Search and Rediscovery

Martino Banchio
Bocconi

Suraj Malladi
Northwestern Kellogg

Abstract

We model search in settings where agents know what can be found but not where to find it. A searcher faces a set of choices arranged by an observable attribute. Each period, she either selects a choice and pays a cost to learn about its quality, or she concludes search to take her best discovery to date. She knows that similar choices have similar qualities and uses this to guide her search. We identify robustly optimal search policies with a simple structure. Search is incremental, recall is never invoked, there is a threshold stopping rule, and the policy at each history depends only on a simple index.

1 Introduction

Making an original discovery (e.g., tackling an open problem or developing a breakthrough innovation) is a process of trial and error in the face of stark uncertainty. Researchers, agencies and firms learn from their past successes and dead-ends when deciding which approach to try next. They also infer from their attempts whether worthwhile discoveries even exist and decide when to give up.

On the other hand, agents attempting rediscovery know that worthwhile discoveries exist, even if they do not know where to find them. A student tackling a homework problem faces much of the same uncertainty as the researcher who first solved it. However, the student knows that the problem has a solution and relies only on material covered in class, whereas the researcher had no such guarantees. Analogously, a non-nuclear state faces significant uncertainty in developing a weapon, but other programs—starting with the Manhattan project—revealed that developing such weapons is possible, even if their methods were not disclosed.

We develop a stylized model of rediscovery and characterize the optimal search process. We crystallize the idea that the process of rediscovery seems simpler than that of original discovery. While original discovery involves learning where to look and inferring what can be found, rediscovery only involves the former. To an agent attempting an original discovery, a bad outcome is potentially a sign that there is nothing good to be found and all known approaches are fruitless. But to an agent attempting rediscovery, a bad outcome is only a sign that a very different approach is needed, because a good outcome is known to exist. Therefore, searching to rediscover is a process of elimination to isolate the approaches that lead to promising outcomes.

Understanding rediscovery may be of interest to economists, because it appears to be a ubiquitous and important driver of innovation. Firms that develop novel technologies, such as self-driving cars or new AI algorithms, often keep their methods as trade secrets instead of publicizing them or filing a patent. Competitors who learn only that such inventions are feasible may embark on their own R&D process to at least partially recreate the innovating firm's success. In the terms of this example, we ask: how do these competitors go about searching the space of possible designs? At what point do they conclude search and release their own novel variants of the original innovation?

Similar questions have been long considered in the management and entrepreneurship literature on the behavioral theory of the firm. In a seminal book, (Cyert and March, 1963) describe *problemistic search* by managers which is triggered by events like “failure to achieve the profit goal" or “innovation by a competitor." Subsequent authors have suggested that this sort of problemistic search by firms often happens over *rugged landscapes*, meaning that the mapping from firm's choices to outcomes is complex and unpredictable (Levinthal, 1997; Billinger et al., 2014; Callander, 2011; Callander et al., 2022). Motivated by this literature, we conceive of rediscovery as problemistic search over rugged landscapes. Whereas much of the economics literature has focused on how social learning *across* firms affects innovation over a rugged landscape, we adapt Malladi (2022) to study the innovation process *within* a rational, forward-looking firm.

In our model, there is a continuum of choices between 0 and 1 that have unknown payoffs. A searcher can learn the payoff to any given choice at a cost. Each period she decides whether to continue searching, and if so, which choice to learn about next. She eventually stops to take the best choice she had discovered so far.

Crucially, the searcher knows that there exists some choice which achieves a certain target payoff (e.g., that it is possible to design an invention of a given quality), but she does not know which. We interpret this as capturing rediscovery: the searcher knows with certainty, perhaps by seeing a predecessor's success, that a good discovery is possible, but she does not know *a priori* where to find it. The searcher also knows that the mapping from options to payoffs is Lipschitz continuous with a known Lipschitz constant. This assumption captures the idea that the searcher explores a rugged landscape, as she entertains a rich set of possibilities about the shape of mapping from options to payoffs. The assumption simultaneously captures learning from past failures and successes through trial and error. The searcher's past discoveries guide where she looks next, as continuity implies that proximate choices yield similar payoffs.

The searcher's utility upon concluding search is the payoff of her best discovery minus the sum of her accumulated search costs. We assume that the searcher follows a plan that, at every history, maximizes her *worst-case* utility upon concluding search. That is, she searches in a way that is robust to the shape of the complex and unpredictable rugged landscape. At each history, and for every plan of search, she evaluates this worst case over the possible shapes of the mapping from choices to payoffs, knowing only that this mapping must be Lipschitz continuous, pass through the points she had previously discovered, and attain the benchmark payoff somewhere.

We find an optimal search policy that is simple in many ways.

First, the searcher follows a threshold stopping rule, meaning she stops if the payoff she discovers exceeds that period's threshold and continues otherwise. Moreover, these thresholds increase with time: the knowledge that good discoveries exist somewhere causes the searcher to become emboldened rather than discouraged by bad discoveries.

Next, search proceeds from left to right, even though the searcher's choice set includes strategies that are not single-directional. The knowledge that good discoveries exist renders search into a process of elimination. Search is incremental, ruling out unfruitful regions of the search space and honing in on the location of more promising choices.

Third, the searcher has perfect recall but never invokes it. She always selects the last option she had discovered rather than returning to a previous discovery she had made.

Finally, while search happens in a non-stationary environment, the optimal search policy depends only on a simple index. We define the *search window* at a given history to be the set of choices which can potentially achieve the benchmark quality that the searcher would ideally rediscover. Both the optimal search and stopping rule are pinned down by the length of the search window.

None of these characterizations hold for the optimal policies identified in Malladi (2022), which studied search in a similar setting but when good discoveries are not guaranteed to exist. The comparison illustrates how rediscovery is procedurally simpler than search for original discovery.

Related Literature Our model contributes to a large literature on search theory. Early papers treat search as a pure stopping problem (McCall, 1970; Rothschild, 1974). Weitzman (1979) considers *ordered search* over a set of independent but not identically distributed items, where the agent can select which item to explore at each history and when to stop. Relaxing independence to allow cross-item learning is notoriously difficult, and a solution is known only in special case of conditionally independent items (Adam, 2001). Callander (2011) and Garfagnini and Strulovici (2016) capture richer learning by modeling the mapping from items to payoffs as the realized path of a Brownian motion. They study ordered search by a sequence of short-lived agents, whereas our model follows Malladi (2022) in solving for a fully forward-looking ordered search policy in the presence of learning.[1] The Brownian framework has also been used to capture aspects of learning in other settings beside sequential ordered search. Importantly, Urgun and Yariv (2024) and Wong (2025) study optimal, forward-looking contiguous search over a Brownian path, where an agent chooses when and how quickly to explore. We relax the constraint of contiguous search by allowing the searcher to freely choose where to explore. Bardhi (2024) and Bardhi and Bobkova (2023) study optimal information acquisition (by a single agent or by delegation to several agents, respectively) about a complex project with correlated attributes. Unlike these models, our agent samples the item set sequentially, and we look for the optimal dynamic search strategy.

Our paper is also related to a large literature on bandits and optimization in computer science and operations (Lattimore and Szepesvári, 2020; Hansen et al., 1992). One key difference is that addition to exploration, our agents face a stopping problem. Another key difference is that we study 'rational', i.e., forward looking and dynamically consistent, agents. This more conventionally economic approach shifts the focus from heuristics to characterizations of optimal strategies and comparative statics.

2 Model

There is a continuum of items arranged along the interval $S \equiv [0,1] \subset \mathbb{R}$. Let $Q \subset [0,1] \to \mathbb{R}$ be the set of potential *quality indices*—mappings from the search space to a measure of quality. There is some true quality index $q \in Q$, so each item $x \in S$ has a quality $q(x) \in [0,1]$.

There is a searcher who knows Q but not the true quality index. She can learn the quality of items in $[0,1]$ through costly search. This way, she narrows down the set of candidate true quality indices in Q.

In each period, $t = 0, 1, 2, 3\ldots$, the searcher takes one of two kinds of actions. She either

[1] In particular, we study sequentially robust search policies. For other recent perspectives on robustness and dynamics, see Li et al. (2024) and Auster et al. (2024).

explores a new item $x_t \in [0,1]$ to learn its quality, $q(x_t)$. Or she concludes her search, $x_t = \emptyset$, and adopts the highest quality item that she had discovered so far, including an outside option of quality zero.

Formally, let $h_t = \{(x_i, z_i)\}_{i=0}^{t-1}$ be the time t partial history when the searcher has not yet concluded search, with $z_i = q(x_i)$. Let H denote the set of all partial histories. Let X_{h_t} be the set of items that were explored at h_t. Let $z^*_{h_0} = 0$ be the outside option, and for $t \geq 1$, $z^*_{h_t} = \max\{0, z_0, \ldots, z_{t-1}\}$. If $x_i \in X_{h_t}$ is such that $z_i = z^*_{h_t}$, then x_i is a *best item at* h_t.

A quality index $\tilde{q} \in Q$ is *consistent* at h_t if $\tilde{q}(x_i) = z_i$ for all $i = 0, \ldots, t-1$. Let $Q_{h_t} \subset Q$ be the set of consistent quality indices at h_t.

We assume the following throughout:

Assumption 1. Q is the set of all L-Lipschitz continuous mappings $q : [0,1] \to \mathbb{R}$ such that $q(x) = 1$ for some $x \in [0,1]$.

In essence, the searcher knows little about the shape of the true quality index. She knows that proximate items in $[0,1]$ cannot be too different in quality. She also knows that there exists some item of at least a certain quality, so search may be worthwhile. But she does not know a priori where to find such an item. We call this known achievable quality the *quality standard* and normalize it to 1. Items that achieve the quality standard are *targets*.

2.1 Payoffs

The searcher's benefit to adopting item x is $q(x)$. The searcher's cost of exploring item x in any period is $c > 0$. The searcher's total payoff at history $h_{t+1} \in \tilde{H}$ such that $x_t = \emptyset$ is given by:

$$p(h_{t+1}) = z^*_{h_{t+1}} - c \cdot t.$$

That is, when an agent stops at history h_t, she will adopt an item of quality $z^*_{h_{t+1}}$.

2.2 Strategies and Policies

A strategy of the searcher is a deterministic mapping $\sigma \colon H \to \Big\{[0,1] \cup \{\emptyset\}\Big\}$. A strategy σ *eventually terminates* if for all $h \in H$ and $\tilde{q} \in Q_h$, σ reaches a terminal history from h when $q = \tilde{q}$. We restrict attention to the set of all strategies Σ that eventually terminate.

We denote by h_q^{+1} the history that follows h if the searcher adopts policy σ and the quality index is q, i.e.

$$h_q^{+1}(\sigma) = h \cup \Big\{\Big(\sigma(h), q(\sigma(h))\Big)\Big\},$$

and similarly h_q^{+2}, h_q^{+3}, etc. The set of reachable histories for a given quality index q is then $H_q^\sigma = \{h_0, (h_0)_q^{+1}(\sigma), (h_0)_q^{+2}(\sigma), \ldots\}$. We denote by $H^\sigma \subset H$ the set of histories that strategy σ can reach along her decision tree (i.e., for some $q \in Q$) starting from the empty history. Formally, $H^\sigma = \{H_q^\sigma : q \in Q\}$. A *search policy* $\sigma|_D$ is a restriction of σ to some domain $D \supset H^\sigma$. A search policy contains sufficient information to describe how search unfolds and when it stops for any true $\tilde{q} \in Q$, because it at least specifies actions for reachable histories. A search policy *terminates* if, for any q there exists $h \in H_q^\sigma$ such that $\sigma(h) = \emptyset$. Such a history h is called *terminal*.

2.3 Objective

To capture the unpredictability about the shape of the true quality index (i.e, the idea of rugged landscapes), we take the view that the searcher does not have a prior over Q. She seeks a strategy that maximizes the eventual payoff that she is guaranteed, starting from *any* history and regardless of which consistent quality index at that history is realized.

Under a strategy $\sigma \in \Sigma$ and starting from a history h, a quality index $q \in Q_h$ induces a terminal history h_q^σ and its corresponding terminal payoff $p(h_q^\sigma)$. A strategy σ^* is *optimal* if at the empty history $h = h_0$

$$\sigma^* \in \operatorname*{argmax}_{\sigma \in \Sigma} \left\{ \min_{q \in Q} p(h_q^\sigma) \right\}.$$

Similarly, a search policy $\sigma^*|_D$ is optimal if it can be extended to a strategy σ^* that satisfies the above condition.

The searcher can be thought of as choosing a fully contingent plan at time zero that maximizes her worst-case eventual payoff upon stopping. A natural question is: would the searcher stick to her (*ex-ante*) optimal plan at every later history even if she were given the chance then to revise it, i.e., is there a dynamically consistent optimal strategy? We show that the answer to this question is yes.

3 An Optimal Search Policy

In this section we characterize an optimal search policy in closed form. Our main result shows that, even though the set of feasible policies for the searcher is rich, and histories can be quite complex, there exist optimal policies that take a fairly simple form.

3.1 Classes of Simple Policies

We begin by formalizing our notion of simplicity of a search strategy. Each property captures an aspect of simplicity, and we point out how that restricts the space of simple policies.

Definition 1. A policy σ is a (left-to-right) *directional policy* if for every $h \in H^{\sigma}$, either $\sigma(h) = \emptyset$ or $\sigma(h) > x$ for all $x \in X_h$.

In words, a directional policy is one where the agent searches along one direction in the search space rather than bouncing back and forth. Note that fixed search direction might reasonably capture a shopper walking though the aisles of a grocery store or a pharmaceutical company experimenting incrementally with drug dosages. While some models assume directionality as a constraint on the search process (e.g., Arbatskaya (2007); Urgun and Yariv (2024); Wong (2025)) we do not.

Definition 2. A policy σ is a *threshold policy* if, for every non-terminal history $h \in H^{\sigma}$, there exists a τ_h such that $\sigma(h_q^{+1}) = \emptyset$ if and only if $z^*_{h_q^{+1}} \geq \tau_h$.

A policy is a threshold policy if, prior to searching, the agent has a threshold in mind such that if search yields a quality exceeding that threshold she will conclude search, and she will otherwise continue searching. While threshold stopping rules are typically optimal in simple search models where agents take independent draws from a known distribution, they need not be optimal when there is learning. For example, a searcher may stop after a sufficiently *bad* draw if this discourages her about the prospect of making good discoveries (e.g., see Rothschild (1978); Malladi (2022)).

Definition 3. A policy σ *ignores past discoveries* if $\sigma(h_t) = \emptyset$ implies $z^*_{h_t} = z_{t-1}$, i.e. the searcher always takes the last item discovered

In the model we assume the searcher has perfect recall and always takes the highest quality item discovered to date. Therefore a strategy can only ignore past discoveries if the searcher continues searching until her best discovery was her last.[2] Recall can be valuable in contexts with learning, as bad draws can cast previous good discoveries in better light (e.g., see Rothschild (1978); Malladi (2022)).

Note that histories are complex and their dimensionality increases with time. Optimal policies may potentially have infinite memory and depend intricately on the sequence

[2]Note that having the threshold property does not imply that the searcher ignores past discoveries. She can, for example, stop regardless of the outcome of her second search (i.e., a threshold stopping rule with a threshold of zero) and take the first discovery if her second one is worse.

of past realizations. Here, we define a class of policies that depend on a simple one-dimensional state variable of any history.

To that end, we introduce the notion of a *search window* at history h:

$$S_h \equiv \left\{x \in [0,1] \,\middle|\, \exists q \in Q_h \text{ s.t. } q(x) = 1\right\}.$$

The search window is the set of items x which are targets under some consistent quality index $q \in Q_h$. Alternatively, if we denote by $\overline{q}_h : [0,1] \to \mathbb{R}$ the upper-envelope of feasible quality indices at history h, then S_h is the set of items x for which $\overline{q}_h(x) \geq 1$.

The search window shrinks with additional searches, so $S_{h_q^{+1}} \subseteq S_h$ for any $q \in Q_h$.[3] At any history h where $(x,z) \in h$ and $z < 1$, the open interval of length $\frac{2(1-z)}{L}$ centered at x lies outside of the S_h. More generally, at history $h = \{(x_0,z_0),\ldots,(x_t,z_t)\}$,

$$S_h = [0,1] - \bigcup_{j=0}^{t} \left(x_j - \frac{(1-z_j)}{L}, x_j + \frac{(1-z_j)}{L}\right).$$

This is depicted in Figure 1.

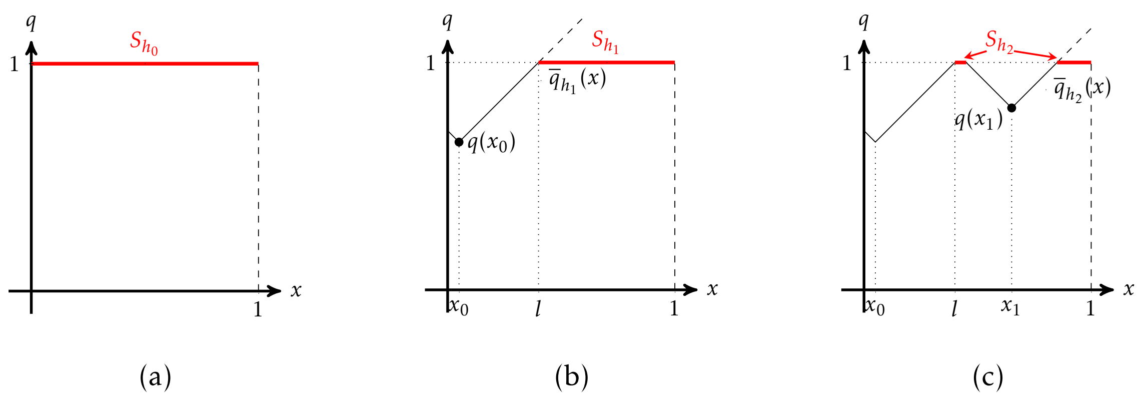

Figure 1: The red line represents the search window. The solid black line instead represents $\overline{q}_{h_t}$, the upper envelope of qualities given the current history.

Let $l_h \equiv \|S_h\|$, i.e., the lebesgue measure of the search window, for all $h \in H$.

Definition 4. A policy σ is an *index policy* if σ is measurable with respect to the lebesgue measure of the search window on H^σ.

Index strategies are those which depend on histories only through the length of the search window. This is particularly attractive as the unique histories grow instead ex-

[3]Observe that the set of feasible quality indices shrinks with additional searches, i.e., $Q_{h_q^{+1}} \subseteq Q_h$, and therefore $\overline{q}_h \geq \overline{q}_{h_q^{+1}}$ everywhere.

ponentially in the number of searches. Optimal search behavior can be fully characterized in terms of a single-dimensional "sufficient statistic" of the exponentially-large state space.

Finally, we return to the question of dynamic consistency. We say that a policy is dynamically consistent if it maximizes the searcher's worst-case eventual payoff after any reachable history, where worst-case is taken over the consistent quality at that history.

Definition 5. A policy σ is *dynamically consistent* if at all histories $h \in H^\sigma$

$$\sigma \in \operatorname*{argmax}_{\sigma \in \Sigma} \left\{ \min_{q \in Q_h} p\left(h_q^\sigma\right) \right\}$$

Note that a dynamically consistent strategy is also optimal, as Definition 5 holds in particular at the empty history. Models that assume Bayesian uncertainty are guaranteed dynamic consistency, while non-Bayesian models often lack this property. For example, Auster et al. (2024) study a pure stopping problem in the context of information acquisition with ambiguity. They finds an optimal policy that is dynamically inconsistent, exhibiting non-monotonicity in beliefs and randomized stopping.

3.2 Main Result

Our main result is that:

Theorem 1. *There exists an optimal policy which is directional, threshold, index and ignores the past. Furthermore, this optimal policy is dynamically consistent.*

The remainder of this section is dedicated to constructing such a policy explicitly. We fix a Lipschitz constant $L = 1$ for ease of notation.

A *(left-to-right) ordered search history* is a history h such that the search window is a right-aligned interval, of the form $[a, 1]$ for some a. We denote the set of such histories as $H^I \subset H$ and their lebesgue measure, which corresponds to their length, is $1 - a$. Figure 1 shows the search windows at three histories: the first two are ordered search histories, because S_{h_0} and S_{h_1} are intervals. The last is not an ordered search history, as S_{h_2} is disconnected.

We define a search policy $\sigma_{L \to R}$ in which the searcher explores S from left to right and stops whenever she makes a discovery exceeding an increasing, history-dependent threshold. To this end, we introduce two auxilliary functions.

Let $N\colon [0,1]^2 \to \mathbb{N}$ be defined as follows:

$$N(c,l) \equiv \begin{cases} 0 & \text{if } c \in \left(1-\frac{l}{2}, 1\right], \\ 1 & \text{if } c \in \left(\frac{l}{2}, 1-\frac{l}{2}\right], \\ n & \text{if } c \in \left(\frac{l}{n(n+1)}, \frac{l}{n(n-1)}\right]. \end{cases} \tag{1}$$

Roughly, N maps search costs and the length of a search window to the maximum number of searches that $\sigma_{L\to R}$ makes for any $q \in Q$. In keeping with this interpretation, N is decreasing in costs. Next, when costs of search are sufficiently low, N is increasing with interval length: more space left to explore means more searches might be needed to discover a good quality item. But when costs are sufficiently high, N is decreasing with interval length increases: more space left to explore discourages the searcher from exploring at all.

When $N(c,l) \neq 0$, define $\phi\colon [0,1] \to \mathbb{R}$ as

$$\phi(l) = 1 - \frac{l}{2N(c,l)} - \frac{N(c,l)-1}{2}c \tag{2}$$

The function ϕ maps the length of a search window to a quality threshold that is used to define the stopping region in $\sigma_{L\to R}$. It is straightforward to check that ϕ is decreasing. As the search window grows larger, more search is potentially required to find a good outcome. Therefore, the searcher is willing to conclude search for lower quality discoveries. When the remaining space to be seached is small, further search is likely to secure items close to the benchmark quality, so the threshold for stopping is higher.

Definition 6. The *left-to-right* search policy $\sigma_{L\to R}\colon H^{\sigma_{L\to R}} \to \Delta\{S \cup \emptyset\}$ is given by

$$\sigma_{L\to R}(h) = \begin{cases} \emptyset & \text{if } z_h^* \geq \phi(l_h) - c, \\ 1 - l_h + 1 - \phi(l_h) & \text{otherwise}, \end{cases}$$

for all $h \in H^{\sigma_{L\to R}}$.

Figure 2 shows that $\sigma_{L\to R}$ has a simple geometric characterization at any ordered search history h. Let $k \leq 2c$ be the largest real number such that the search window S_h can be partitioned in balls of diameter $k, k+2c, k+4c, \ldots$. [4] The number of such balls is $N(c, l_h)$. Order these balls from the largest to the smallest in the search window. The

[4]Note that such a k is unique.

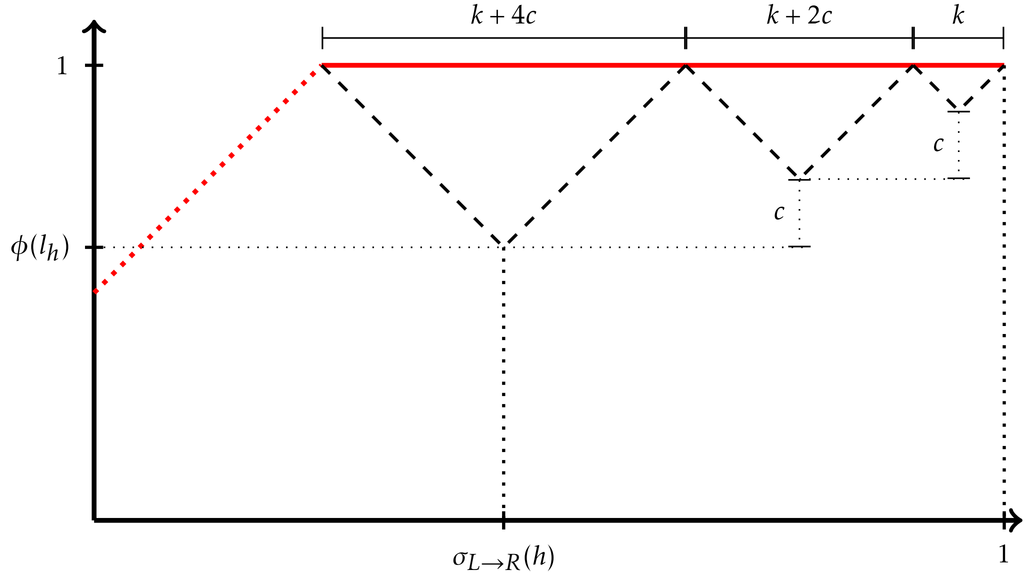

Figure 2: Visualizing $\sigma_{L\to R}$ at some ordered search history h, where the search window is given by the solid red line.

left-to-right policy searches at the center of the largest ball at history h and stops if and only if $q(\sigma_{L\to R}(h))$ attains a value of at least the peak of the corresponding triangle.

Although $\sigma_{L\to R}$ is well-defined at all histories, the left-to-right policy stops at the first history that is not ordered.

Lemma 1. *For any $q \in Q$ and non-terminal $h \in H^{\sigma_{L\to R}}$, exactly one of the following conditions is satsified:*

1. *$h_q^{+1}(\sigma_{L\to R})$ is a terminal history, or*
2. *$h_q^{+1}(\sigma_{L\to R})$ is an ordered search history.*

The proof of this Lemma is in Appendix A, but Figure 3 provides some intuition. Suppose search reveals that $q(x) > \phi(l)$, as in Figure 3a. The history following this observation is no longer an ordered search history, but $\sigma_{L\to R}$ stops here. Instead, suppose that the searcher observes a quality $q(x)$ lower than the threshold, as in Figure 3b. Then, the history that follows is an ordered search history, and $\sigma_{L\to R}$ determines what to do next.

Proposition 1. *The left-to-right policy $\sigma_{L\to R}$ is directional, threshold, index, and ignores past discoveries.*

Proof. First, the left-to-right policy is *threshold* by definition.

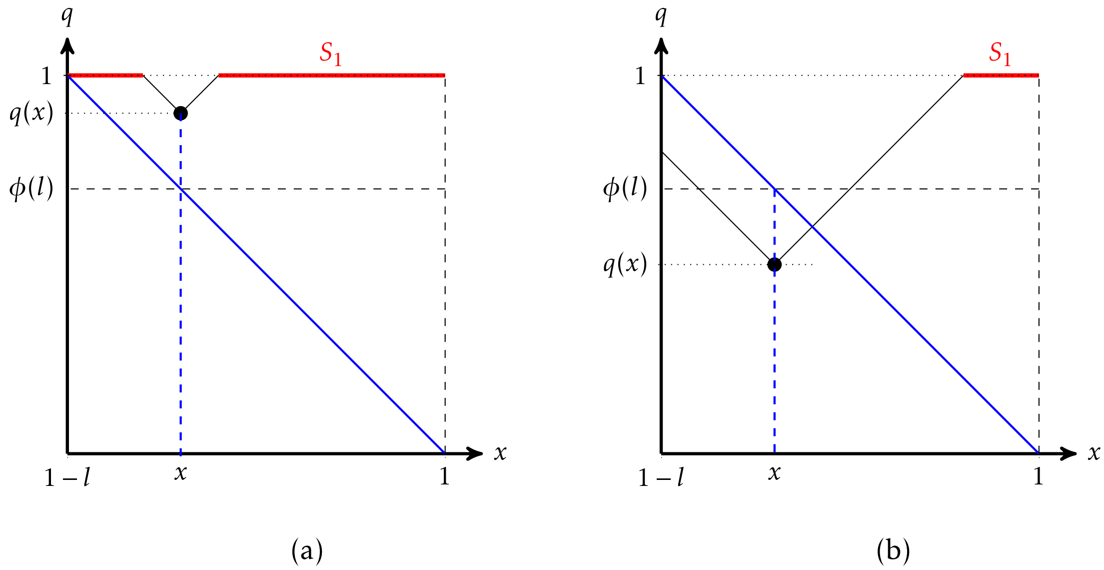

Figure 3

Second, Lemma 1 shows that if the policy doesn't stop, the search window shrinks: $l_{h_q^{+1}} < l_h$. Because $\phi(\cdot)$ is decreasing, $\phi(l_{h_q^{+1}}) > \phi(l_h)$, so the thresholds are increasing and thus the left-to-right policy *ignores past discoveries*.

Third, we can further improve Lemma 1's inequality $l_{h_q^{+1}} < l_h$: in fact, if the policy doesn't stop, then $l_{h_q^{+1}} \leq l_h - 2(1-\phi(l_h))$. Then

$$\begin{aligned}\sigma_{L\to R}(h_q^{+1}) > 1 - l_{h_q^{+1}} &\geq 1 - l_h + (2 - \phi(l_h)) \\ &> 1 - l_h + 1 - \phi(l_h) = \sigma_{L\to R}(h),\end{aligned}$$

where the last inequality follows from the fact that $\phi(l) \leq 1$ for any l (see Lemma 4 in Appendix B for a formal proof). This proves that the left-to-right policy is *directional*.

Finally, the threshold depends on the history only through the length of the search window l_h, so $\sigma_{L\to R}$ is *index*. □

We then prove that the left-to-right policy is optimal, and in the process we also show that any optimal policy is dynamically consistent.

Theorem 2. *$\sigma_{L\to R}$ is an optimal, dynamically consistent search policy.*

3.3 Discussion

The optimal search policy bears some resemblance to the observations made in the literature on problemistic search and rugged landscapes.

First, we find that the searcher is never discouraged by a poor payoff draw, because she knows good discoveries exist. By contrast, she becomes more ambitious and sets a higher stopping threshold for her next search. The idea that searchers attempting rediscovery are not easily discouraged corresponds with the observation by Cyert and March (1963) that, in problemistic search by firms and managers, *"[so] long as the problem is not solved, search will continue."*

The stopping threshold in our model keeps increasing if unmet, generating a behavior we defined "ignoring past discoveries". Contrast this dynamic with the classic insight of the literature on reference points and aspiration formation (see e.g. Dalton et al. (2018), Selten (1998), Karandikar et al. (1998)), where unrealized aspirations negatively affect utility and therefore adapt to what is learned to be feasible. Instead, here the existence of a benchmark that the searcher knows to be feasible acts in the opposite direction. Low quality discoveries do not affect expectations, but contribute to the spatial learning component, pushing the searcher's aspirations higher.

On the other hand, the searcher in our model does not begin to explore the space unless the quality standard lies sufficiently higher than the status quo. It reflects the finding by Baum and Dahlin (2007) that *"[performance] below aspirations [...] leads decision makers to initiate experimentation to identify new ways of doing things and new things to do, while satisfactory performance does not."*

Finally, we briefly mentioned that multiple prior models like ours often lack dynamic consistency. The literature, starting with Epstein and Schneider (2003), has derived sufficient conditions for dynamic consistency to hold for maxmin expected utility (MEU), variational and smooth ambiguity preferences. Such conditions boil down to requiring prior-by-prior updating of the decision maker's beliefs and *rectangularity* of the set of priors. One can indeed verify that rectangularity holds in our model. First, the set of priors for our decision maker is the set of dirac measures over the space of quality indices. The decision maker updates her beliefs prior-by-prior: the set of posteriors is exactly the set of dirac measures over quality indices that are consistent with her observation. Finally, rectangularity requires that any combination of marginal beliefs with *any* conditional probability distribution is a plausible prior. Marginal beliefs in our model are fully concentrated Dirac measures, hence every such combination is simply a posterior distribution, which is itself a Dirac measure contained in the set of priors.

4 Analysis of the Two Period Case

We begin by solving the model for costs at which the agent would never search more than twice. This special case of the general problem illustrates how the searcher hedges against uncertainty over the shape of the mapping by her choice of initial search location. It also illustrates for what quality realizations the searcher continues or stops, highlighting the key tradeoff between the quality and informativeness of discoveries. Extending the ideas developed here, Section 5 solves the general case where the agent may face lower costs and search many times.

4.1 Relevant Costs

First, we identify a range of costs for which the agent searches at least once and most twice under any quality index $q \in Q$.

Claim 1. If $c \in (\frac{1}{4}, \frac{1}{2})$, the agent searches at least once and at most twice in any optimal search policy, for every $q \in Q$.

Proof. If the agent explores item 1/2 and stops, then her payoff is at least $1/2 - c > 0$ for any $q \in Q$. Therefore, the agent searches at least once.

Next note that if the agent searches item 1/4 and next searches item 3/4, she can guarantee herself a payoff of at least $3/4 - 2c$. Therefore, a lower bound on the agent's payoff if she explores twice is $3/4 - 2c$. An upper bound on the searcher's payoff if she explores $k \geq 3$ times is $1 - kc$. Because $1/4 < c$, we have $1 - kc \leq 1 - 3c < 3/4 - 2c$. Therefore, the agent never searches more than twice in an optimal search policy. □

Henceforth in this section we assume $c \in (\frac{1}{4}, \frac{1}{2})$.

4.2 Bifurcation Risk

An agent faces *bifurcation risk* at a history where she had searched once and her initial search, x, is of sufficiently high quality ($q(x) \geq \max\{x, 1 - x\}$) but it is below the quality standard ($q(x) < 1$). A discovery of this quality leaves open the possibility that targets exist either to the left or right of the initial search; see the left panel of Figure 4. The agent finds herself at a crossroads in deciding where to conduct her next search. In the worst case, the direction she picks leads to a poor outcome, revealing that high quality discoveries instead exist on the other side of the search space.

Claim 2. An optimal search policy concludes when the agent faces bifurcation risk.

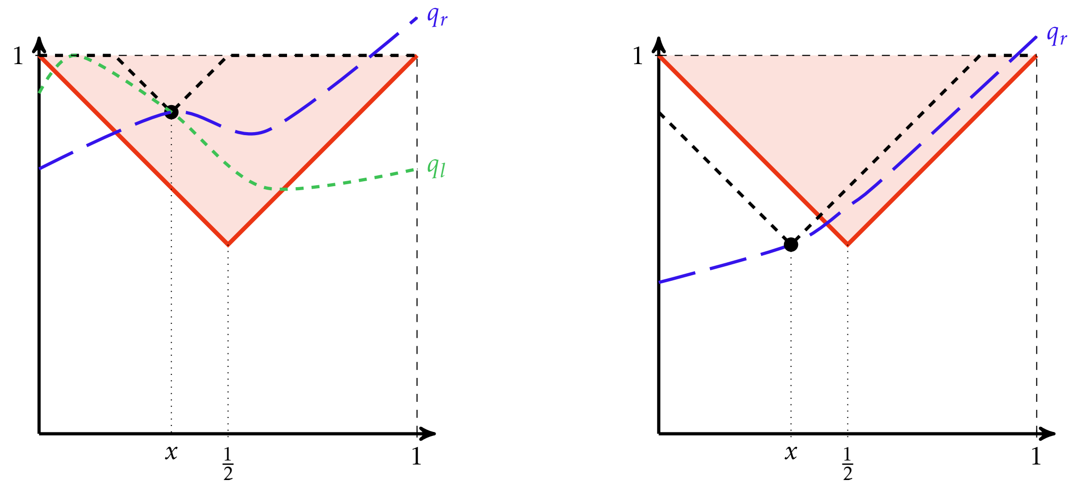

Figure 4: The agent in the first figure explores x and faces bifurcation risk. Given what she knows, it is possible that target locations exist either exclusively to the left (e.g., if $q = q_l$) or right of her initial search (e.g., if $q = q_r$), and she guesses the wrong side to search next. The agent in the second figure does not face bifurcation risk. Due to the Lipschitz constraint, any feasible q must lie on or below the dashed black line. Therefore, target locations must exist exclusively near the right end of the search space.

Proof. Suppose the agent faces bifurcation risk after searching at x. Let

$$q_l(y) = \begin{cases} q(x), \text{ if } y \leq x \\ \min\{q(x) + y - x, 1\}, \text{ if y > x.} \end{cases}$$

Note that q_l is feasible at this history. If $q = q_l$, the agent's payoff if she searches to the left of x and then concludes is $q(x) - 2 \cdot c < q(x) - c$. If she searches to the left of x, then at 1 and then concludes, she gets a payoff of $1 - 3c < 1/2 - c \leq q(x) - c$. Therefore, stopping at x improves on the agent's worst case payoff to searching to the left of x, regardless of what she does afterward. Symmetrically, concluding search is also better than exploring to the right of x. □

4.3 Directional Risk

If the agent searches too close to 0 or 1, she risks having searched on the wrong side of the interval. Discovering a sufficiently low quality is valuable, because it narrows down the location of the target. If the quality the agent discovers is sufficiently low, further search is bound by Lipschitz continuity to yield an item close to the target. This is the case on the left panel of Figure 5. Instead, if the quality she discovers is sufficiently high, the search region remains large and an additional search cannot, in the worst case, be profitable. The right panel of Figure 5 shows such an example. We say that the agent faces *directional risk* when she does not face bifurcation risk and she discovers a quality $q(x) \geq \min\left\{\frac{2}{3}(1-c)+\frac{x}{3}, \frac{2}{3}(1-c)-\frac{x-1}{3}\right\}$. Figure 5 highlights such region in blue.

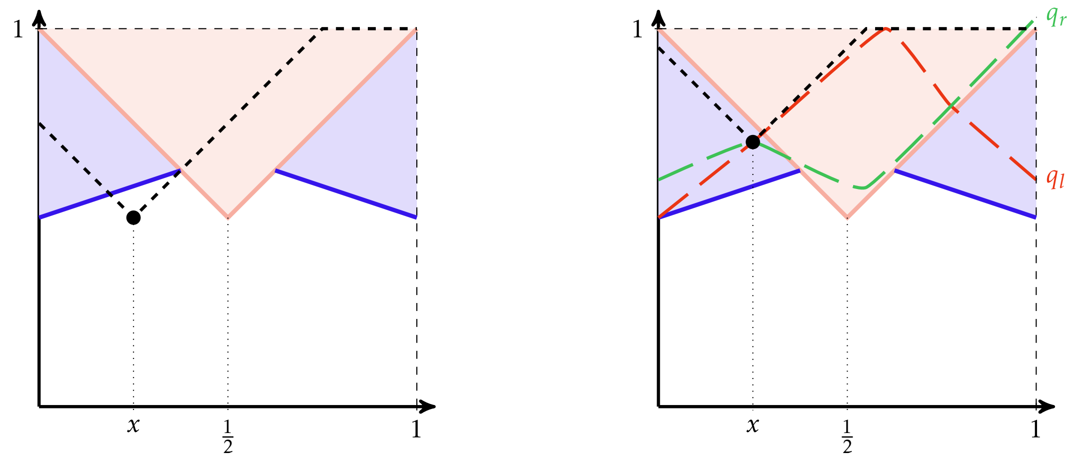

Figure 5: The agent explores x and in the first figure she discovers a low quality $q(x)$. Lipschitz continuity then narrows down the possible target locations on the right end of the interval. Instead, in the second figure the agent discovers a high quality, which does not help significantly in narrowing down the location of the target.

Claim 3. An optimal search policy concludes when the agent faces directional risk.

Proof. Suppose the agent discovered a quality such that she is facing directional risk. The worst-case quality index is similar to the one pictured in the right panel of Figure 5. If the first search was to the left of $\frac{1}{2}$, the searcher is left with a search window of measure $q(x)-x$. Since the agent will optimally search at most twice, the worst-case quality she anticipates over this region is $1-\frac{q(x)-x}{2}$. She will search for a second time only if this quality minus the cost of search exceeds the quality $q(x)$ she already discovered. That is,

$$q(x) < 1 - \frac{q(x)-x}{2} - c \iff q(x) < \frac{2}{3}(1-c) + \frac{x}{3}.$$

This is a contradiction, since the agent faces directional risk only if the quality she discovered is $q(x) \geq \min\left\{\frac{2}{3}(1-c)+\frac{x}{3}, \frac{2}{3}(1-c)-\frac{x-1}{3}\right\} = \frac{2}{3}(1-c)+\frac{x}{3}$. Similarly, if the first search is to the right of $\frac{1}{2}$, she is left with a search region of measure $x+q(x)-1$, and thus an expected worst-case quality of $\frac{3-x-q(x)}{2}$. Then, she will search a second time only if

$$q(x) < \frac{3-x-q(x)}{2} - c \iff q(x) < \frac{2}{3}(1-c) - \frac{x-1}{3},$$

which is again a contradiction. □

4.4 Optimal Search Policy

Suppose the agent is choosing which location to the left of $\frac{1}{2}$ to search. When considering a location, the agent worries about the worst-case quality index that she could be facing. In particular, she is not worried about discovering extremely low qualities, because that would help narrowing down the target's location for future search. In fact, Claim 3 proves that when an agent receives quality $q(x_0) < \frac{2}{3}(1-c) + \frac{x_0}{3}$ for her first search, the quality of her second search will be at least $1 - \frac{q(x_0)-x_0}{2} > \frac{2}{3}(1-c) + \frac{x_0}{3}$.

Naturally the agent doesn't worry about discovering extremely high qualities either: she will stop searching at quality $q(x) > \min\left\{\frac{2}{3}(1-c) + \frac{x}{3}, 1-x\right\}$. The worst-case quality at x is exactly $q(x) = \min\left\{\frac{2}{3}(1-c) + \frac{x}{3}, 1-x\right\}$, where the agent is either indifferent between stopping and continuing search or strictly prefers to stop. The agent then selects a search location expecting the quality to fall on the lower envelope of the region highlighted in Figure 6.

It is clear from the figure that there are two optimal search locations. At these points, the agent is hedging optimally against directional and bifurcation risk. When searching closer the center, the agent risks discovering a low quality with bifurcation exposure. When searching closer to the ends of the search space, the agent risks discovering a low quality with directional exposure. At the optimal search locations, the agent balances the two risks. To find the optimal search locations, it suffices to solve the following system of equations:

$$\begin{cases} q(x) = \min\left\{\frac{2}{3}(1-c) + \frac{x}{3}, \frac{2}{3}(1-c) - \frac{x-1}{3}\right\}, \\ q(x) = \max\{x, 1-x\}, \end{cases}$$

which implies that $x_0 \in \left\{\frac{1}{4} + \frac{c}{2}, \frac{3}{4} - \frac{c}{2}\right\}$.

At these locations, the agent stops searching if the worst case materializes. If the true quality of location x is larger than the worst case, the agent will also stop searching. She is now facing bifurcation, and her next search is worthless, but the quality she discovered is sufficiently large. If the true quality of location x is instead lower than the worst case, the agent can rule out a large portion of the search space. The target must lie at the other

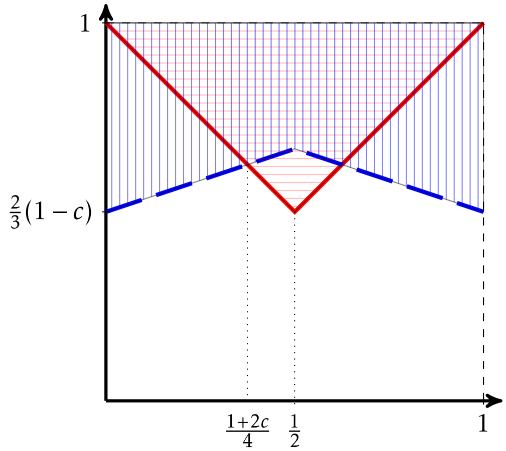

Figure 6: In horizontal red lines we depict the region where the agent experiences bifurcation risk. It is the region above the solid red line, $y = \max\{x, 1-x\}$. Instead, in vertical blue lines we depict the region where the agent experiences directional risk. It is the region above the dashed blue line, $y = \min\left\{\frac{2}{3}(1-c) + \frac{x}{3}, \frac{2}{3}(1-c) - \frac{x-1}{3}\right\}$. The minimum between these two lines is the worst-case quality for all locations.

extreme of the search space, and one future search at the center of the updated search window is sufficient to discover a sufficiently high-quality item.

It is instructive to compare the optimal search policy in our sequential search model to the optimal policy in a simultaneous search model. If the searcher had to pick two locations at once to search simultaneously, she would choose locations 1/4 and 3/4. This configuration ensures an item of quality 1 cannot be farter than distance 1/4 from one of the chosen locations, ensuring that the searcher gets a payoff of at least $3/4 - 2c$. The first search of the optimal sequential policy cleaves closer to the center. The difference arises because in sequential search, there is option value in deciding whether or not to continue with a second search. Getting a lower draw is more informative when searching closer (but not exactly at) the center, because it generates more information for a second search. Under the optimal policy, the searcher gets a payoff of at least $3/4 - 1.5c$ (see Figure 6).

5 Proof of the General Result

We convey the key ideas for the proof of Theorem 2 in the general case. For any strategy $\sigma \in \Sigma$, the continuation value at any history $h \in H$ is

$$V_q^\sigma(h) \equiv p(h_q^\sigma) - (p(h) - z_h^*).$$

First, we argue that at a non-terminal ordered search history h and conditional on following $\sigma_{L\to R}$, the quality that minimizes the searcher's continuation payoff after searching at $\sigma_{L\to R}(h)$ is $\phi(l_h)$:

Lemma 2. *For any non-terminal $h \in H^{\sigma_{L\to R}}$ and for all $q \in Q_h$,*

$$\phi(l_h) - c \leq V_q^{\sigma_{L\to R}}(h).$$

Clearly, the searcher is better off if she discovers a quality above the threshold, as $\phi(l_h)$ is the lowest quality for which the policy recommends stopping. The argument that discovering any quality below the threshold improves the searcher's eventual payoff upon stopping is less immediate. On the one hand, discovering low quality is *informative*: the search window shrinks, narrowing the location of target items. On the other hand, exploiting this information requires additional costly searches.

To understand why the tradeoff always goes in the searcher's favor, recall the geometric depiction of $\sigma_{L\to R}$ at an ordered search history h in Figure 2. Suppose searching at $\sigma_{L\to R}(h)$ reveals a quality lower than $\phi(l_h)$. The search window at this new history, h_q^{+1}, shrinks to a region smaller than the interval covered by the triangles centered at x_2 and x_3, shown in Figure 7. At h_q^{+1}, consider the 'non-responsive' strategy σ that searches at x_2, stops if $x_2 \geq \phi(l_h) + c$, and searches and stops at x_3 otherwise. This strategy nets the searcher either $\phi(l_h) + c - c = \phi(l_h)$ after the first search or $\phi(l_h) + 2c - 2c = \phi(l_h)$ in the second search. A discovery of quality lower than $\phi(l_h) + 2c$ is impossible if x_3 is searched on path: the Lipschitz bounds would imply that the search window is empty at such a history, constradicting the fact that a target item exists. Strategy $\sigma_{L\to R}$ performs even better than the non-responsive strategy at history h_q^{+1} by adapting the search location to the smaller search window, proving Lemma 2.

Next we argue that at any ordered search history, the searcher can do no better by using a strategy other than $\sigma_{L\to R}$:

Lemma 3. *For any non-terminal $h \in H^{\sigma_{L\to R}}$ and $\sigma \in \Sigma$, there is a $\tilde{q} \in Q_h$ with*

$$V_{\tilde{q}}^{\sigma}(h) \leq \phi(l_h) - c.$$

Suppose σ is a strategy that always searches inside the search window in any period. We construct a feasible quality index $\tilde{q}$ such that the latest discovery at any step is c better than the previous.[5] Such a quality index at most maintains the searcher's con-

[5] If σ searches outside the search window, the greatest feasible improvement in quality is even smaller; compare $q(x_3)$ and $q(x_2)$ in Figure 8.

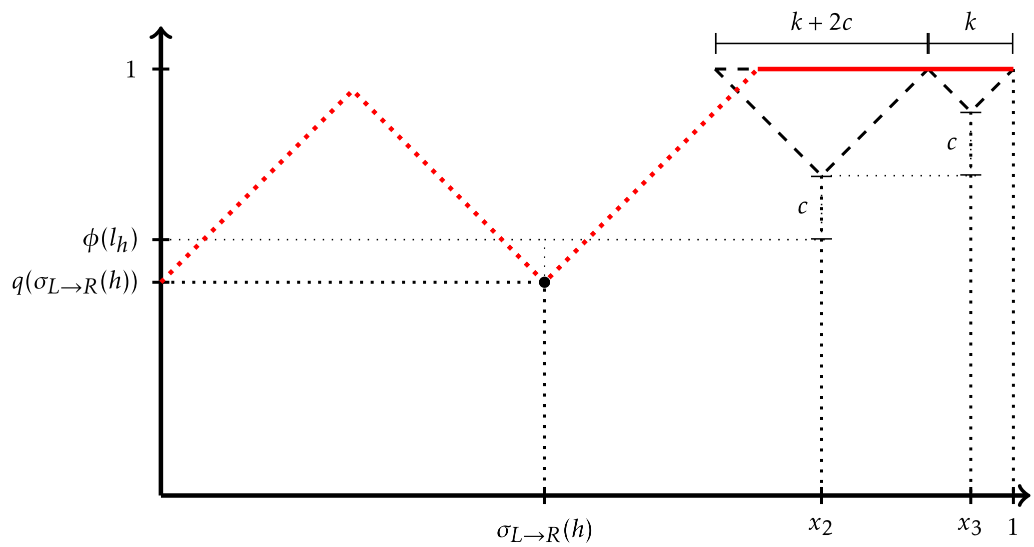

Figure 7: When the searcher discovers a quality $q(\sigma_{L\to R}(h)) < \phi(l_h)$, the search window shrinks to the solid red region in this image, with size $l_{h_q^{+1}}$ less than $k+(k+2c)$. Suppose the searcher chose to search locations x_2 and x_3 sequentially. If she discovered worse qualities than those at the vertices of the respective triangles, the search window would be empty, which is a contradiction. By searching at the peaks of the triangles, the searcher must then be able to secure a continuation utility $V_q^{\sigma}(h)$ larger than $\phi(l_h)-c$.

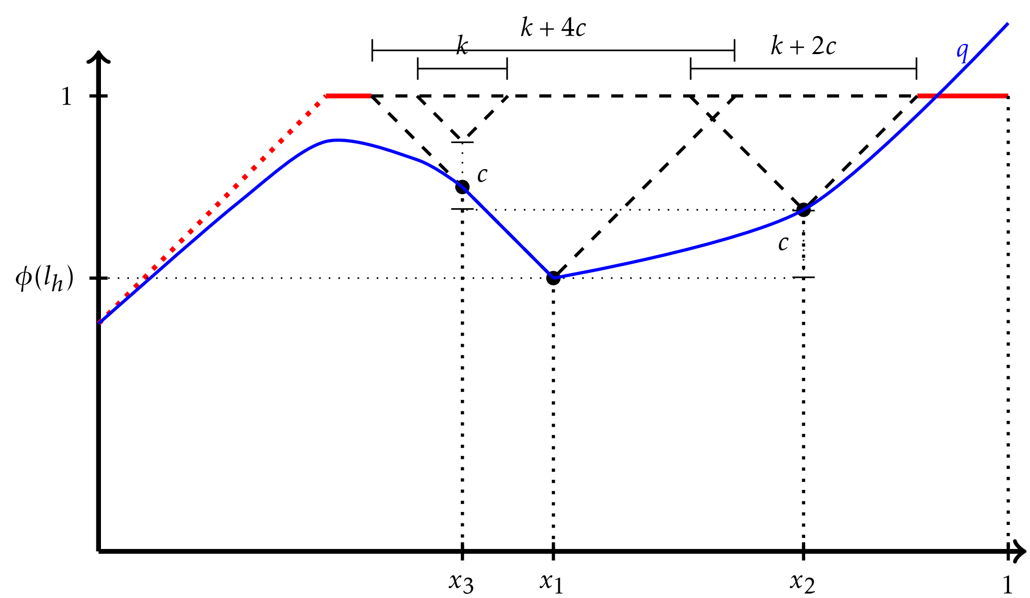

Figure 8: Searching in $x_1 \neq \sigma_{L\to R}(h)$, x_2, and x_3 leaves a non-empty search window. Then, there exist quality indices that pass through the items discovered and attain the target. The searcher will never do better than the threshold $\phi(l_h)$.

tinuation value constant, proving that $\phi(l_h) - c$ is an upperbound to the searcher's belief about her continuation payoff. Interestingly, the threshold is such that the searcher will never search more than $N(c, l_h)$ items, thus making $N(c, 1)$ the effective game horizon (and guaranteeing that any optimal strategy must terminate).

That this policy is dynamically consistent follows from the recursive nature of the proof. Both Lemma 2 and Lemma 3 apply to *all* non-terminal histories, thus the agent is incentivized to follow her contingent plan of action after any discovery. It is key here that the agent cannot hedge against worst-case realizations by choosing a randomized policy, or equivalently, that the agent evaluates her payoff guarantees ex-post, after any realization of the uncertainty. If instead she evaluated her payoff ex-ante, randomization would improve the payoff guarantee at the cost of dynamic consistency.

6 Conclusion

Running a mile under 4 minutes was a feat that was seriously attempted by athletes since the 1880s. For many decades, it was unclear whether the barrier to achieving this feat were physical or psychological. However, in the 1954, Roger Bannister, an unlikely and iconoclastic runner who eschewed the training wisdom of the day finally ran a mile in under 4 minutes. Several other soon followed him, and today it is an impressive but not rare feat. A tribute to Bannister in the *Harvard Business Review* writes (Taylor (2018)):

> "...what goes for runners goes for leaders running organizations. In business, progress does not move in straight lines. Whether it's an executive, an entrepreneur, or a technologist, some innovator changes the game, and that which was thought to be unreachable becomes a benchmark, something for others to shoot for. That's Roger Bannister's true legacy..."

We model search for rediscovery and confirm the idea that simply knowing that some target is achievable dramatically changes how agents look for it. Having a target to shoot for renders search for rediscovery into a process of elimination. It simplifies the search process and ensures that agents who find it worthwhile to explore at least come close to hitting their target.

References

Klaus Adam. 2001. Learning while searching for the best alternative. *Journal of Economic Theory* 101, 1 (2001), 252–280.

Maria Arbatskaya. 2007. Ordered search. *The RAND Journal of Economics* 38, 1 (2007), 119–126.

Sarah Auster, Yeon-Koo Che, and Konrad Mierendorff. 2024. Prolonged learning and hasty stopping: the wald problem with ambiguity. *American Economic Review* 114, 2 (2024), 426–461.

Arjada Bardhi. 2024. Attributes: Selective learning and influence. *Econometrica* 92, 2 (2024), 311–353.

Arjada Bardhi and Nina Bobkova. 2023. Local evidence and diversity in minipublics. *Journal of Political Economy* 131, 9 (2023), 2451–2508.

Joel AC Baum and Kristina B Dahlin. 2007. Aspiration performance and railroads' patterns of learning from train wrecks and crashes. *Organization Science* 18, 3 (2007), 368–385.

Stephan Billinger, Nils Stieglitz, and Terry R Schumacher. 2014. Search on rugged landscapes: An experimental study. *Organization Science* 25, 1 (2014), 93–108.

Steven Callander. 2011. Searching and learning by trial and error. *American Economic Review* 101, 6 (2011), 2277–2308.

Steven Callander, Nicolas Lambert, and Niko Matouschek. 2022. *Innovation and competition on a rugged technological landscape*. Technical Report. Working paper Northwestern University.

R. M. Cyert and J. G. March. 1963. A behavioral theory of the firm. PrenticeHall/Pearson Education.

Patricio S. Dalton, Julius Ruschenpohler, and Bilal Husnain Zia. 2018. *Determinants and dynamics of business aspirations : evidence from small-scale entrepreneurs in an emerging market*. Policy Research Working Paper Series 8400. The World Bank. https://ideas.repec.org/p/wbk/wbrwps/8400.html

Larry G Epstein and Martin Schneider. 2003. Recursive multiple-priors. *Journal of Economic Theory* 113, 1 (2003), 1–31.

Umberto Garfagnini and Bruno Strulovici. 2016. Social experimentation with interdependent and expanding technologies. *Review of Economic Studies* 83, 4 (2016), 1579–1613.

Pierre Hansen, Brigitte Jaumard, and Shi-Hui Lu. 1992. Global optimization of univariate Lipschitz functions: I. Survey and properties. *Mathematical programming* 55, 1 (1992), 251–272.

Rajeeva Karandikar, Dilip Mookherjee, Debraj Ray, and Fernando Vega-Redondo. 1998. Evolving aspirations and cooperation. *journal of economic theory* 80, 2 (1998), 292–331.

Tor Lattimore and Csaba Szepesvári. 2020. *Bandit algorithms*. Cambridge University Press.

Daniel A Levinthal. 1997. Adaptation on rugged landscapes. *Management science* 43, 7 (1997), 934–950.

Zihao Li, Jonathan Libgober, and Xiaosheng Mu. 2024. Sequentially Optimal Pricing under Informational Robustness. arXiv:2202.04616 [econ.TH]

Suraj Malladi. 2022. Searching in the Dark and Learning Where to Look. *Available at SSRN 4084113* (2022).

John Joseph McCall. 1970. Economics of information and job search. *The Quarterly Journal of Economics* (1970), 113–126.

Michael Rothschild. 1974. Searching for the Lowest Price When the Distribution of Prices Is Unknown. *Journal of Political Economy* 82, 4 (1974), 689–711.

Michael Rothschild. 1978. Searching for the lowest price when the distribution of prices is unknown. In *Uncertainty in Economics*. Elsevier, 425–454.

Reinhard Selten. 1998. Aspiration adaptation theory. *Journal of mathematical psychology* 42, 2-3 (1998), 191–214.

Bill Taylor. 2018. What Breaking the 4-Minute Mile Taught Us About the Limits of Conventional Thinking. *Harvard Business Review* (2018). https://hbr.org/2018/03/what-breaking-the-4-minute-mile-taught-us-about-the-limits-of-conventional-thinking

Can Urgun and Leeat Yariv. 2024. Contiguous Search: Exploration and Ambition on Uncharted Terrain. *Journal of Political Economy* (2024), Forthcoming.

Martin L Weitzman. 1979. Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society* (1979), 641–654.

Yu Fu Wong. 2025. Forward-looking Experimentation of Correlated Alternatives. *Theoretical Economics* (2025), Forthcoming.

A Proofs

Recall that all proofs are developed in continuation value space. Unless necessary to avoid confusion, we drop the dependence of reachable histories $h_q^{+k}(\sigma)$ on the policy σ.

Proof of Lemma 1.

Because h is non-terminal, $\sigma_{L\to R}(h) = 1 - l_h + 1 - \phi(l_h)$. We consider two cases.

Case 1: $q(\sigma_{L\to R}(h)) \geq \phi(l_h)$. Then,

$$S_{h_q^{+1}} = [1 - l_h, \sigma_{L\to R}(h) - 1 + q(\sigma_{L\to R}(h))] \cup [\sigma_{L\to R}(h) + 1 - q(\sigma_{L\to R}(h)), 1],$$

where, by the assumption in this case,

$$\begin{aligned}\sigma_{L\to R}(h) - 1 + q(\sigma_{L\to R}(h)) &\geq \sigma_{L\to R}(h) - 1 + \phi(l_h)\\ &= 1 - l_h + 1 - \phi(l_h) - 1 + \phi(l_h)\\ &= 1 - l_h.\end{aligned}$$

Therefore, h_q^{+1} is not an ordered search history. Let $n = N(c, l_h)$. Then, the length of the search window is

$$\begin{aligned}l_{h_q^{+1}} &> q(\sigma_{L\to R}(h)) - \sigma_{L\to R}(h)\\ &= 2(\phi(l_h)) - 2 + l_h\\ &= \frac{n-1}{n} l_h - (n-1)c.\end{aligned}$$

Because $\phi(\cdot)$ is decreasing, $\phi(l_{h_q^{+1}}) < \phi\Big(\frac{n-1}{n} l_h - (n-1)c\Big)$. By Equation 1, $n(n-1)c \leq l_h < n(n+1)c$. This implies that $(n-1)(n-2)c \leq \frac{n-1}{n} l_h - (n-1)c < (n-1)nc$. Therefore,

$$\phi\Big(\frac{n-1}{n} l_h - (n-1)c\Big) = 1 - \frac{1}{2(n-1)}\Big(\frac{n-1}{n} l_h - (n-1)c\Big) - \frac{n-2}{2} c$$

$$= 1 - \frac{l_h}{2n} - \frac{n-1}{2}c + c$$
$$= \phi(l_h) + c.$$

Therefore,

$$\phi(l_{h_q^{+1}}) - c < \phi(l_h) \leq q(\sigma_{L\to R}(h)).$$

Because $\sigma_{L\to R}$ ignores past discoveries, this directly implies that h_q^{+1} is terminal according to Definition 6.

Case 2: Suppose instead that $q(\sigma_{L\to R}(h)) < \phi(l_h)$. Then,

$$S_{h_q^{+1}} = [\sigma_{L\to R}(h) + 1 - q(\sigma_{L\to R}(h)), 1],$$

so h_q^{+1} is an ordered search history. By assumption $l_{h_q^{+1}} < \frac{n-1}{n} l_h - (n-1)c$, so $\phi(l_h^{+1}) > \phi(l_h) + c$. Then, combining, we get $q(\sigma_{L\to R}(h)) < \phi(l_h) < \phi(l_h^{+1}) - c$, therefore h_q^{+1} is not a terminal history. □

Proof of Lemma 2

Let $h \in H^{\sigma_{L\to R}}$ be a non-terminal history and $q \in Q_h$.

Base Case: $N(c, l_h) = 1$. Because h is non-terminal, $S_h = [a, 1]$ for some $a \in (0, 1]$, so $\sigma_{L\to R}(h) = a + \frac{l_h}{2}$. The threshold $\phi(l_h)$ is $1 - \frac{l_h}{2}$. There is no feasible $\tilde{q} \in Q_h$ such that $\tilde{q}(\sigma_{L\to R}(h)) < \phi(l_h)$, because such a $\tilde{q}$ does not attain a value of 1 anywhere. Therefore, $q(\sigma_{L\to R}(h)) \geq \phi(l_h)$, so $\sigma_{L\to R}$ stops, so $\phi(l_h) - c \leq V_q^{\sigma_{L\to R}}(h)$.

Induction Hypothesis: Suppose that for any $h' \in H^{\sigma_{L\to R}}$ such that $N(c, l_{h'}) \leq n - 1$ and for all $q \in \Omega_{h'}$,

$$\phi(l_{h'}) - c \leq V_q^{\sigma_{L\to R}}(h').$$

Inductive Step: $N(c, l_h) = n$. The case where $q(\sigma_{L\to R}(h)) \leq \phi(l_h)$ is trivial, so suppose that $q(\sigma_{L\to R}(h)) < \phi(l_h)$. Then,

$$S_{h_q^{+1}} = [\sigma_{L\to R}(h) + 1 - q(\sigma_{L\to R}(h)), 1].$$

Then

$$\begin{aligned} l_{h_q^{+1}} = q(\sigma_{L\to R}(h)) - \sigma_{L\to R}(h) &< \phi(l_h) - \sigma_{L\to R}(h) \\ &= \phi(l_h) - (1 - l_h + 1 - \phi(l_h)) \end{aligned}$$

$$
\begin{aligned}
&= 2\phi(l_h) + l_h - 2 \\
&= \frac{n-1}{n}(l_h - nc) \\
&< n(n-1)c,
\end{aligned}
$$

where the last inequality is because $l_h < n(n+1)c$ by Equation 1. This implies that $N(c, l_{h^{+1}}(q)) = n-k$, for some $1 \le k < n$. By the inductive hypothesis,

$$V_q^{\sigma_{L\to R}}(h_q^{+1}) \ge \phi(l_{h_q^{+1}}) - c.$$

Moreover, $l_{h_q^{+1}} < (n-k)(n-k+1)c$ by Equation 1. Because ϕ is decreasing,

$$
\begin{aligned}
\phi(l_{h_q^{+1}}) - c \ge \phi\Big((n-k)(n-k+1)c\Big) - c &= 1 - \frac{(n-k)(n-k+1)c}{2(n-k)} - \frac{(n-k-1)}{2}c - c \\
&= 1 - (n-k)c - c
\end{aligned}
$$

Finally, because $l_h \ge n(n-1)c$ by Equation 1,

$$\phi(l_h) \le 1 - \frac{n(n-1)}{2n}c - \frac{n-1}{2}c = 1 - (n-1)c \le 1 - (n-k)c - c.$$

Therefore, $V_q^{\sigma_{L\to R}}(h_q^{+1}) \ge \phi(l_h)$. Because h is non-terminal, $V_q^{\sigma_{L\to R}}(h) = V_q^{\sigma_{L\to R}}(h_q^{+1}) - c$, so

$$V_q^{\sigma_{L\to R}}(h) \ge \phi(l_h) - c.$$

□

Proof of Lemma 3

Fixing a σ,[6] we construct a following quality index[7] q' that satisfies the following:

$$q'(\sigma(h_q^{+k})) = \min\left\{\phi(l_h) + kc, \max_{q \in Q_{h_{q'}^{+k}}} q(\sigma(h_{q'}^{+k}))\right\}. \tag{3}$$

If the searcher follows σ and the true quality index is q' then $V_{q'}^{\sigma}(h) < \phi(l_h) - c$. This follows immediately from the definition of q': for all $k \ge 1$ the best item at $h_{q'}^{+k}$ has quality at most $\phi(l_h) + (k-1)c$, but the searcher had to pay kc to get to history $h_{q'}^{+k}$. Her continuation payoff

[6]We omit the dependence on σ of the histories h_q^{+k} for the rest of the proof, to declutter notation

[7]The reader will notice that we are characterizing an equivalence class of quality indices, pinned down by the qualities of the items searched by the searcher on path.

at history $h_{q'}^{+k}$ is therefore at most

$$\phi(l_h) + (k-1)c - kc = \phi(l_h) - c.$$

Before we conclude, we must prove that the quality index q' defined in Equation (3) is feasible at every history, i.e. $q' \in Q_{h_{q'}^{+k}}$ for every k.

First, feasibility requires that q' is Lipschitz continuous at every history. The minimum in the defining equation of q' ensures that $q'(\sigma(h_q^{+k}))$ is always below the upperbound generated by the Lipschitz constraints. Since the upperbound is above the lowerbound and since $\phi(l_h) + kc$ is by definition larger than previous discoveries, the lowerbound is below $q'(\sigma(h_{q'}^{+k}))$.

Second, feasibility requires that, at any history $h_{q'}^{+k}$ there always exists an item x such that $q'(x) = 1$. Let $n = N(c, l_h)$. For $k > n$, the quality of item $\sigma(h_{q'}^{+k})$ is

$$\begin{aligned}
\phi(l_h) + kc > \phi(l_h) + nc &= 1 - \frac{l_h}{2n} - \frac{n-1}{2}c + nc \\
&> 1 - \frac{n(n+1)}{2n} - \frac{n-1}{2}c + nc = 1
\end{aligned}$$

if it is feasible, and the largest possible quality otherwise. Thus, the quality index q' will always achieve quality 1 for some item from the $n+1$ search onward. We prove this also for the first n searches.

Recall that the search window S_h is the set of items x such that there exists a consistent quality index $q \in Q_h$ that attains the quality standard at x. We can reinterpret the search window as the set of items where the quality index q' could attain the quality standard. We then need to show that the search window is non-empty at any history $h_{q'}^{+k}$. Because the search window can only shrink after a search, it suffices to show that $S_{h_{q'}^{+n}}$ is non-empty, or equivalently, $l_{h_{q'}^{+n}} = \mu(S_{h_{q'}^{+n}}) \neq 0$. The Lebesgue measure of the initial search window is $\mu(S_h) = l_h$. Each successive search reduces the search window by a measure at most equal to its informativeness. That is, the k-th search at most eliminates from the search window a set as large as a ball of radius $1 - q'(\sigma(h_{q'}^{+k}))$. The measure of the search window $S_{h_{q'}^{+n}}$ must then be

$$\begin{aligned}
\mu(S_{h_{q'}^{+n}}) \geq \mu\Big(S_h \setminus \cup_{k=0}^{n-1} B(h_{q'}^{+k})\Big) &\geq \mu(S_h) - \sum_{k=0}^{n-1} 2\Big(1 - q'(\sigma(h_{q'}^{+k}))\Big) \\
&\geq l_h - \sum_{k=0}^{n-1} 2\Big(1 - \phi(l_h) - kc\Big)
\end{aligned}$$

$$= l_h - \sum_{k=0}^{n-1} 2\left(\frac{l_h}{n} + \frac{n-1}{2}c - kc\right)$$
$$= l_h - l_h + \frac{n(n-1)}{2}c - \frac{n(n-1)}{2}c = 0,$$

where the informativeness upperbound generates the first two inequalities and the definition of q' in Equation (3) generates the third. The search window at history $h_{q'}^{+n}$ has always weakly positive measure. If the measure is strictly positive, we conclude. Otherwise, note that since each ball $B(h_{q'}^{+k})$ is an open set, the search window cannot be empty, but must contain at least an isolated point x, where $q'(x) = 1$. This concludes the proof. □

B Technical Lemmas

Lemma 4. *The threshold $\phi(l)$ is always bounded above by* 1.

Proof. From equation 1 we know that $l \geq N(c,l)(N(c,l)-1)c$. Then,

$$\phi(l) = 1 - \frac{l}{2N(c,l)} - \frac{N(c,l)-1}{2}c \leq 1 - \frac{N(c,l)(N(c,l)-1)}{2} - \frac{N(c,l)-1}{2}c$$
$$= 1 - (N(c,l)-1)c \leq 1$$

as long as $N(c,l) > 0$. □