1; S)\right].$$

The function

$$x \longmapsto x + \delta \Phi^I_{T-1}(x, r+1; S)$$

is convex. Conditional Jensen's inequality therefore gives

$$Q^I_{2,T}(S) \geq \mathbb{E}\left[\Psi^I_{r+2} + \delta \Phi^I_{T-1}(\Psi^I_{r+1}, r+1; S)\right] = Q^I_{1,T}(S).$$

Thus,

$$Q^I_{2,T}(S) \geq Q^I_{1,T}(S)$$

for every finite T. The passage to the infinite horizon is the same as in the proof of Lemma 1. Since

$$n_t \leq r + 2(t+1) \qquad \text{and} \qquad |\Psi^I_n| \leq n$$

almost surely, the discounted truncation error converges to zero uniformly over admissible policies. Letting $T \to \infty$ yields

$$Q^I_2(S) \geq Q^I_1(S).$$

□

Proof of Lemma 4. Fix an untried occupation n and a candidate fair charge q. The arm must first be activated. After its payoff Y^I_n is observed, its only descendant is the safe arm $\mathsf{S}(Y^I_n)$. It is optimal within the single-arm descendant problem to operate this safe arm forever when $Y^I_n > q$ and to retire the family otherwise. Its fair-charge excess value is therefore

$$V^H_n(q) = \mathbb{E}\left[Y^I_n - q + \frac{\delta}{1-\delta}\left(Y^I_n - q\right)^+\right] = -q + \frac{\delta}{1-\delta}\mathbb{E}\left[\left(Y^I_n - q\right)^+\right],$$

where the second equality uses $\mathbb{E}[Y^I_n] = 0$. The function V^H_n is continuous and strictly decreasing,

$$V^H_n(0) > 0,$$

and

$$V^H_n(q) \longrightarrow -\infty \qquad \text{as } q \to +\infty.$$

It therefore has a unique positive root. By the fair-charge representation, that root is γ_n, giving

$$\delta \mathbb{E}\left[\left(Y^I_n - \gamma_n\right)^+\right] = (1-\delta)\gamma_n.$$

For every fixed charge q, the function

$$y \longmapsto (y-q)^+$$

is convex. Since, for every $n \leq k$, Y^I_k is a mean-preserving spread of Y^I_n, it follows

$$V^H_n(q) \leq V^H_k(q) \qquad \text{whenever } n \leq k.$$

Since each excess-value function is strictly decreasing and vanishes at its index,

$$\gamma_n \leq \gamma_k.$$

Finiteness of b_r follows from discounted integrability. To obtain the lower bound, consider the following feasible policy in the descendant problem initiated by F^I_r. Activate the frontier once, ignore the newly created hole and future frontier, and manage only the safe arm generated by the draw at position $r+2$. This reproduces the single-arm policy available from H^I_{r+2}. Hence

$$b_r \geq \gamma_{r+2}.$$

Finally,

$$\gamma_{r+2} > 0$$

because Y^I_{r+2} is nondegenerate. □

Proof of Proposition 4. Lemma 3 eliminates the one-position frontier move. Consider any untried occupation n behind the frontier. Since $n \leq r+1$, Lemma 4 gives

$$I\big(\mathsf{H}_n^I\big) = \gamma_n \leq \gamma_{r+2} \leq b_r.$$

Every untried occupation behind the frontier is therefore weakly dominated by the career-frontier arm.

Among the previously tried occupations, one attaining m has the highest safe index, equal to m. The optimal action is consequently obtained by comparing

$$m \qquad \text{and} \qquad b_r.$$

The frontier is selected when $m < b_r$, while the tie-breaking rule selects a safe occupation when $m \geq b_r$.

Selecting a known occupation leaves the informational state unchanged. Hence, once a safe occupation is optimal, it remains optimal in every subsequent period. Settlement is therefore permanent. □

References

Martino Banchio and Suraj Malladi. Rediscovery. *arXiv preprint arXiv:2504.19761*, 2026.

Jesse Bruhn, Jacob Fabian, Luke Gallagher, Matthew Gudgeon, Adam Isen, and Aaron R. Phipps. Path dependence in the labor market: The long-run effects of early career occupational experience. NBER Working Paper 34463, National Bureau of Economic Research, November 2025.

Steven Callander. Searching and learning by trial and error. *American Economic Review*, 101(6): 2277–2308, 2011.

Daniel Fershtman and Alessandro Pavan. *Searching for arms: Experimentation with endogenous consideration sets*. Foerder Institute for Economic Research, Tel Aviv University, The Eitan . . . , 2026.

Lee Fleming. Recombinant uncertainty in technological search. *Management science*, 47(1):117–132, 2001.

Umberto Garfagnini and Bruno Strulovici. Social experimentation with interdependent and expanding technologies. *The Review of Economic Studies*, 83(4):1579–1613, 2016.

Kevin D Glazebrook. On a sufficient condition for superprocesses due to whittle. *Journal of Applied Probability*, 19(1):99–110, 1982.

Anders Humlum, Jakob R. Munch, and Pernille Plato. Changing tracks: Human capital investment after loss of ability. *American Economic Review*, 115(5):1520–1554, 2025. doi: 10.1257/aer.20231067.

Boyan Jovanovic and Rafael Rob. Long waves and short waves: Growth through intensive and extensive search. *Econometrica: Journal of the Econometric Society*, pages 1391–1409, 1990.

Bryan Kelly, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. Measuring technological innovation over the long run. *American Economic Review: Insights*, 3(3):303–320, 2021.

Leonid Kogan, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. Technological innovation, resource allocation, and growth. *The quarterly journal of economics*, 132(2):665–712, 2017.

Suraj Malladi. Searching in the dark and learning where to look. *Available at SSRN 4084113*, 2022.

Peter Nash. *Optimal allocation of Resources Between Research Projects.* PhD thesis, University of Cambridge, 1973.

Dominika Kinga Randle and Gary Paul Pisano. The evolutionary nature of breakthrough innovation: An empirical investigation of firm search strategies. *Strategy Science*, 6(4):290–304, 2021.

Can Urgun and Leeat Yariv. Contiguous search: Exploration and ambition on uncharted terrain. *Journal of Political Economy*, 133(2):522–567, 2025.

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

Yu Fu Wong. Forward-looking experimentation of correlated alternatives. *Theoretical Economics*, 20(3):883–909, 2025.

Xianyi Wu and Xian Zhou. Open bandit processes with uncountable states and time-backward effects. *Journal of Applied Probability*, 50(2):388–402, 2013.