Searching for "Arms": Experimentation with Endogenous Consideration Sets*

Daniel Fershtman† Alessandro Pavan‡

July 2025

Abstract

A decision-maker alternates between exploring alternatives in a consideration set and searching for new ones. The problem admits an index solution, featuring a novel recursive index for the expansion of the consideration set. When the expansion technology is stationary or improving, newly discovered alternatives replace existing ones. When it deteriorates, previously discovered alternatives are revisited, and each expansion is effectively treated as the last. We apply the results to the design of procurement mechanisms in environments where a buyer faces uncertainty about the number, costs, and quality of potential suppliers, and must trade off learning about existing suppliers against soliciting new ones. On the positive side, the analysis shows that firms discovered later in the process enjoy higher expected profits, all else equal, and it identifies primitive conditions under which the buyer becomes more (or less) lenient with incumbents after each expansion. On the normative side, the analysis reveals that low-cost firms contribute to inefficient entrenchment with incumbents, whereas high-cost firms contribute to excessive solicitation.

Keywords: Search, experimentation, learning, endogenous consideration sets, multi-armed bandit problems, optimal procurement, dynamic firms' solicitation.

*The paper supersedes previous versions circulated under the titles "Searching for Arms" and "Sequential Learning with Endogenous Consideration Sets." We are grateful to various participants at conferences and workshops where the paper was presented. Nageeb Ali, Arjada Bardhi, Dirk Bergemann, Eddie Dekel, David Dillenberger, Laura Doval, Kfir Eliaz, Teddy Kim, Stephan Lauermann, Charles Manski, Benny Moldovanu, Xiaosheng Mu, Derek Neal, Michael Ostrovsky, David Pearce, Philip Reny, Andrew Rhodes, Eran Shmaya, Andy Skrzypacz, Rani Spiegler, Bruno Strulovici, Asher Wolinsky, and Jidong Zhou provided very useful comments and suggestions. We also thank Matteo Camboni and Tuval Danenberg for outstanding research assistance. Fershtman gratefully acknowledges funding from ISF grant #1202/20. The usual disclaimer applies.

†Department of Economics, Emory University. Email: danfershtman@gmail.com

‡Department of Economics, Northwestern University. Email: alepavan@northwestern.edu

1 Introduction

Classic models of sequential experimentation or learning typically involve a decision-maker (DM) who explores a fixed set of alternatives with unknown characteristics. In many dynamic environments, however, the set of feasible alternatives is expanded over time as the DM gathers information about the alternatives already in her consideration set (CS).

In this paper, we study the tradeoff between exploring alternatives already in the CS and expanding the CS through costly search. A key distinction between exploration and CS expansion lies in their directness. The DM can "point to" and select a particular alternative within the CS for exploration. In contrast, she cannot target specific alternatives outside the CS for inclusion and evaluation. This inability may stem from inherent randomness in the search process, whereby the alternatives discovered may differ from those the DM intended to find. Alternatively, search may produce batched arrivals, and the expected composition of each batch may shape the DM's expansion decisions. Finally, the DM may be uncertain about the distribution of alternatives outside the CS or the efficacy of the search process itself, and may update her beliefs about the underlying search technology based on past outcomes.

To analyze the tradeoff between exploring alternatives already in the CS and expanding it, we study a generalization of the classic multi-armed bandit problem in which the set of "arms" is *endogenous*. Pulling an arm—i.e., exploring an alternative already in the CS—yields a flow payoff and generates information about the arm's characteristics, such as the distribution from which payoffs are drawn. Uncertainty is resolved gradually, with each arm potentially requiring repeated exploration before its attributes are fully revealed. Expanding the CS—i.e., engaging in search—entails a cost and results in the arrival of a random set of new arms whose characteristics, learned upon arrival, may depend on past search outcomes.

We show that the solution to the problem takes the form of an index policy. Each alternative in the CS is assigned a history-dependent number—its index—which depends only on the state of that alternative. This index coincides with the one introduced by Gittins and Jones' (1974) in the classic bandit problem with a fixed set of arms. In addition, the decision to expand the consideration set (i.e., to search) is also assigned an index, which depends on the state of the search technology. This state summarizes the outcomes of past searches and the DM's beliefs about the stochastic process governing the arrival of new alternatives. Crucially, the search index is independent of the information generated by the exploration of any specific alternative already in the CS. While distinct from the value the decision-maker attaches to expansion, the search index is linked to the expected indexes of the alternatives that future searches may uncover. We provide a recursive characterization of the search index and use it to derive many of our results. The optimal policy prescribes choosing, in every period, the option—either an alternative in the CS or search—with the highest index.

Our environment can be viewed as a special case of the branching bandit problem studied in the operations research literature, where activating certain arms causes them to disappear and generates a set of new arms (see, e.g., Weiss, 1988, Weber, 1992, and Keller and Oldale, 2003). In our setting, the decision to expand the CS corresponds to activating a branching arm that yields negative rewards (i.e., search costs) and brings a stochastic set of new alternatives, drawn from a distribution that depends on past search outcomes. Our proof of indexability departs from previous work by leveraging a novel recursive characterization of the search index. This formulation facilitates computation and reveals key dynamic properties of the exploration-expansion tradeoff with relevance for economic applications.

First, at any point in time, the decision to expand the CS depends on the current set only through (i) the index of the alternative with the highest value, and (ii) the state of the search technology. This is true even though the opportunity cost of search—that is, the value forgone by suspending exploration—depends on the entire history of exploration outcomes and past expansions. Likewise, conditional on forgoing search, the decision of which alternative to explore depends solely on each alternative's individual state and is independent of the search technology—even when newly discovered alternatives may resemble some current alternatives more closely than others.

Second, if the search technology is stationary or improving (in a sense formalized below), alternatives present at the time of expansion are never explored again. In effect, each search replaces the current CS with a new one.

Third, when the search technology deteriorates—e.g., because the DM becomes pessimistic about the likelihood of discovering valuable alternatives—the current set is not discarded but put on hold. Alternatives may be revisited after the expansion, and the decision to expand is made as if no further expansions will take place thereafter.

These properties can be viewed as generalizations of the classic independence of irrelevant alternatives (IIA) property from multi-armed bandit problems. What sets the present setting apart from the standard model enriched with a "meta" arm—representing the collection of alternatives that search brings into the CS—is that evaluating such a meta arm requires solving a nested dynamic problem. Specifically, doing so involves determining not only whether to activate the arm but also how to explore the alternatives it yields over time and how to alternate between exploring them and expanding the CS further. In other words, the selection of the meta search arm also involves the choice of "how to use it" and not merely "for how long to use it". Dynamic problems of this kind, in which arms encapsulate future decision processes, seldom admit an index solution. See also the discussion in Section 4.

The framework developed in this paper applies to a broad class of experimentation and sequential learning problems. As an illustrative application, we study the design of sequential procurement mechanisms. In many settings, buyers face uncertainty not only about the quality

and cost of goods or services, but also about the composition of the supplier pool. A salient example is that of local municipalities that repeatedly procure infrastructure, waste management, and other public services from a shifting set of vendors. These agencies expand their pool of potential suppliers gradually over time—through advertising campaigns, public tenders, or requests for proposals (RFPs)—each of which incurs administrative costs and generates a stochastic inflow of new bidders. Suppliers' costs are revealed at bidding, while service quality is learned through repeated interaction. The key tradeoff faced by the buyer—between experimenting with known suppliers and incurring costs to identify new ones—maps naturally into our multi-armed bandit model with an endogenous CS.

We use the model to characterize the properties of the optimal procurement mechanism in such environments. Assuming that each supplier's cost of service provision is privately known, we derive novel implications for the inefficiencies that arise in the solicitation process under profit maximization—that is, when the DM seeks to maximize buyer surplus rather than total welfare.

On the positive side, we identify primitive conditions under which, following each expansion, the buyer becomes more (or less) lenient toward incumbent suppliers—that is, more (or less) willing to tolerate unsatisfactory performance before initiating further search. This translates into a delay (or acceleration) in the expansion of the CS. We also show that, all else equal—that is, fixing the number of competitors, their costs, and the history of experimentation—suppliers who enter the mechanism at later stages earn higher profits than earlier entrants. In other words, the mere delay in their discovery is good news for new entrants.

On the normative side, we show that all suppliers, regardless of type, contribute to inefficiencies in both solicitation and experimentation, albeit in systematically different ways. Specifically, low-cost firms tend to induce inefficient delays in the expansion of the vendor pool, resulting in excessive entrenchment with incumbents. Formally, these firms are tolerated through an inefficiently large number of unsatisfactory performances before the buyer initiates the search for new suppliers. Conversely, high-cost firms lead to an inefficiently rapid expansion of the CS: they are given too few chances to deliver satisfactory service before the solicitation of new firms. To the best of our knowledge, these normative and positive implications are novel and, in principle, open to empirical investigation.

Our framework accommodates general search (i.e., CS expansion) and exploration technologies. In particular, it allows for gradual resolution of uncertainty regarding the alternatives already in the CS—following the classic structure in Gittins and Jones, 1974—as well as flexible, history-dependent processes governing the discovery of new alternatives. This generality is essential for capturing applications such as the sequential procurement setting discussed above, in which each alternative must be explored over multiple periods, the DM alternates across multiple options, and beliefs about the returns to CS expansion evolve endogenously based on prior search outcomes.

Outline. The rest of the paper is organized as follows. The remainder of this section reviews the relevant literature. Section 2 presents the model. Section 3 characterizes the optimal policy and highlights key properties of the dynamics of experimentation and CS expansion. Section 4 discusses indexability and considers generalizations beyond those captured in Section 2. Section 5 contains the application to procurement. Section 6 concludes. All proofs are contained in the Appendix. Additional material is provided in an online Supplement. In particular, the Supplement shows how the indexability result extends to certain decision problems in which the DM, in addition to learning about existing alternatives and searching for new ones, can, or must, irreversibly commit to an option, thus ending the exploration process. As an illustration, we show how Theorem 1 and Proposition 1 can be used to solve an extension of Weitzman's (1979) Pandora's box problem, in which the set of boxes is gradually constructed over time via a general search technology.

1.1 Related literature

To the best of our knowledge, the problem analyzed in this paper—where the DM alternates between exploring "arms" already in the CS and stochastically expanding the latter—is novel. As noted above, this setup can be viewed as a special case of the branching problems studied in the operations-research literature (see, e.g., Weiss, 1988, Weber, 1992, and Keller and Oldale, 2003). Our contribution includes a new proof of indexability, based on a novel recursive characterization of the search (i.e., CS expansion) index. This characterization exploits a classification of alternatives into "categories," which summarize all features relevant to the dynamics of the CS expansion technology. This recursive structure not only facilitates computation of the index but also underpins the key results that follow.

This paper is also related to a large body of work on experimentation and sequential learning with an exogenous choice set—see, for example, Weitzman (1979), Austen-Smith and Martinelli (2018), Fudenberg, Strack and Strzalecki (2018), Gossner, Steiner and Stewart (2021), Ke and Villas-Boas (2019), and Ke, Shen, and Villas-Boas (2016).[1] In these papers, the DM acquires costly information about a fixed set of alternatives before stopping and selecting one. Related contributions include Che and Mierendorff (2019), who study optimal sequential allocation of attention to two biased signal sources, and Liang, Mu, and Syrgkanis (2022), who analyze dynamic information acquisition about an unknown Gaussian state. In all of these models, however, the set of alternatives is fixed from the outset. By contrast, in our framework the DM endogenously expands the choice set in response to accumulated information about existing alternatives. Also related are Garfagnini and Strulovici (2016) and Carnehl and Schneider (2023). The former

[1]See also Bergemann and Välimäki (2008) for an overview of economic applications of multi-armed-bandit problems.

studies forward-looking experimentation with endogenous technologies, where trying "radically" new technologies reduces the cost of experimenting with similar ones, effectively expanding the set of affordable options. The latter examines the time-risk tradeoff faced by a DM who chooses between implementing an existing method or developing a new one under a deadline. While these papers also consider environments in which the choice set expands over time, both their modeling approaches and substantive questions differ significantly from those in our analysis.

The application in Section 5 contributes to three strands of the literature: procurement, "bandit auctions," and bidder solicitation. The bandit auction literature has focused on environments where the set of bidders—that is, the CS—is fixed over time; see, for example, Bergemann and Välimäki (2010), Kakade, Lobel, and Nazerzadeh (2013), Pavan, Segal, and Toikka (2014), and Fershtman and Pavan (2017). In contrast, the auction literature on bidder solicitation has considered static environments with a single solicitation decision, and has abstracted from experimentation—for recent contributions, see Lauermann and Wolinsky (2017), Lauermann and Wolinsky (2025), and references therein. Our analysis contributes a novel angle to these three literatures by allowing the designer to endogenously expand the pool of bidders over time, based on the bids and experimentation outcomes of firms already participating in the mechanism. The resulting dynamics uncover novel positive and normative insights bridging those in these literatures.

The Pandora's boxes problem, as introduced by Weitzman (1979), is a special case of the multi-armed bandit framework characterized by the immediate resolution of uncertainty. In this sense, it can be viewed as a particular instance of the broader class of bandit problems with gradual learning, as studied by Gittins and Jones (1974). Despite its many applications, the Pandora's boxes framework has seen relatively few extensions in the literature. Notable exceptions include Olszewski and Weber (2015), Choi and Smith (2016), and Doval (2018)—all of which assume a fixed set of boxes. A related contribution in the marketing literature is Greminger (2022), who examines a consumer search problem with an endogenous product set (i.e., a dynamic set of boxes), focusing on the comparison between directed and undirected search strategies.

Finally, this paper contributes to the rapidly growing literature on consideration sets (CS). Eliaz and Spiegler (2011) examine how different (exogenously given) CS influence firm behavior. By contrast, Manzini and Mariotti (2014) and Masatlioglu, Nakajima, and Ozbay (2012) develop methods to infer CS from observed choice behavior. Caplin, Dean, and Leahy (2019) provide necessary and sufficient conditions under which rationally inattentive agents optimally limit attention to a subset of available options, thereby endogenizing the CS. Earlier, Simon (1955) proposed a model of sequential search in which alternatives are evaluated until one meets a satisfactory threshold, a behavior that Caplin, Dean, and Martin (2011) later showed to be optimal under information costs. Our analysis complements this body of work by offering a dy-

namic microfoundation for an endogenous CS. Rather than selecting a CS upfront and evaluating its elements, the DM gradually expands the CS in response to the exploration outcomes of its current members.

2 Model

In each period $t = 0, 1, 2, ...$, the DM chooses between exploring one of the alternatives in the CS and expanding the latter by searching for additional alternatives. Exploring an alternative generates information about it and yields a (possibly negative) flow payoff. Expanding the CS yields a stochastic set of new alternatives, which are added to the CS and can be explored in subsequent periods.

Consideration sets. Denote by $C_t \equiv (0, ..., n_t)$ the period-t CS, with $n_t \in \mathbb{N}$.[2] C_t comprises all alternatives $i = 0, ..., n_t$ that the DM can explore in period t, with the initial set $C_0 \equiv (0, ..., n_0)$ specified exogenously and with alternative 0 corresponding to the selection of the DM's outside option, yielding a payoff equal to zero. Given C_t, expansion of the CS in period t (that is, search) brings a set of new alternatives $C_{t+1} \backslash C_t = (n_t + 1, ..., n_{t+1})$ which are added to the current CS and expand the latter from C_t to C_{t+1}.

Alternatives, categories, learning, and payoffs. Each alternative belongs to a fixed *category* $\xi \in \Xi$ that is observed by the DM when the alternative is brought to the CS, with the set Ξ measurable but not necessarily finite. A category contains information about an alternative's experimentation technology and payoff process. Let $\mu \in \mathbb{R}$ denote a fixed unknown parameter about the alternative that the DM is learning about, with μ drawn from a distribution Γ_ξ. When the DM explores the alternative, she observes a signal realization about μ. Let $m \in \mathbb{N}$ index the explorations of an alternative, and denote by $\vartheta^{m-1} \equiv (\vartheta_s)_{s=0}^{m-1} \in \mathbb{R}^m$ its history of past signal realizations, with $\vartheta^{m-1} \equiv \emptyset$ for $m = 0$.[3] When the DM explores the alternative for the m-th time, she receives an additional signal ϑ_m about it, drawn from some distribution $G_\xi(\vartheta^{m-1}; \mu)$ and updates her beliefs about μ using Bayes' rule. Importantly, signal realizations are drawn independently across alternatives, given the alternatives' categories. The flow payoff u that the DM obtains from exploring an alternative from category ξ with parameter μ for the m-th time is drawn from a distribution $L_\xi(\vartheta^{m-1}; \mu)$.

The assumption that ξ is observable implies that the distribution Γ_ξ from which μ is drawn is known to the DM after the alternative's category ξ is learned (which occurs at the time the alternative is brought to the CS). Note, however, that the distribution $G_\xi(\vartheta^{m-1}; \mu)$ from which the m-th signal ϑ_m is drawn, as well as the distribution $L_\xi(\vartheta^{m-1}; \mu)$ from which the m-th reward is drawn, are not fully known to the DM because they depend on μ, which is unknown to the

[2] We adopt the convention of letting the set $\mathbb{N}$ include 0.

[3] We assume that μ and and ϑ_s are real numbers just for convenience; the results do no hinge on this assumption. We also label the exploration by $m = 0$ to ease some of the formulas below.

DM.

The search (i.e., CS expansion) technology. When the DM searches for the k-th time, she incurs a cost c_k and discovers alternatives of different categories. Let $E_k = (n_k(\xi) : \xi \in \Xi)$ denote the complete description of the alternatives identified through the k-th search, with $n_k(\xi) \in \mathbb{N}$ representing the number of category-ξ alternatives discovered. Let $(c_k, E_k)_{k=0}^{m-1}$ denote the history of the past $m-1$ search outcomes with $(c_k, E_k)_{k=0}^{m-1} \equiv \emptyset$ for $m = 0$. Given $(c_k, E_k)_{k=0}^{m-1}$, the m-th search outcome (c_m, E_m) is drawn from a distribution $J((c_k, E_k)_{k=0}^{m-1})$ that is independent of calendar time. The dependence of J on the history of past search outcomes allows us to capture, for example, learning about the effectiveness of search (e.g., about the number of potential firms in the market, as in the application in Section 5), as well as changes in the DM's ability to find new alternatives (e.g., because of learning by doing and/or fatigue).

The classification of alternatives into categories allows us to keep track of all relevant information about the evolution of the search technology. In particular, it allows each search outcome (both the search cost and the set of new alternatives identified) to be drawn from a distribution that depends on the composition of the CS while still permitting an index characterization of the optimal policy. In an environment with an exogenous CS, categories play no role and one can simply let each alternative belong to its own category. With an endogenous CS, instead, categories permit us to identify common information among the alternatives in the CS that is responsible for the outcomes of future searches.

Objective. A *policy* χ for the decision problem described above is a rule specifying, for each period t, whether to experiment with one of the alternatives in the CS C_t or expand the latter through search. A policy χ is *optimal* if, after each period t, it maximizes the expected discounted sum $\mathbb{E}^{\chi}\left[\sum_{s=t}^{\infty} \delta^s U_s | \mathcal{S}_t\right]$ of the flow payoffs, where $\delta \in (0,1)$ denotes the discount factor, U_s denotes the flow period-s payoff (with the latter equal to the search cost in case search is conducted in period s), $\mathcal{S}_t$ denotes the state of the problem in period t (the latter specifies, for each alternative in the CS, the history of signals, along with the history of all past search outcomes; see Section 3 for the formal definition) and $\mathbb{E}^{\chi}\left[\cdot | \mathcal{S}_t\right]$ denotes the expectation under the endogenous process for the flow payoffs obtained by starting from the state $\mathcal{S}_t$ and following the policy χ at each period $s \geq t$. To guarantee that the process of the expected payoffs is well behaved, we assume that, for any t, any $\mathcal{S}_t$ and any χ, $\delta^t \mathbb{E}^{\chi}\left[\sum_{s=t}^{\infty} \delta^s U_s | \mathcal{S}_t\right] \to 0$ as $t \to \infty$. This property is immediately satisfied if payoffs and costs are uniformly bounded; its role is to guarantee that the solution to the Bellman equation of the above dynamic program coincides with the true value function.

Remark. The model above describes an infinite-horizon experimentation problem (with endogenous set of alternatives) in which payoffs are accumulated alongside learning. However, flow payoffs and learning need not be intertwined. In Section 4 and in the Supplement we discuss settings in which the DM sequentially decides between learning about alternatives in the CS and expanding the CS, until a final choice is made among the alternatives in the CS, ending the

decision problem. For example, Theorem 1 can be used to characterize the optimal policy in a generalization of Weitzman's Pandora's boxes problem where the set of boxes is endogenized through the DM's search for new boxes. More generally, the results in Theorem 1 extend to a broader family of problems where the DM needs to irreversibly stop learning in order to be able to accumulate rewards; see Section 4 and the Supplement for further details.

3 Optimal policy and key implications

To facilitate the characterization of the optimal policy, we start by introducing the following notation. Denote by θ a generic sequence of signal realizations about an alternative; that is, θ is given by $\vartheta^{m-1} \equiv (\vartheta_s)_{s=0}^{m-1}$ for some m. Denote by $\omega^P = (\xi, \theta)$ an *alternative's state*, and by Ω^P the set of all possible states of an alternative.[4] While the category ξ is fixed, the history θ of past signal realizations changes over time as the result of the information the DM accumulates about the alternative through past explorations. Similarly, the *state of the search technology* is given by the history of past search outcomes, that is, $\omega^S = (c_k, E_k)_{k=0}^m$ for some m. Denote the set of the possible states of search by Ω^S.

The *state of the decision problem* is given by the pair $\mathcal{S} \equiv (\omega^S, \mathcal{S}^P)$, where $\mathcal{S}^P$ is the *state of the current CS*; formally, $\mathcal{S}^P : \Omega^P \to \mathbb{N}$ is a counting function that specifies for each possible state of an alternative $\omega^P \in \Omega^P$, the number of alternatives in the CS in that state. Let $\Omega \equiv \Omega^P \cup \Omega^S$ and note that $\Omega^P \cap \Omega^S = \emptyset$. Denote by $\mathcal{S}_t$ the state of the decision problem at the beginning of period t. This representation of the decision problem keeps track of all relevant information in a parsimonious way and, as will become clear below, greatly facilitates the analysis.

Remark. The time-varying component θ of each alternative's state $\omega^P = (\xi, \theta)$ admits interpretations other than the signals about a fixed unknown parameter μ. In particular, all of our results apply to a broader class of problems where θ evolves as the result of "shocks" that need not reflect the accumulation of information. For example, such shocks may reflect endogenous variations in preferences, as in certain habit-formation or learning-by-doing models. Furthermore, because no assumptions are made on the distributions $L_\xi(\vartheta^{m-1}; \mu)$ and $G_\xi(\vartheta^{m-1}; \mu)$ from which the payoffs and the signals are drawn, the analysis accommodates for cases where payoffs themselves carry information, as well as cases where information arrives without any accompanying rewards.

3.1 Optimal policy

We now characterize the optimal policy and discuss its implications for the dynamics of experimentation and CS expansion. Recall that a policy χ for the decision problem above specifies,

[4] The initial state of each alternative from category ξ, before the DM explores it, is $(\xi, \emptyset)$. The superscript P in ω^P is meant to highlight the fact that this is the state of a "physical" alternative in the CS, not the state of the search technology, or the overall state of the decision problem, defined below.

for each period t and each period-t state $\mathcal{S}_t$, whether to experiment with one of the alternatives in the CS or expand the latter through search. Clearly, because the entire decision problem is time-homogeneous (independent of calendar time), so is the optimal policy. That is, for any two periods t and t' such that $\mathcal{S}_t = \mathcal{S}_{t'}$, the decisions specified by the optimal policy for the two periods are the same.

For each state ω^P of an alternative, let

$$\mathcal{I}^P(\omega^P) \equiv \sup_{\tau>0} \frac{\mathbb{E}\left[\sum_{s=0}^{\tau-1} \delta^s u_s | \omega^P\right]}{\mathbb{E}\left[\sum_{s=0}^{\tau-1} \delta^s | \omega^P\right]}, \tag{1}$$

denote the "index" of each alternative in the CS in state ω^P. The definition in (1) is equivalent to the one in Gittins and Jones (1974). The expectations in (1) are under the process obtained by selecting the given alternative in all periods. The process τ in (1) is a stopping time (that is, a rule prescribing when to stop, as a function of the observed signal realizations). The flow payoff u_s in (1) is the one generated by the s-th exploration and, depending on the application, can be either positive or negative. As is well known, the optimal stopping rule in the definition of the index is the first period (after the one at which the index is computed) at which the index falls weakly below the value at the time the index was computed (see, e.g., Mandelbaum, 1986).

Given each state $\mathcal{S} = (\omega^S, \mathcal{S}^P)$ of the decision problem, let

$$\mathcal{I}^*(\mathcal{S}^P) \equiv \max_{\omega^P \in \{\hat{\omega}^P \in \Omega^P : \mathcal{S}^P(\hat{\omega}^P) > 0\}} \mathcal{I}(\omega^P).$$

denote the maximal index among the alternatives within the CS.

We now define an index for search (i.e., expansion of the CS). Analogously to the indexes defined above, the index for search is defined as the maximal expected average discounted net payoff, per unit of expected discounted time, obtained between the current period and an optimal stopping time. Contrary to the standard indexes, however, the maximization is not just over the stopping time, but also over the rule governing the selection among the new alternatives brought to the CS by the current and further searches as well as the decision of when to further expand the CS. Denote by τ a stopping time, and by π a rule prescribing, for any period s between the current one and the stopping time τ, either the selection of one of the new alternatives brought to the CS by search or further search. Importantly, π selects only among search and alternatives that are not already in the CS when the decision to search is made.[5]

[5]Suppose the index for search is computed in period t when the state of the search technology is ω^S. Then, for each period $t < s < \tau$, π selects between further search and the selection of alternatives in the CS at period s that were not in the CS in period t.

Formally, given the state of the search technology ω^S, the index for search is defined by

$$\mathcal{I}^S(\omega^S) \equiv \sup_{\pi,\tau} \frac{\mathbb{E}^\pi \left[\sum_{s=0}^{\tau-1} \delta^s U_s | \omega^S\right]}{\mathbb{E}^\pi \left[\sum_{s=0}^{\tau-1} \delta^s | \omega^S\right]}, \tag{2}$$

where U_s denotes the flow payoff from the s-th decision taken under the rule π (with this decision taking the form of further search – in which case U_s is the stochastic cost of search – or exploration of one of the alternatives brought to the CS by searches following the one for which the index is computed, in which case U_s is the stochastic payoff associated with the exploration of the alternative), and where the expectations are under the process generated by the rule π.

Definition 1 (**Index policy**). *The **index policy** χ^* selects at each period t the option with the greatest index given the overall state $\mathcal{S}_t = (\omega^S, \mathcal{S}^P)$ of the decision problem: search if $\mathcal{I}^S(\omega^S) \geq \mathcal{I}^*(\mathcal{S}^P)$, and an arbitrary alternative with index $\mathcal{I}^*(\mathcal{S}^P)$ if $\mathcal{I}^S(\omega^S) < \mathcal{I}^*(\mathcal{S}^P)$.*

Ties between alternatives are broken arbitrarily. In order to maintain consistency throughout the analysis, we assume that, when $\mathcal{I}^S(\omega^S) = \mathcal{I}^*(\mathcal{S}^P)$, search is carried out. To characterize the optimal policy, we first introduce the following notation. For any $v \in \mathbb{R}$, let $\kappa(v) \in \mathbb{N} \cup \{\infty\}$ denote the first time at which, when the DM follows the index policy χ^*, (a) the search technology reaches a state in which its index is no greater than v, and (b) *all* alternatives in the CS – regardless of when they were introduced into it – have an index no greater than v. That is, $\kappa(v)$ is the minimal number of periods until all indexes are weakly below v ($\kappa(v) = \infty$ if this event never occurs).[6]

Let $\mathcal{V}^*(\mathcal{S}_0) \equiv (1-\delta) \sup_\chi \mathbb{E}^\chi \left[\sum_{t=0}^{\infty} \delta^t U_t | \mathcal{S}_0\right]$ denote the supremum expected per-period payoff the DM can attain across all feasible policies χ, given the initial state $\mathcal{S}_0$.

Theorem 1 (**Optimal policy**).

1. *The index policy χ^* is optimal in the sequential experimentation problem with endogenous CS described above.*

2. *The index for search (i.e., CS expansion), as defined in (2), satisfies the following recursive characterization. For any $\omega^S \in \Omega^S$,*

$$\mathcal{I}^S(\omega^S) = \frac{\mathbb{E}^{\chi^*} \left[\sum_{s=0}^{\tau^*-1} \delta^s U_s | \omega^S\right]}{\mathbb{E}^{\chi^*} \left[\sum_{s=0}^{\tau^*-1} \delta^s | \omega^S\right]}, \tag{3}$$

   where τ^* is the first time (strictly after the one at which the index is computed) at which $\mathcal{I}^S$ and the indexes of all the alternatives brought to the CS by the current and future searches

[6] Note that between the current period and the first period at which all indexes are weakly below v, if the DM searches, new alternatives are added to the CS, in which case the evolution of their indexes is also taken into account in the calculation of $\kappa(v)$.

fall weakly below the value $\mathcal{I}^S(\omega^S)$ of the search index at the time the latter is computed, and where all the expectations are under the process induced by the index policy χ^*.

3. *The DM's expected (per-period) payoff under the index policy χ^* is equal to*

$$\int_0^\infty \left(1 - \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|\mathcal{S}_0\right]\right) dv. \tag{4}$$

As in the classic multi-armed bandit problem with an exogenous CS, independence across alternatives is the key assumption behind the optimality of the index policy. That is, the payoffs (and the signals) from the various alternatives are drawn independently across the alternatives, given the latter's categories, and the set of new alternatives brought to the CS at each expansion is invariant in the signals generated by the past explorations of the individual alternatives in the CS. Under such assumptions, the theorem establishes a generalization of the index Theorem, according to which selecting in each period the alternative, or search, with the highest index is optimal.

Part (2) characterizes the stopping time in the index of search by exploiting its recursive formulation. It uses the fact that the rule (for the selection of the alternatives brought to the CS by current and future searches and for the timing of future CS expansions) maximizing the expected payoff per unit of expected discounted time starting from the state ω^S is itself an index rule. This implies that the optimal stopping in the definition of the search index, starting from ω^S, occurs the first moment at which the index of each new alternative brought to the CS by current and future searches as well as the value of the search index itself fall weakly below the value of the search index $\mathcal{I}^S(\omega^S)$ at the time the latter is computed (i.e., before launching the current search, starting from ω^S). Such a recursive representation, the validity of which we establish in the proof of Theorem 1 in the Appendix, facilitates an explicit characterization of the index in applications, and permits us to identify various properties of the dynamics of experimentation and CS expansion that are useful in applications and for our proof of indexability.

Finally, Part (3) offers a convenient representation of the DM's payoff under the optimal rule that can be used, among other things, to determine the DM's willingness to pay for changes in the search technology with limited knowledge about the details of the environment (see also the discussion in the next subsection). Because all indexes represent expected payoffs per expected discounted unit of time, the integral (over all values v) of the time it takes for all indexes to fall below v is a concise statistic of all the signal and reward processes responsible for the DM's payoff under the optimal rule.

3.2 Implications for dynamics of exploration and CS expansion

We first describe a few properties that the search technology may satisfy.

Definition 2 (**Search technology**). *(i) A search technology is* ***stationary*** *if, given any two states of the search technology* $\omega^S = (c_j, E_j)_{j=0}^m$ *and* $\hat{\omega}^S = (\hat{c}_j, \hat{E}_j)_{j=0}^{\hat{m}}$, $J(\omega^S) = J(\hat{\omega}^S)$. *(ii) A search technology is* ***deteriorating*** *if, given any state* $\omega^S = (c_j, E_j)_{j=0}^m$ *and subsequent state* $\hat{\omega}^S = \left((c_j, E_j)_{j=0}^m, (c_j, E_j)_{j=m+1}^{m+s}\right)$, $m, s \in \mathbb{N}$, *the distribution* $J(\omega^S)$ *first-order stochastically dominates the distribution* $J(\hat{\omega}^S)$. *(iii) A search technology is* ***improving*** *if, for any state* ω^S *and subsequent state* $\hat{\omega}^S$, *as defined in Part (ii),* $J(\hat{\omega}^S)$ *first-order stochastically dominates* $J(\omega^S)$.[7]

The next result uses the recursive characterization in Theorem 1 to identify various properties of the dynamics of exploration and CS expansion, which are useful in applications.

Proposition 1 (**Dynamics of exploration and CS expansion**).

1. ***Invariance of expansion to CS composition****: At any period, the decision to expand the CS depends on the state* $\mathcal{S} = (\omega^S, \mathcal{S}^P)$ *of the system only through the value* $\mathcal{I}^*(\mathcal{S}^P)$ *of the alternative with the highest index, and the state* ω^S *of the search technology.*

2. ***Independence of irrelevant alternatives****: At any period* t, *for any pair of alternatives* $i, j \in C_t$ *with* $i \neq j$, *the choice between exploring alternative* i *or exploring alternative* j *is invariant to the period-*t *state* ω^S *of the search technology.*

3. ***Possible irrelevance of improvements in search technology****: An improvement in the search technology increasing the probability of finding alternatives of positive expected value (vis-a-vis the outside option) need not affect the decision to expand the CS even at histories at which, prior to the improvement, the DM is indifferent between expanding the CS and exploring one of the alternatives already in it.*

4. ***Stationary value function****: If the search technology is stationary, for any two states* $\mathcal{S}$, $\mathcal{S}'$ *at which the DM expands the CS,* $\mathcal{V}^*(\mathcal{S}) = \mathcal{V}^*(\mathcal{S}')$.

5. ***Stationary replacement****: If the search technology is stationary or improving and search is carried out at period* t, *without loss of optimality, the DM never comes back to any alternative in the CS at period* t.

6. ***Single search ahead****: If the search technology is stationary or deteriorating, at any history, the decision to expand the CS is the same as in a fictitious environment in which the DM expects she will have only one further opportunity to search.*

[7] That is, the search technology is deteriorating if, regardless of the outcome of past searches, for any k and any upper set $A \subset \mathbb{R} \times \mathbb{N}^{|\Xi|}$ (that is, any set $A \subset \mathbb{R} \times \mathbb{N}^{|\Xi|}$ such that for each $a_1, a_2 \in \mathbb{R} \times \mathbb{N}^{|\Xi|}$ with $a_2 \geq a_1$, $a_2 \in A$ if $a_1 \in A$), one has that $\Pr((-c_{k+1}, E_{k+1}) \in A) \leq \Pr((-c_k, E_k) \in A)$. This definition is quite strong. In more specific environments, where there is an order on the set of categories Ξ, weaker definitions are consistent with the results below.

7. ***Pricing formula****: Consider two states $\mathcal{S}_0 = (\mathcal{S}^P, \omega^S)$ and $\hat{\mathcal{S}}_0 = (\mathcal{S}^P, \hat{\omega}^S)$ that differ only in terms of the state of the search technology. The DM's willingness-to-pay to change the state of the search technology from ω^S to $\hat{\omega}^S$ is equal to*

$$\mathcal{P}^*(\mathcal{S}^P, \omega^S, \hat{\omega}^S) = \int_0^\infty \left(\mathbb{E}\left[\delta^{\kappa(v)} | (\mathcal{S}^P, \hat{\omega}^S)\right] - \mathbb{E}\left[\delta^{\kappa(v)} | (\mathcal{S}^P, \omega^S)\right] \right) dv.$$

Part (1) of the proposition is an implication of Theorem 1. The result is not trivial because the *opportunity cost* of expanding the CS (i.e., the value of continuing with the current CS) typically depends on the entire state $\mathcal{S}$ of the problem, beyond the information contained in $\mathcal{I}^*(\mathcal{S}^P)$ and ω^S.

Part (2) also follows from Theorem 1. Starting with each period t, the *relative* amount of time the DM spends on each pair of alternatives in the period-t CS is invariant to the type of alternatives the DM expects to find by expanding the CS. This is true despite the fact that further expansions of the CS may bring alternatives that are more similar to one alternative than the other.

Part (3) follows from the fact that improvements in the search technology need not imply an increase in the index of search. This is because, as shown in Part (2) of Theorem 1, the optimal stopping time in the index of search is the first time at which the index of search and the indexes of all alternatives brought to the CS by the current and future searches fall weakly below the value of the search index at the time the current search is launched. As a result, any improvement in the search technology affecting only those alternatives whose index at the time of arrival is below the value of the search index at the time search is launched does not affect the value of the search index, and hence the decision to expand the CS. This is true even if these alternatives are explored with positive probability under the optimal rule.

Part (4) of the proposition says that, when the search technology is stationary, the continuation value when search is launched is invariant to the state of the system $\mathcal{S}$ at the time search is launched. This follows from the fact that, without loss of optimality, the DM never comes back to any alternative in the CS after search is launched and the outcome of any future search is invariant in $\mathcal{S}$.

For Part (5), note that, because the state of an alternative changes only when the DM selects it, if, in period t, $\mathcal{I}^S(\omega^S) \geq \mathcal{I}^*(\mathcal{S}^P)$, under a stationary or improving search technology, the same inequality remains true in all subsequent periods. Hence, in this case, search corresponds to disposal of all alternatives in the current CS. Each time the DM searches, she starts fresh.

Part (6) follows again from the recursive characterization of the stopping time in the index of search, as per Part (2) of Theorem 1. Recall that this time coincides with the first time at which the index of any physical alternative brought to the CS by the current or future searches, and the index of search itself, drop weakly below the value of the search index at the time the

current search is launched. If the search technology is stationary, or deteriorating, the index of search falls (weakly) below its current value immediately after search is launched. Hence, $\mathcal{I}^S(\omega^S)$ is invariant to the outcome of any search following the current one, conditional on ω^S.

Part (7) follows from Part (3) in Theorem 1, and can be used to price changes in the search technology, with limited knowledge about the details of the environment. To see this, suppose that the econometrician, the analyst, or a search engine, have enough data about the average time it takes for an agent with an exogenous outside option equal to $v \in \mathbb{R}_+$ to exit and take the outside option, under different search technologies. Then by integrating over the relevant values of the outside option one can compute the maximal price $\mathcal{P}^*(\mathcal{S}^P, \omega^S, \hat{\omega}^S)$ that the DM is willing to pay to change the state of the search technology from ω^S to $\hat{\omega}^S$.

4 Discussion

Indexability and "meta-arms." Indexability of the optimal policy is not obvious in our setting because search acts as a meta-arm—an action that opens up multiple downstream possibilities (over and above the length of utilization). Specifically, choosing to search not only involves selecting an arm but also specifying how to use it—whether to explore one of the physical alternatives it brings into the CS or to continue searching. While our results establish that search can effectively be treated as a meta-arm endowed with its own index, this outcome is not a priori guaranteed. Indeed, in many problems where actions correspond to meta-arms—that is, subproblems with their own internal decisions (sometimes called super-processes)—an index solution generally does not exist, even when subproblems are independent and their solutions known. Similarly, correlation or dependence among alternatives often precludes indexability, even when subsets evolve independently of one another and optimal within-subset decision rules are known.

To see why multi-armed bandit problems in which alternatives take the form of meta-arms typically do not admit an index solution, consider the following extension of the environment described in Section 2. Suppose there are $k \in \mathbb{N}$ distinct sets of arms, $K_1, ..., K_k$. Arms in different sets evolve independently, but arms within a given set may have interdependent states. More generally, suppose each arm is a meta-arm: activating it triggers a distinct decision process that involves not just timing (i.e., when to stop), but also how to act across multiple stages. Assume that each meta-arm evolves independently of the others and that the optimal decision rule for each meta-arm can be computed in isolation.

It might then be tempting to conjecture that one could assign an independent index to each meta-arm (or to each arm set K_i) and solve the overall problem via an index policy—that is, by selecting the meta-arm with the highest index in each period. However, this intuition fails. Even when arms (or sets of arms) are independent in the standard probabilistic sense, the internal complexity of the decision process within each meta-arm can prevent the existence of a global

index policy. In other words, independence across meta-arms is not sufficient for indexability when the arms themselves are structured decision problems. The following example illustrates this failure.

Example. There are two arms. Arm 1 yields a reward of 1,000 when pulled for the first time. In all subsequent pulls, it yields a reward of λ, where $\lambda \in \{1, 10\}$ is initially unknown and drawn with equal probability. The value of λ is revealed with certainty after the first pull and remains fixed thereafter.

Arm 2 is a meta-arm that corresponds to a sub-decision problem. When this arm is pulled for the first time, the DM must also choose one of two modes of operation:

- 2(A): The arm yields a reward of 100 for one period only, and then becomes inactive (i.e., no rewards thereafter).
- 2(B): The arm yields a reward of 11 in each period it is pulled, indefinitely.

Crucially, the choice between 2(A) and 2(B) must be made upon the first activation of Arm 2 and is irreversible.

Assume a discount factor $\delta = 0.9$. The optimal policy is as follows:

- In period 1, pull Arm 1.
- If $\lambda = 10$, then pull Arm 2 in version 2(A) for one period (yielding 100), and revert to Arm 1 in all subsequent periods.
- If $\lambda = 1$, then switch to Arm 2 in version 2(B) and continue pulling it indefinitely.

This strategy is optimal but clearly not implementable via an index policy. Under an index policy, each arm is assigned a numerical index that is independent of the state or outcome of other arms, and the highest-indexed arm is pulled in each period. In this example, however, the optimal choice of how to use Arm 2—whether to activate it in mode 2(A) or 2(B)—depends on the outcome of Arm 1's first pull. Thus, the optimal utilization of one arm is contingent on information obtained from another, violating the separability required for indexability.

This example illustrates the broader point: when arms are meta-arms with internal decision processes, even full independence across arms and knowledge of the optimal utilization of each meta-arm in isolation are not sufficient to guarantee that the optimal policy for the entire problem takes the form of an index rule.

Stopping and irreversible choice. In many decision problems, the DM faces not only the task of learning about existing options and searching for new ones, but also the possibility—or necessity—of irreversibly committing to one of the alternatives, thereby ending the exploration process. In general, such problems do not admit an index solution. In the Supplement, we provide a sufficient condition under which the optimality of an index policy extends to these

settings. Specifically, we assume that each alternative of category ξ must be explored at least $M_\xi \geq 0$ times before the DM is allowed to commit to it. For instance, a hiring committee may need to collect a minimum number of positive signals about a candidate's qualifications before making an offer, or a consumer might need to visit a vendor's website at least once to complete a transaction.

The sufficient condition ensures that, once an alternative reaches a state in which commitment is feasible, its retirement value—i.e., the value of irreversibly selecting it—either (i) falls below the value of the outside option (e.g., when a candidate is revealed to be unqualified), or (ii) increases (weakly) with additional exploration. This property is reminiscent of a condition identified in Glazebrook (1979), who establishes the optimality of index policies in a class of bandit problems with stoppable processes. Our proof, however, differs in that it explicitly accounts for the endogenous evolution of the set of alternatives over time.

A special case of the problem described above arises when exploring an alternative reveals its value immediately. In this case, the setting reduces to the well-known Pandora's boxes problem, but with an endogenous (time-varying) set of boxes. We now show how the results from Theorem 1 and Proposition 1 can be used to solve an extension of Weitzman's problem in which the set of boxes evolves over time in response to the outcomes of past explorations. In this extended problem, each alternative corresponds to a "box" that belongs to a category $\xi \in \Xi$. Each category ξ is associated with a pair (F^ξ, λ^ξ), where F^ξ is the distribution from which the box's prize v is drawn, and λ^ξ is the cost of inspecting (i.e., opening) the box. As in Weitzman's original formulation, each box's prize v is drawn independently (conditional on the category) and is revealed upon first inspection.

At each period, the DM can choose one of three actions: (a) search for additional boxes to add to the current CS; (b) open one of the boxes already in the CS to reveal its prize; or (c) stop and either recall the prize from one of the previously opened boxes or take the outside option (whose value is normalized to zero). Either choice in (c) brings the decision process to an end.

For simplicity, assume that each search yields exactly one new box. The category ξ of the box is drawn from the set Ξ according to a distribution $\rho(m) \in \Delta(\Xi)$, which may vary with the number of past searches but is independent of the realizations of those searches. The draws from the distributions $\rho(m)$ are independent across searches.[8]

We assume that $\Xi \subset \mathbb{N}$, with higher ξ denoting superior boxes, in the sense that, for any $\xi', \xi'' \in \Xi$ with $\xi'' > \xi'$, $F^{\xi''} \succeq_{FOSD} F^{\xi'}$ and $\lambda^{\xi''} \leq \lambda^{\xi'}$ (with one of the two relationships strict). Let $\underline{\xi} \equiv \inf \Xi$ and $\overline{\xi} \equiv \sup \Xi$. The cost of the m-th expansion of the CS is $c(m)$, where $c(\cdot)$ is a

[8]Note that, even though each search yields exactly one new arm, the problem cannot be reduced to a standard multi-armed bandit formulation by treating each search as the first pull of a newly arrived arm. The key reason is that the outcome of this initial "pull" depends on the number of arms previously explored, thereby violating a central condition for indexability in the standard bandit framework—namely, that the evolution of each arm be independent of the history of other arms.

positive and increasing function. In addition, we assume that, for all m, $\rho(m) \succeq_{FOSD} \rho(m+1)$; that is, the distribution $\rho(m) \in \Delta(\Xi)$ from which the category of the m-th box is drawn first-order-stochastically dominates, weakly, the distribution $\rho(m+1) \in \Delta(\Xi)$ from which the category of the $(m+1)$-th box is drawn. The combination of the assumption that $c(m)$ is weakly increasing in m and the distribution $\rho(m) \in \Delta(\Xi)$ from which the boxes are drawn "decreases" with m in a FOSD sense implies that the index of search $\mathcal{I}^S(m)$ defined below is decreasing in m and can be characterized using the same recursive properties as when the search technology deteriorates in the sense of Definition 2 (as per Part (6) of Proposition 1). We denote by $\rho^\xi(m)$ the probability that the m-th search brings a ξ-box, with $\sum_{\xi\in\Xi} \rho^\xi(m) = 1$ for all m.[9]

As in the baseline model, the DM discounts the future at rate δ.

The setting described above is one in which the DM's problem terminates upon choosing to walk away with the prize from an opened box or the outside option. By contrast, the framework in Section 2 features an infinite horizon, with the DM continuously choosing among evolving alternatives. Despite this distinction, the optimal solution to the box problem takes the form of an index policy closely related to the one in Definition 1 (see Proposition S.1 and its proof in the Supplement).

The key intuition is that the Pandora's boxes problem with an endogenously expanding set of boxes can be mapped into an auxiliary problem that conforms to the general structure analyzed in Section 1. Theorem 1 and Proposition 1 can then be applied to characterize the structure of the optimal policy, which we describe next.

For any ξ-box that has not been opened yet (i.e., for which $\omega^P = (\xi, \emptyset)$) let the reservation prize $\mathcal{I}^P(\xi, \emptyset)$ be given by the solution to:

$$\mathcal{I}^P(\xi,\emptyset) = \frac{-\lambda^\xi + \delta \int_{\frac{\mathcal{I}^P(\xi,\emptyset)}{1-\delta}}^{\infty} v \mathrm{d}F^\xi(v)}{1 + \frac{\delta}{1-\delta}\left(1 - F^\xi\left(\frac{\mathcal{I}^P(\xi,\emptyset)}{1-\delta}\right)\right)}. \tag{5}$$

For any $l \in \mathbb{R}$, let $\Xi(l) \equiv \left\{\xi \in \Xi : \mathcal{I}^P(\xi,\emptyset) > l\right\}$ denote the set of boxes whose reservation prize exceeds l. For any m, the reservation prize of search $\mathcal{I}^S(m)$ is given by the solution to:[10]

$$\mathcal{I}^S(m) = \frac{-c(m) + \delta \sum_{\xi\in\Xi(\mathcal{I}^S(m))} \rho^\xi(m) \left(-\lambda^\xi + \delta \int_{\frac{\mathcal{I}^S(m)}{1-\delta}}^{\infty} v \mathrm{d}F^\xi(v)\right)}{1 + \sum_{\xi\in\Xi(\mathcal{I}^S(m))} \rho^\xi(m) \left[\delta + \frac{\delta^2}{1-\delta}\left(1 - F^\xi\left(\frac{\mathcal{I}^S(m)}{1-\delta}\right)\right)\right]}. \tag{6}$$

The solution to the Pandora's boxes problem with an endogenously evolving CS takes the following form. If the highest reservation prize among all unopened boxes in the CS exceeds both the reservation prize of search, $\mathcal{I}^S(m)$, and the flow value $(1-\delta)v$ of each opened box (as well as

[9] All the results extend to the case where Ξ is infinite.

[10] Because all the relevant information about the state of the search technology is summarized in the number of past searches, we abuse notation and let $\mathcal{I}^S(m)$ denote the index for the m-th search.

the value of the outside option), then the DM opens one of the unopened boxes with the highest reservation prize. If, instead, the reservation prize of search $\mathcal{I}^S(m)$ exceeds the reservation prize $\mathcal{I}^P(\xi, \emptyset)$ of every unopened box, the flow value $(1-\delta)v$ of each opened box, and the outside option, then the DM chooses to search. If neither of these two cases applies, the DM stops. In that event, she selects the opened box whose flow value $(1-\delta)v$ is highest, provided it exceeds the outside option; otherwise, she takes the outside option.

As in Weitzman's problem, the reservation prizes $\mathcal{I}^P(\xi, \emptyset)$ of the boxes that have not been opened yet have the following interpretation.[11] Suppose there are only two alternatives: one is an unopened ξ-box, and the other is a hypothetical box yielding a constant flow payoff K per period. Then, $\mathcal{I}^P(\xi, \emptyset)$ is the value of K for which the DM is indifferent between choosing the hypothetical annuity (with continuation value $K/(1-\delta)$) and inspecting the unopened ξ-box while retaining the option to recall the annuity after observing the realized prize v.

The reservation prize of search, $\mathcal{I}^S(m)$, extends this logic. Consider again two options: the hypothetical annuity yielding known K, and the opportunity to search for a new box. The reservation prize of search is the value of K for which the DM is indifferent between taking the annuity immediately and performing one additional search, retaining the option to take the annuity later either:

(a) immediately after discovering the category ξ of the newly found box, if $\mathcal{I}^P(\xi, \emptyset) \leq K$, or

(b) after inspecting the box—if $\mathcal{I}^P(\xi, \emptyset) > K$—in the event that the realized prize $v \leq K/(1-\delta)$.

See the Supplement for a more general treatment of irreversible problems with an endogenous set of alternatives for which the optimal policy is indexable.

Relative length of expansion. To incorporate frictions in the search for new alternatives, we assume that each time the DM searches, she forgoes the opportunity to explore any of the alternatives currently in the CS, with search requiring the same amount of time as a single exploration. All our results extend to environments in which the duration of search and exploration varies stochastically with the state. Specifically, they generalize to semi-Markov settings where time is not slotted into discrete periods, and decision times are modeled as random variables. In such a setting, the time required to "pull an arm" can be stochastic, heterogeneous across arms, and different from the time needed to expand the CS. Similarly, the time required for each search may depend on the outcomes of previous searches. These durations must therefore be incorporated into the state variables: for explorations, into each arm's state ω^P; for search, into the state of the search technology ω^S.

Moreover, because the time cost of search can be made arbitrarily small through appropriate

[11] Weitzman defines the reservation prize $\hat{\mathcal{I}}^P(\omega^P)$ for $\omega^P = (\xi, \emptyset)$ as the solution to $\lambda^\xi = \delta \int_{\hat{\mathcal{I}}^P(\omega^P)}^{\infty} (v - \hat{\mathcal{I}}^P(\omega^P)) \mathrm{d}F^\xi(v) - (1-\delta)\hat{\mathcal{I}}^P(\omega^P)$, which yields $\hat{\mathcal{I}}^P(\omega^P) = [-\lambda^\xi + \delta \int_{\hat{\mathcal{I}}^P(\omega^P)}^{\infty} v \mathrm{d}F^\xi(v)]/[1 - F^\xi(\hat{\mathcal{I}}^P(\omega^P))]$. The reservation prizes in (5) are thus equal to those in Weitzman (1979) multiplied by $(1-\delta)$; that is, $\mathcal{I}^P(\omega^P) = (1-\delta)\hat{\mathcal{I}}^P(\omega^P)$.

rescaling of payoffs and adjustment of the discount factor, the framework also accommodates environments in which search and learning proceed almost in parallel. That is, our results continue to apply even when search causes only negligible delays in the exploration of existing alternatives so that search and learning occur "almost" in parallel.

No discounting. The results above assume that $\delta < 1$. However, they extend to $\delta = 1$ (i.e., no discounting). As noted in Olszewski and Weber (2015), bandit problems in which $\delta = 1$ can be thought of as problems with non-discounted "target processes" where arms reaching a certain (target) state stop delivering payoffs. A well-known result for such problems is that the finiteness they impose allows one to take the limit as $\delta \to 1$ (see e.g., Dumitriu, Tetali, and Winkler, 2003).

General "states" and multiple search options. The general model presented in Section 2 is highly flexible. As noted earlier, the results in Theorem 1 and Proposition 1 also apply to environments in which the state of each alternative evolves due to factors beyond information acquisition—for instance, through exogenous shocks such as firms improving their products over time. Moreover, the framework accommodates settings in which the DM has access to multiple search opportunities and can choose where and how to expand the CS. The results in Theorem 1 and Proposition 1 extend to such settings provided that each "search arm" brings forth a set of alternatives independently of all other search arms.

The results also extend to settings in which search arms function as "meta-arms" that do not directly introduce new alternatives, but instead reveal information relevant to an entire group of alternatives—information that is independent of that revealed by other meta-arms for other groups. The proof proceeds along lines similar to those used in establishing Theorem 1. The parallel with the case of multiple independent search arms becomes evident when interpreting the information shared across a group of alternatives as a category, following the classification introduced in Section 2. Under this interpretation, the problem corresponds to one in which the CS is initially empty, and each independent meta-arm, when activated, yields a fixed set of alternatives that share a common category, which is revealed at the time of activation.

5 Application: Experimentation and Solicitation in Sequential Procurement Auctions

In many procurement settings, buyers face uncertainty not only about the quality and cost of services offered by different suppliers but also about the set of firms in the market that could potentially provide those services. For instance, municipalities and other public agencies regularly procure goods, infrastructure, and services from private companies to meet public needs. These agencies engage in repeated procurement from multiple vendors while also investing in advertising and other solicitation efforts to identify new potential suppliers. The uncertainty surrounding such campaigns arises from limited knowledge about how many firms are active in the market,

whether their services meet the required standards, and the cost of procuring from them.

In this section, we show how the procurement problem—with a set of suppliers endogenously built over time—naturally maps into our model. We then use Theorem 1 and Proposition 1 to analyze the properties of the optimal mechanism.

We assume that each supplier's unit cost of providing the service is privately known to the supplier. This assumption allows us, among other things, to derive novel predictions about the dynamics of inefficiencies in bidder solicitation under profit maximization (i.e., when the designer maximizes buyer surplus rather than total welfare). While, under the optimal mechanism, suppliers' costs are revealed through bidding when they join the mechanism, the quality of each supplier's service is learned by the buyer gradually through experimentation (i.e., through repeated purchases).

In addition to the procurement literature, the analysis contributes to the "bandit auctions" literature by endogenizing the set of bidders, and to the recent literature on bidder solicitation in auctions by incorporating dynamics in which solicitation efforts depend on the outcomes of past campaigns and the performance of previously contracted firms.

While the analysis is motivated by the procurement of goods and services by municipalities, the results apply more broadly to both public and private procurement problems, as well as to the design of screening mechanisms with an endogenous set of agents.

5.1 Environment

A buyer repeatedly procures services from firms that have joined the procurement mechanism—that is, firms that are part of the consideration set (CS). Initially, the CS is empty. In each period, the buyer can either search for new firms (e.g., by conducting an advertising campaign or hiring a consultant to solicit bids), or procure from one of the firms already in the CS.

The buyer does not know how many firms are in the market for the desired service. Let $N \in \mathbb{N}$ denote this number. The buyer initially believes that N is drawn from a distribution with cumulative distribution function (cdf) F and probability density function (pdf) f. That is, for any $N \in \mathbb{N}$, $f(N)$ represents the buyer's belief about the probability that there are exactly N firms in the market that could potentially be brought into the procurement mechanism.

Whenever the buyer searches for new suppliers, she identifies each firm in the market that has not yet been identified with probability $\rho \in (0, 1)$. In other words, if there are N firms in the market and $n \leq N$ have already been identified, then the probability that the m-th search yields $k \in \{0, N - n\}$ new firms is given by

$$b(k; N, n, \rho) \equiv \binom{N-n}{k} \rho^k (1-\rho)^{N-n-k}. \tag{7}$$

Conditional on N, the number of new firms brought into the procurement mechanism as a result of the m-th search depends on the outcomes of past searches only through the total number n of firms identified thus far.

For simplicity, assume that each search entails no direct monetary cost to the buyer, aside from the opportunity cost of pausing procurement for one period.[12] Given that n firms have been identified through past searches, the buyer updates her belief about the remaining number of firms in the market, $N - n$, accordingly.

Each firm incurs a unit cost c each time it supplies its service to the buyer. The service is assumed to be indivisible and can be supplied an unlimited number of times. Each firm's cost c is drawn from a cumulative distribution function G with strictly positive density g over the interval $[\underline{c}, \bar{c}]$. Costs are drawn independently across firms and are independent of all other random variables in the model. Each firm's cost is constant over time and privately known to the firm. We assume that the function G/g is increasing, and we define the "virtual cost" of a firm with true cost c as $\gamma(c) \equiv c + G(c)/g(c)$.

The buyer does not initially know the quality of any firm's service—formally captured by the distribution from which the buyer's flow payoff is drawn—and gradually learns about each firm's quality through experimentation. Firms, in turn, are also unaware of the buyer's valuation for their services. To keep the analysis tractable, we assume that service quality is binary. Let $\mu \in \{\underline{\mu}, \overline{\mu}\}$ denote a firm's quality, where $\overline{\mu}$ corresponds to high quality and $\underline{\mu}$ to low quality. Each firm's quality is drawn independently across firms and independently from all other random variables in the model. Let p denote the ex-ante probability that a firm's quality is high, i.e., $\Pr(\tilde{\mu} = \overline{\mu}) = p$.[13]

Each time the buyer procures the service from a firm, she receives a signal $\vartheta \in \{0, 1\}$ about the firm's quality. For simplicity, we assume that these signals coincide with the buyer's flow payoffs. Specifically, for any $m \in \mathbb{N}$, the probability the m-th signal is equal to 1 is given by $\Pr(\tilde{\vartheta}_m = 1|\tilde{\mu} = \underline{\mu}) = 0$, whereas $\Pr(\tilde{\vartheta}_m = 1|\tilde{\mu} = \overline{\mu}) = q \in (0, 1)$. That is, a low-quality firm never generates a positive signal (equivalently, a satisfactory service), while a high-quality firm does so with probability q. Signals are drawn independently across time and across firms, conditional on firm quality. In other words, the experimentation process follows the "no news is bad news" (NNBN) framework commonly used in the literature, where "news" corresponds to "satisfactory experiences" or, equivalently, "successes." We note that most of the results presented below generalize to other signal structures, as we will explain in due course.

Let $\bar{u}(\vartheta^{m-1})$ denote the buyer's expected flow payoff (gross of any payments) from procuring for the m-th time from a firm with signal history $\vartheta^{m-1} \equiv (\vartheta_s)_{s=0}^{m-1}$. When ϑ^{m-1} is such that $\vartheta_l = 1$ for some $l \leq m - 1$, $\bar{u}(\vartheta^{m-1}) = q$. When, instead, ϑ^{m-1} is such that $\vartheta_l = 0$ for all $l \leq m - 1$, or

[12] This assumption is made solely for expositional convenience and does not play a critical role in the results that follow.

[13] We use tildes to denote random variables.

$m = 0$ (in which case $\vartheta^{m-1} = \emptyset$),

$$\bar{u}(\vartheta^{m-1}) = q \cdot \frac{p(1-q)^m}{p(1-q)^m + 1 - p},$$

where $p(1-q)^m/[p(1-q)^m + 1 - p]$ is the posterior probability that the quality of the firm's service is $\overline{\mu}$ given that the firm never delivered a satisfactory service in the past.

5.2 Profit- and Welfare-Maximizing Mechanisms

Whenever the buyer identifies a new firm, she invites it to join the procurement mechanism and submit a bid. This bid determines the payment the buyer will make to the firm each time its service is procured. Additionally, the buyer commits to paying a lump-sum transfer T to the firm upon joining the mechanism, where T depends on the firm's bid (details follow). This transfer remains fixed regardless of the firm's subsequent utilization. Firms identified in period t submit their bids at the end of that period.

To facilitate the mapping of this application to the general model in the paper, let $\mathcal{S}$ denote the state of the procurement mechanism. This state specifies, for each past search and each bid b, the number of firms that joined the mechanism after that search with bid b, as well as the history of signals $\theta = \vartheta^{m-1}$ for each firm in the mechanism. If a search is conducted at period t, the state $\mathcal{S}$ at the end of period t also includes, for each bid b, the number of new firms identified in period t that submitted bid b.

The buyer aims to maximize either welfare—the expected discounted value of the services procured net of firms' costs—or profits—the expected discounted value of the services procured net of payments made to the firms. We refer to the latter objective as "profits" to facilitate comparison with other screening problems, particularly those studied in the recent literature on bidder solicitation in auctions; see, e.g., Lauermann and Wolinsky (2017) and Lauermann and Wolinsky (2025). In our procurement application, the buyer maximizes profits when acting as a firm procuring inputs for the production of a final good. Conversely, when the buyer represents a local public authority—such as a municipality procuring services for its citizens—it is sufficient to interpret profits as "buyer surplus."

Consistent with the dynamic mechanism design literature, we say that a procurement mechanism is optimal if: (a) it is interim incentive compatible (IC) and individually rational (IR), meaning that, without knowledge of the current state of the mechanism when solicited to join, each firm—regardless of its cost—finds it optimal to join and bid truthfully, assuming that all other firms do the same; and (b) it maximizes the buyer's objective over the entire class of direct revelation mechanisms that satisfy interim IC and IR.

For each of the objectives described above—welfare and profit maximization—the optimal mechanism implements the same allocations as those sustained under an index policy analogous

to the one in Theorem 1. Specifically, suppose each firm bids truthfully (i.e., each firm with cost c submits bid $b = c$ upon joining), and consider the following experimentation problem, which is a special case of those covered by Theorem 1:

1. Each firm corresponds to an "arm" whose category is given by the firm's bid (equivalently, its cost), and whose state $\omega^P(b, \theta)$ combines the firm's bid b with the history of past signal realizations $\theta = \vartheta^{m-1}$;

2. The state ω^S of the search technology specifies, for each previous search and each bid b, the number of firms that joined the mechanism after that search with bid b. Since search incurs no direct cost, we omit the search cost from the description of ω^S.

3. The search technology is characterized by the distribution over the number of new firms added to the mechanism following each expansion (as described above), along with the following two assumptions: (a) the probability that a firm is identified after each expansion is independent of its cost and service quality; and (b) all firms bid truthfully. Note that, unlike in the settings studied in previous sections, each arm's category here is endogenous, corresponding to the firm's exogenous cost under truthful bidding.

4. The flow expected payoff from procuring from a firm in state $\omega^P(b, \theta)$, when $\theta = \vartheta^{m-1}$, is given by $\bar{u}(\vartheta^{m-1}) - b$ under welfare maximization, and by $\bar{u}(\vartheta^{m-1}) - \gamma(b)$ under profit maximization.

Let χ_W and χ_{PR} denote the index policies characterized in Theorem 1 for this experimentation problem, respectively for welfare and profit maximization.

Finally, for each bid b and each state of the mechanism $\mathcal{S}$ (defined exactly as in Section 2), let $\mathcal{S}^\wedge(b)$ denote the state that results from adding to the outcome of the most recent expansion one additional firm that bids b. Then, let $Q_W(b; \mathcal{S}^\wedge(b))$ and $Q_{PR}(b; \mathcal{S}^\wedge(b))$ denote the expected discounted number of times a firm with bid b is utilized (from the moment it joins the mechanism onward), given that the state of the mechanism at the end of the period in which it joins is $\mathcal{S}^\wedge(b)$, respectively under policies χ_W and χ_{PR}.

Lemma 1 below follows from Theorem 1 along with standard properties of screening models with a continuum of types.

Lemma 1. *Under the optimal mechanism, for any $\mathcal{S} \equiv (\omega^S, \mathcal{S}^P)$ and any b, a firm joining the mechanism with bid b receives a lump-sum transfer equal to*

$$T_W(b; \mathcal{S}^\wedge(b)) = K + \int_b^{\bar{c}} Q_W(y; \mathcal{S}^\wedge(y))dy$$

under welfare maximization, and equal to

$$T_{PR}(b; \mathcal{S}^\wedge(b)) = \int_b^{\bar{c}} Q_{PR}(y; \mathcal{S}^\wedge(y))dy$$

under profit maximization, where $K \geq 0$ is an arbitrary constant, and $\mathcal{S}^{\wedge}(b)$ denotes the state of the mechanism at the end of the period in which the firm joins. In each period, the buyer follows the index policy χ_W under welfare maximization and the index policy χ_{PR} under profit maximization. Under both mechanisms, participation followed by truthful bidding is weakly dominant for every firm. Furthermore, without loss of optimality, the buyer may disclose the current state of the mechanism to any identified firm prior to its joining.

5.3 Dynamics under Optimal Mechanisms

We now establish several properties of the procurement dynamics induced by the optimal mechanisms.

5.3.1 Preliminaries

We begin by establishing several useful properties of the search (or expansion) index. Let $\succ$ denote the strict dominance relation under the Monotone Likelihood Ratio order (MLR); that is, $F \succ \hat{F}$ if and only if $f(N)/\hat{f}(N)$ is increasing in N. Let $\mathcal{I}_x^S$ denote the search index, where $x = W$ under welfare maximization and $x = PR$ under profit maximization. Lemma 2 below characterizes key properties of these indexes.

The following definition proves useful:

Definition 3 (Ultra log concavity and convexity)**.** A distribution F is **UL-concave** (respectively, **UL-convex**) if $L(N) \equiv (N+1)f(N+1)/f(N)$ is decreasing (respectively, increasing) in N.

The Poisson distribution provides a useful benchmark. When F is Poisson with parameter λ, the posterior distribution over the total number N of firms in the market conditional on $n \in \mathbb{N}$ firms having been identified through $m \in \mathbb{N}$ past searches is itself Poisson with parameter $\lambda(1-\rho)^m$. This posterior is decreasing in m but invariant in n.[14] Hence, under a Poisson prior, the search index deteriorates deterministically with each additional expansion. For the Poisson distribution, $L(N) = \lambda$. Distributions that are UL-concave (respectively, UL-convex) are more (respectively, less) concentrated around the mode of the probability mass function than a Poisson distribution with the same mean. In the proof of Lemma 2, we show that these properties are preserved by the posteriors over the number of firms yet to be discovered, which in turn has important implications for the behavior of the search index.

[14]This follows from the general property that, when the distribution of the random variable $\tilde{X}$ is Poisson with parameter λ and, conditional on $\tilde{X} = x$, the random variable $\tilde{Y}|x$ is binomial with parameters (x, p), then the unconditional distribution of $\tilde{Y}$ is Poisson with rate λp. In the current model, $\tilde{X}$ is the total number of firms in the market, and $\tilde{Y}$ is the number of firms that remains undiscovered after m searches. The probability that a firm in the market remains undiscovered after m searches is $(1-\rho)^m$. Hence, conditional on $\tilde{X} = N$, the random variable $\tilde{Y}|N$ is binomial with parameters $(N, (1-\rho)^m)$. We conclude that the posterior distribution over the total number of firms in the market that remain to be found when the first m searches identified n firms is invariant in n and is Poisson with parameter $\lambda(1-\rho)^m$.

Lemma 2 (Properties of the search index $\mathcal{I}_x^S$)**.** *The following properties about $\mathcal{I}_x^S$ are true for both $x = W$ and $x = PR$:*

1. *$\mathcal{I}_x^S$ depends on the state ω^S of the search technology only through the number of past searches (m) and the number of firms identified (n);*

2. *$\mathcal{I}_x^S$ is decreasing in m;*

3. *$\mathcal{I}_x^S$ is decreasing (respectively, increasing) in n if F is UL-concave (respectively, UL-convex).*[15]

Part (1) establishes that the search index $\mathcal{I}_x^S$ depends on the history of past search outcomes only through two statistics: the number of searches m conducted thus far, and the total number n of firms identified. With a slight abuse of notation, we will therefore write $\mathcal{I}_x^S(m, n)$ to denote the search index associated with the $(m + 1)$-th search.

Part (2) establishes that, for a given number n of firms identified through past searches, the search index $\mathcal{I}_x^S(m, n)$ is decreasing in m: the more searches it takes to identify n firms, the lower the resulting search index. This monotonicity arises because, holding n fixed, the posterior distribution over the number of remaining firms in the market becomes less optimistic (in the Monotone Likelihood Ratio–MLR–order) as the number of searches m increases.

Part (3) follows from the recursive characterization of the search index provided in Part (2) of Theorem 1, together with the property that, when the distribution F is UL-concave (respectively, UL-convex), a larger number n of firms identified through a fixed number m of searches makes the buyer more pessimistic (respectively, optimistic) about the outcome of the $(m+1)$-th search (i.e., the number of new firms identified, in the MLR order). Namely, the result derives from this comparative statics property together with two additional facts: (a) stopping in the computation of the search index occurs at the first time at which the search index itself and the indexes of all the alternatives brought to the CS by the current and future searches fall weakly below the value of the search index at the time the latter is computed; and (b) the search index equals the expected discounted sum of the flow payoffs generated by the alternatives found through search, normalized by the average discounted time required to realize them, with all the expectations under the process induced by the index policy χ_W (welfare maximization) or χ_{PR} (profit maximization). It is important to note that none of the results in this lemma rely on the assumption that the signal-generating process satisfies the "no news is bad news" (NNBN) property.

[15] If F is Poisson, $\mathcal{I}_x^S$ is invariant in n.

5.3.2 Positive Implications

We are now ready to derive several implications of two key features of the optimal procurement mechanism—whether designed to maximize welfare or profit. First, the mechanism takes the form of a sequence of auctions, used to elicit firms' cost information. Second, the buyer's experimentation and solicitation strategies are governed by an index policy.

A key implication is that, when the distribution F is Poisson or UL-concave, the search index declines with each additional search. This implies that the buyer becomes increasingly lenient over time in her willingness to tolerate underperformance from the firms already in the mechanism. The notion of such leniency is particularly sharp when the signal-generating process satisfies the NNBN (No News is Bad News) property. To formalize this idea, let $A_x(c; m, n)$ denote the number of failed attempts the buyer allows a firm with cost c (revealed via its bid upon entry) before initiating a new search, conditional on the fact that it took m searches to identify n firms. This is defined under the optimal mechanism for $x = W, PR$, corresponding to welfare and profit maximization, respectively. That is, when the first m searches have identified n firms, any firm with cost c is granted at most $A_x(c; m, n)$ opportunities to supply a satisfactory service (i.e., to generate a signal $\vartheta = 1$) before the $(m+1)$-th search is triggered. Importantly, the $(m+1)$-th search will not occur as long as there exists a firm in the mechanism with state $\omega^P = (c, \vartheta^{s-1})$ such that either $\sum_{l=0}^{s-1} \vartheta_l > 0$ (the firm has already provided a satisfactory service at least once), or $s - 1 < A_x(c; m, n)$ (the firm has not yet exhausted its allowed number of attempts).

Lemma 3. *The function $A_x(c; m, n)$ satisfies the following properties, for $x = W, PR$:*

1. *It is decreasing in c and increasing in m;*
2. *If F is UL-concave (respectively, UL-convex), it is increasing (respectively, decreasing) in n.*[16]

That $A_x(c; m, n)$ is decreasing in c follows from the fact that a higher cost increases the buyer's unit cost of procurement, which naturally leads the buyer to be less tolerant of underperformance from higher-cost firms. In contrast, all other properties of the function $A_x(c; m, n)$ are nontrivial and follow from the characterization of the search index in Lemma 2.

Definition 4. Under objective $x = W, PR$, the buyer is said to be *more lenient* if she grants a larger number of failed attempts A_x to each firm in the mechanism before initiating a new search, and *less lenient* if she grants a smaller number of such attempts.

Proposition 2. *The following properties hold for both $x = W$ and $x = PR$:*

[16] If F is Poisson, $A_x(c; m, n)$ is invariant in n.

1. *Monotonicity of leniency in search depth under Poisson or UL-concave priors:*

   When F is either Poisson or UL-concave, the buyer becomes more lenient as the number of searches increases, regardless of how many firms have been identified. That is, for any cost c, and for all $m' > m$, $n' \geq n$, $A_x(c; m', n') \geq A_x(c; m, n)$.

2. *Monotonicity of leniency in identified firms under UL-convex priors:*

   When F is UL-convex, for any fixed number of past searches m, the buyer becomes less lenient the more firms have been identified. That is, for any cost c, and for all $n' > n$, $A_x(c; m, n') \leq A_x(c; m, n)$.

While the result above is for the NNBN case, it extends more generally. Under the optimal mechanism for $x = W, PR$, the $(m+1)$-th search is not conducted as long as there exists at least one firm in the mechanism whose index satisfies $\mathcal{I}_x^P(c, \vartheta^{s-1}) \geq \mathcal{I}_x^S(m, n)$. That is, the buyer postpones additional search until all firms in the current mechanism have indexes strictly below the search index. This property follows directly from the indexability of the optimal policy, as established in Lemma 1.

Moreover, because the monotonicity properties of the search index $\mathcal{I}_x^S(m, n)$ in (m, n), as established in Lemma 2, do not rely on the specifics of the signal-generating process, the qualitative implications extend beyond the NNBN case. In more general environments, the buyer can still be interpreted as becoming more lenient if she tolerates lower values of the firm-level index $\mathcal{I}_x^P(c, \vartheta^{s-1})$ before initiating new searches.

In the NNBN setting, the index $\mathcal{I}_x^P(c, \vartheta^{s-1})$ for a firm that has never delivered a satisfactory service is a deterministic function of the firm's cost c and the number of failed attempts. Under more general signal structures, by contrast, the index depends on the buyer's posterior belief that the firm is able to deliver a satisfactory service, which may evolve stochastically with the signal history ϑ^{s-1}.[17]

Our second result (Proposition 3 below) establishes that firms joining the procurement mechanism at a later stage—i.e., after a greater number of searches—earn higher expected profits than those entering earlier, holding constant the number of competitors present at the time of entry, the firms' costs, and the individual history of experimentation outcomes. In other words, the longer it takes the buyer to assemble the pool of potential suppliers, the more favorable the expected profit becomes for a newly joining firm.

Proposition 3. *Consider two states of the procurement mechanism, $\mathcal{S} = (\omega^S, \mathcal{S}^P)$ and $\mathring{\mathcal{S}} = (\mathring{\omega}^S, \mathring{\mathcal{S}}^P)$, such that the total number of firms in the mechanism and each firm's state $\omega^P =$*

[17] In settings with more general payoff structures, the index $\mathcal{I}_x^P(c, \vartheta^{s-1})$ of a firm with cost c and signal history ϑ^{s-1} remains defined as the expected discounted sum of the buyer's flow payoffs until stopping, normalized by the expected discounted duration of time. However, unlike in simpler environments, this index need not admit a sufficient statistic—such as the posterior belief that the firm's product is safe or that the service is of high quality. Instead, it may depend on the entire signal history ϑ^{s-1}, without reducing to a lower-dimensional summary.

(c, ϑ^{s-1}) are identical in $\mathcal{S}$ and $\mathring{\mathcal{S}}$ (i.e., $\mathring{\mathcal{S}}^P = \mathcal{S}^P$). Let m and $\mathring{m}$ denote the number of searches conducted under $\mathcal{S}$ and $\mathring{\mathcal{S}}$, respectively, and assume that $\mathring{m} > m$. Consider the perspective of a firm with cost c, after being informed of the state of the procurement mechanism at the end of the period it joins. Regardless of whether the buyer's objective is welfare or profit maximization, the firm's expected discounted number of utilizations and hence its expected profits are greater after learning that the state is $\mathring{\mathcal{S}}$ than after learning that the state is $\mathcal{S}$.

The result follows from Theorem 1, Lemmas 1 and 2, and the arguments in the proof of Lemma 2, which establish that the distribution over the outcome of each search depends on the state of the search technology only through the statistics (m, n). Since a firm's index is a deterministic function of its state ω^P, every firm has the same index under the two states. In contrast, the search index is lower in $\mathring{\mathcal{S}}$ than in $\mathcal{S}$, as it took more searches to identify the same set of firms under $\mathring{\mathcal{S}}$. This last property, together with the fact that the distribution over future search outcomes depends only on (m, n) and is such that the likelihood of discovering new firms decreases with the number of past searches, implies that the joining firm's expected number of utilizations is higher under $\mathring{\mathcal{S}}$ than under $\mathcal{S}$. Finally, because the firm's expected profit under the optimal mechanism is strictly increasing in its expected utilization (by Lemma 1), it follows that its expected profit is also higher in state $\mathring{\mathcal{S}}$.

As in Lemma 2, the result does not rely on the assumption that the signal-generating process is NNBN. One may conjecture that the reason why a longer search process benefits a joining firm is that it signals a greater likelihood that the buyer has been disappointed by the firms previously utilized. While this property may also be relevant in specific settings, it plays no role in Proposition 3. What makes the result compelling is that the only dimension differentiating the two states is the length of the search process. The composition of the CS—formally captured by the function mapping each firm's state $\omega^P = (c, \vartheta^{s-1})$ to the number of firms in that state—is identical across $\mathcal{S}$ and $\mathring{\mathcal{S}}$ (i.e., $\mathring{\mathcal{S}}^P = \mathcal{S}^P$). In particular, note that if no further searches were to occur, the joining firm's expected profits would be the same under both $\mathcal{S}$ and $\mathring{\mathcal{S}}$. Thus, the difference in expected profits arises solely from the implications of a longer search process for the buyer's future behavior, not from differences in the current pool of firms.

5.3.3 Normative Implications

We now analyze the inefficiencies that arise under profit-maximizing procurement (namely, when the buyer does not account for firms' profits), focusing on those linked to the expansion of the pool of potential suppliers (CS), that is, the solicitation of new firms. The objective is to understand how the buyer's behavior diverges from the socially optimal one when she prioritizes her own surplus rather than total surplus. Define

$$\Delta(c; m, n) \equiv A_W(c; m, n) - A_{PR}(c; m, n)$$

as the difference in the number of failed attempts granted to a firm with cost c under welfare and profit maximization, conditional on the first m searches having identified n firms. Also, let $\hat{c}(m,n)$ denote the solution to

$$\mathcal{I}_W^S(m,n) - \mathcal{I}_{PR}^S(m,n) = G(c)/g(c),$$

that is, the cost level at which the wedge between the welfare-based and profit-based search indexes exactly offsets the handicap $G(c)/g(c)$ in the virtual cost $\gamma(c)$, which accounts for the implications of firm utilization on informational rents.

We start with the following observation:

Lemma 4. $\Delta(c;m,n)\cdot(c-\hat{c}(m,n))>0$.

The lemma establishes that, prior to the $(m+1)$-th expansion, low-cost firms are granted an inefficiently large number of failed attempts relative to what would be optimal under welfare maximization, while the opposite holds for high-cost firms. The result follows from the fact that, under both the profit-maximizing and the welfare-maximizing mechanism, the buyer follows an index policy, along with the recursive structure of the search index characterized in Part (2) of Theorem 1.

To understand the result, note first that, under profit maximization, the search index is distorted downward by a fixed positive amount to account for the informational rents that the buyer must leave to the identified firms. Specifically, each firm's profit-based index is lower than its welfare-based counterpart by an amount equal to

$$\mathcal{I}_W^P(c,\vartheta^{s-1}) - \mathcal{I}_{PR}^P(c,\vartheta^{s-1}) = \frac{G(c)}{g(c)},$$

reflecting the marginal rent the buyer must pay to the firms to induce truthful revelation. Since the search index aggregates across the indexes of the firms identified through search—as established in Part (2) of the theorem—the distortion in the search index satisfies

$$0 < \mathcal{I}_W^S(m,n) - \mathcal{I}_{PR}^S(m,n) < \frac{G(\bar{c})}{g(\bar{c})},$$

where $\bar{c}$ is the highest cost in the support of F. Other things equal, this wedge in the search indexes contributes to delay in the solicitation of new firms.

Next note that the distortion in each firm's index is smaller than the distortion in the search index for low-cost firms, but larger for high-cost firms. This asymmetry arises because the handicap $G(c)/g(c)$ in the virtual costs is increasing in c, while the distortion in the search index reflects an average across all identified firms. Since, under both welfare and profit maximization, a new expansion is triggered if and only if all firms currently in the mechanism have indexes

below the search index, these properties imply that—prior to launching the next expansion—the buyer, under profit maximization, tolerates an inefficiently large deterioration in the indexes of low-cost firms and an inefficiently small deterioration in the indexes of high-cost firms. The lemma formalizes this insight by leveraging these monotonicity properties along with the structure of the NNBN setting, in which the deterioration in each firm's index is solely determined by the number of failed attempts.

Note that the threshold $\hat{c}$ is generally non-monotonic in m. To see why, consider that when m is small, search is expected to yield many new firms, as shown in the proof of Lemma 2. Some of these firms are likely to have low costs. Since such firms are the most influential in determining the search index—and because their handicap $G(c)/g(c)$ is small—it follows that the wedge $\mathcal{I}^S_W(m,n) - \mathcal{I}^S_{PR}(m,n)$ is also small for small m.[18] Similarly, when m is large, the expected number of new firms identified in future searches is very small (again by Lemma 2), so the wedge between the welfare- and profit-based search indexes is again small. Thus, in both early and late stages of the procurement process, $\hat{c}(m,n)$ is close to $\underline{c}$, implying that most firms are granted an inefficiently low number of failed attempts under profit maximization relative to the welfare benchmark.

Next, fix m and consider how variations in n affect the threshold $\hat{c}(m,n)$. When the prior is UL-concave or Poisson, $\hat{c}(m,n)$ is typically non-monotonic in n. The reason is that, in these cases, the probability of identifying new firms decreases with the number of firms already found. As a result, the same logic used to explain the non-monotonicity of $\hat{c}(m,n)$ in m applies here: when few or many firms have already been identified, the wedge $\mathcal{I}^S_W(m,n) - \mathcal{I}^S_{PR}(m,n)$ between the search indexes is small which implies that $\hat{c}(m,n)$ is close to $\underline{c}$.

By contrast, when the distribution is UL-convex, identifying more firms increases the buyer's expectation that future searches will yield even more new firms, including those with low costs. In this case, the wedge $\mathcal{I}^S_W(m,n) - \mathcal{I}^S_{PR}(m,n)$ may decrease with n, implying that $\hat{c}(m,n)$ may also decrease in n. Under UL-convexity, it is thus possible that too many firms are granted too few failed attempts when the pool of participating firms becomes large.

Proposition 4. *The following observations characterize the inefficiencies induced by profit maximization:*

- *Low-cost firms (i.e., those with $c < \hat{c}(m,n)$) contribute to inefficient entrenchment—the buyer delays the solicitation of new suppliers longer than would be efficient under welfare maximization. Conversely, high-cost firms (with $c > \hat{c}(m,n)$) contribute to inefficient impatience with incumbents, prompting premature search relative to the welfare benchmark.*

[18] That low-cost firms are the most salient in the determination of the search index follows from its recursive structure established in Theorem 1, which implies that (a) only the indexes of firms whose index exceeds the search index contribute to its computation, and (b) the optimal policy π used in the definition of the search index is itself an index policy, which prioritizes firms with the lowest costs.

- *An analyst who lacks information about firms' indexes but knows that their bids are below $\hat{c}(m,n)$ can predict excessive delay in the expansion of the pool of potential suppliers, relative to the efficient benchmark.*

As anticipated in the Introduction, and in contrast to standard screening problems, when the pool of firms is endogenous and gradually constructed over time, all firms contribute to search inefficiencies, regardless of their type. However, low- and high-cost firms contribute in opposite directions: low-cost firms tend to generate inefficient delays in the search for new suppliers, while high-cost firms induce excessively early expansions.

6 Conclusions

We introduce a model of dynamic experimentation in which the decision maker alternates between exploring alternatives in the consideration set and searching for new alternatives to explore in the future. Each search stochastically delivers a new set of alternatives of varying types, which are then added to the consideration set. As a result, the consideration set is constructed gradually in response to the information revealed over time. We characterize the optimal policy and show how the trade-off between exploring existing alternatives and expanding the consideration set depends on the properties of the search technology. This evolving trade-off is governed by a comparison of independent indexes, with the search/CS-expansion index computed recursively to reflect future optimal decisions.

We illustrate how the results can be put to work in concrete applications by considering the design problem of a buyer repeatedly procuring a service from multiple firms. The buyer faces uncertainty about the number of potential suppliers and learns the quality of their services through experimentation, while firms possess private information about their costs. The optimal procurement mechanism—whether the buyer maximizes profit or total surplus—takes the form of a sequence of auctions with endogenous and time-varying solicitation. Upon entry, firms bid for the unit price at which they are willing to supply the service and receive a lump-sum transfer designed to elicit truthful reporting. In each subsequent period, the buyer alternates between procuring from firms with the highest index and soliciting new firms via advertising or other outreach efforts. On the positive side, the analysis explains why a buyer may become more lenient or more demanding with incumbent suppliers over time, and why later entrants may earn higher expected profits, holding all else equal. On the normative side, we show that all firms contribute to inefficiencies in the expansion and experimentation process, irrespective of their type—but that low- and high-cost firms contribute in opposite directions.

The model can be used to study many dynamic problems in which the set of feasible alternatives is not known ex ante—either due to limited attention or because alternatives are revealed sequentially, for example by intermediaries such as search engines or online platforms. Exploring

such applications offers a promising direction for future research.

References

Austen-Smith, D. and C. Martinelli (2018). "Optimal exploration." Working paper.

Bergemann, D. and J. Välimäki (2008). "Bandit problems". In: *The New Palgrave Dictionary of Economics*. Ed. by Steven N. Durlauf and Lawrence E. Blume. Basingstoke: Palgrave Macmillan.

Bergemann, D. and J. Välimäki (2010). "The dynamic pivot mechanism." *Econometrica*, 78(2), 771-789.

Caplin, A., Dean, M. and D. Martin (2011). "Search and satisficing." *American Economic Review*, 101(7), 2899-2922.

Caplin, A., Dean, M. and J. Leahy (2019). "Rational inattention, optimal consideration sets, and stochastic choice," *The Review of Economic Studies*, 86(3), 1061-1094.

Che, Y. K. and K. Mierendorff (2019). "Optimal dynamic allocation of attention." *American Economic Review*, 109(8), 2993-3029.

Choi, M. and L. Smith (2016). "Optimal sequential search among alternatives." mimeo University of Wisconsin.

Doval, L. (2018). "Whether or not to open Pandora's box." *Journal of Economic Theory* 175, 127–158.

Dumitriu, I., Tetali, P. and P. Winkler (2003). "On playing golf with two balls," *SIAM Journal on Discrete Mathematics*, 16(4), 604-615.

Eliaz, K. and R. Spiegler (2011). "Consideration sets and competitive marketing," *The Review of Economic Studies*, 78(1), 235-262.

Fershtman, D. and A. Pavan (2017). "Pandora's Auctions: Dynamic Matching with Unknown Preferences". *American Economic Review P&P*, 107(5), 186-190.

Fershtman, D. and A. Pavan (2021). "Soft Affirmative Action and Minority Recruitment," *American Economic Review: Insights*, 3(1), 1-18.

Fudenberg, D., Strack P. and T. Strzalecki (2018). "Speed, accuracy, and the optimal timing of choices," *American Economic Review* 108 (12), 3651–3684.

Garfagnini, U. and B. Strulovici (2016). "Social experimentation with interdependent and expanding technologies," *Review of Economic Studies*, 83(4), 1579-1613.

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

Glazebrook, K. D. (1979). "Stoppable families of alternative bandit processes." *Journal of Applied Probability* 16.4: 843-854.

Gossner, O., Steiner, J. and C. Stewart (2021). "Attention, please!," *Econometrica*, 89(4), 1717-1751.

Greminger, R. (2022). "Optimal search and discovery," *Management Science* 68.5, 3904-3924.

Kakade, S. M., Lobel, I., and H. Nazerzadeh (2013). "Optimal dynamic mechanism design and the virtual-pivot mechanism". *Operations Research*, 61(4), 837-854.

Ke, T. T., Z.-J. M. Shen and J. M. Villas-Boas (2016). "Search for information on multiple products," *Management Science* 62 (12), 3576-3603.

Ke, T. T. and J. M. Villas-Boas (2019). "Optimal learning before choice," *Journal of Economic Theory* 180, 383-437.

Keller, G. and A. Oldale (2003). "Branching bandits: a sequential search process with correlated pay-offs," *Journal of Economic Theory*, 113(2), 302-315.

Lauermann, S., and A. Wolinsky (2017). "Bidder solicitation, adverse selection, and the failure of competition." *American Economic Review*, 107(6), 1399-1429.

Lauermann, S., and A. Wolinsky (2025). "Auctions with frictions: Recruitment, entry, and limited commitment." *Forthcoming, Review of Economic Studies.*

Liang, A., Mu, X., and V. Syrgkanis (2022). "Dynamically aggregating diverse information," *Econometrica*, 90(1), 47-90.

Mandelbaum, A. (1986). "Discrete multi-armed bandits and multi-parameter processes," *Probability Theory and Related Fields 71*(1), 129–147.

Manzini, P. and M. Mariotti (2014). "Stochastic choice and consideration sets," *Econometrica*, 82(3), 1153-1176.

Masatlioglu, Y., Nakajima, D. and E. Y. Ozbay (2012). "Revealed attention," *American Economic Review*, 102(5), 2183-2205.

Olszewski, W. and R. Weber (2015). "A more general Pandora rule?," *Journal of Economic Theory* 160, 429–437.

Pavan, A., Segal, I., and J. Toikka (2014). "Dynamic mechanism design: A myersonian approach". *Econometrica*, 82(2), 601-653.

Carnehl, C. and J. Schneider (2023). "On risk and time pressure: When to think and when to do," *Journal of the European Economic Association,* 21(1), 1-47.

Simon, H. A. (1955). "A behavioral model of rational choice," *The Quarterly Journal of Economics*, 69(1), 99-118.

Weber, R., (1992). "On the Gittins Index for Multiarmed Bandits," *The Annals of Applied Probability*, Vol. 2(4), 1024-1033.

Weiss, G. (1988). "Branching bandit processes," *Probability in the Engineering and Informational Sciences*, 2(3), 269-278.

Weitzman, M. (1979). "Optimal search for the best alternative," *Econometrica* 47 (3), 641–654.

A Proofs

Proof of Theorem 1. Below we first establish the result in Part (2) of the theorem and then use the recursive representation of the search index in (3) to show that, when the DM follows an index policy, her expected (per-period) payoff satisfies the representation in (4), thus establishing Part (3) of the theorem. We then show how the representation of the DM's payoff in (4), along with the recursive representation of the search index in Part (2) of the theorem and an appropriate description of the state space that exploits the classification of the alternatives into categories, permits us to establish Part (1) of the theorem, i.e., the optimality of the index policy, by means of a novel proof that shows that the DM's payoff under such a policy satisfies the Bellman equation for the dynamic program under consideration.

Part (2). Let $\hat{\tau}$ be the optimal stopping time in the definition of $\mathcal{I}^S(\omega^S)$. Note that, at $\hat{\tau}$, the index of each alternative brought to the CS by the search under consideration (initiated in state ω^S), as well as the index of search itself, must be weakly smaller than $\mathcal{I}^S(\omega^S)$. Otherwise, by continuing to search, or by selecting one of the alternatives brought to the CS by the search under consideration for which the index is larger than $\mathcal{I}^S(\omega^S)$ and stopping optimally from that moment onward, the DM would attain an average payoff per unit of average discounted time

$$\frac{\mathbb{E}^{\pi}\left[\sum_{s=0}^{\tau-1}\delta^s U_s|\omega^S\right]}{\mathbb{E}^{\pi}\left[\sum_{s=0}^{\tau-1}\delta^s|\omega^S\right]}$$

strictly greater than $\mathcal{I}^S(\omega^S)$, contradicting the optimality of $\hat{\tau}$ in the definition of $\mathcal{I}^S(\omega^S)$.[19] This implies that $\hat{\tau}$ is weakly greater than τ^*, where the latter is the first time at which the index of search and the index of each alternative brought to the CS by the search under consideration are weakly below $\mathcal{I}^S(\omega^S)$. Moreover, since at τ^* the index of search and of each alternative brought to the CS by the search under consideration are weakly below $\mathcal{I}^S(\omega^S)$, if $\hat{\tau} > \tau^*$, the average payoff per unit of average discounted time between τ^* and $\hat{\tau}$ must be equal to $\mathcal{I}^S(\omega^S)$. Hence, under the optimal selection rule in the definition of $\mathcal{I}^S(\omega^S)$, the average payoff per unit of average discounted time from 0 to τ^* must also be equal to $\mathcal{I}^S(\omega^S)$. This implies that the optimal stopping time in the definition of $\mathcal{I}^S(\omega^S)$ can be taken to be τ^*. Because the index policy χ^* selects in each period between 0 and τ^* the alternative for which the average payoff per unit of average discounted time is the largest (including search), we have that the optimal selection rule π in the definition of $\mathcal{I}^S(\omega^S)$ must coincide with the index policy χ^*. That $\mathcal{I}^S(\omega^S)$ satisfies the recursive representation in Part (2) then follows from the arguments above.

Part (3). We construct the following stochastic process based on the values of the indexes, and the expansion of the CS through search, *under the index policy* χ^*. Starting with the initial

[19] Since infinity is allowed as a value of the stopping time, the supremum in the definitions of $\mathcal{I}^S$ (and $\mathcal{I}^P$) is attained, that is, an optimal stopping time exists (the arguments are similar to those in Mandelbaum, 1986, and hence omitted).

state $\mathcal{S}_0 = (\mathcal{S}_0^P, \omega_0^S)$, let $v^0 \equiv \max\{\mathcal{I}^*(\mathcal{S}_0^P), \mathcal{I}^S(\omega_0^S)\}$. Let $t(v^0)$ be the first time at which, when the DM follows the policy χ^*, all indexes are *strictly* below v^0, with $t(v^0) = \infty$ if this event never occurs. Note that $t(v^0)$ differs from $\kappa(v^0)$, as $\kappa(v^0) = 0$ is the first time at which all indexes are *weakly* below v^0. Next let $v^1 \equiv \max\{\mathcal{I}^*(\mathcal{S}^P_{t(v^0)}), \mathcal{I}^S(\omega^S_{t(v^0)})\}$ be the value of the largest index at $t(v^0)$, where $\mathcal{S}_{t(v^0)} = (\mathcal{S}^P_{t(v^0)}, \omega^S_{t(v^0)})$ is the state of the decision problem in period $t(v^0)$. Note that, by construction, $t(v^0) = \kappa(v^1)$. Furthermore, when $t(v^0) < \infty$, if $v^0 > \mathcal{I}^S(\omega_0^S)$, then $\omega^S_{t(v^0)} = \omega_0^S$. We can proceed in this manner to obtain a strictly decreasing sequence of values $(v^i)_{i\geq 0}$, with corresponding stochastic times $(\kappa(v^i))_{i\geq 0}$. Note that the values v^i are all non-negative, as the DM's outside option is normalized to zero.

Next, for any $i = 0, 1, 2, ...$, let $\eta^i \equiv \sum_{s=\kappa(v^i)}^{\kappa(v^{i+1})-1} \delta^{s-\kappa(v^i)} U_s$ denote the discounted sum of the net payoffs between periods $\kappa(v^i)$ and $\kappa(v^{i+1}) - 1$, when the DM follows the index policy, and let $(\eta^i)_{i\geq 0}$ denote the corresponding sequence of discounted accumulated net payoffs, with $\eta^i = 0$ if $\kappa(v^i) = \infty$.

Denote by $\mathcal{V}(\mathcal{S}_0)$ the expected (per-period) net payoff under the index policy χ^*, given the initial state of the problem $\mathcal{S}_0$. That is, $\mathcal{V}(\mathcal{S}_0) = (1-\delta)\mathbb{E}^{\chi^*}\left[\sum_{t=0}^{\infty} \delta^t U_t | \mathcal{S}_0\right]$. By definition of the processes $(\kappa(v^i))_{i\geq 0}$ and $(\eta^i)_{i\geq 0}$, $\mathcal{V}(\mathcal{S}_0) = (1-\delta)\mathbb{E}^{\chi^*}\left[\sum_{i=0}^{\infty} \delta^{\kappa(v^i)} \eta^i | \mathcal{S}_0\right]$. Next, using the definition of the indexes in (1) and (2), observe that

$$v^i = \frac{(1-\delta)\mathbb{E}^{\chi^*}\left[\eta^i | \mathcal{S}_{\kappa(v^i)}\right]}{\mathbb{E}^{\chi^*}\left[1 - \delta^{\kappa(v^{i+1})-\kappa(v^i)} | \mathcal{S}_{\kappa(v^i)}\right]}. \tag{8}$$

To see why (8) holds, recall that, at period $\kappa(v^i)$, given the state of the decision problem $\mathcal{S}_{\kappa(v^i)}$, the value of the highest index is v^i. Now suppose that the alternative corresponding to v^i is a physical alternative and that all other physical alternatives' indexes, as well as the index of search, are strictly below v^i. Recall that the optimal stopping time τ in the definition of the index of the physical alternative corresponding to v^i in (1) is the first period (strictly above $\kappa(v^i)$) at which the alternative's index falls below v^i. While it is convenient to take this fall to be weak, it is well known that one can equivalently take the fall to be strict. That is, stopping at the first period at which the index reaches a value equal to or smaller than the value at the time the index was computed is optimal, but so is stopping at the first period at which the index reaches a value strictly below the one at the time the index was computed. Now recall that $t(v^i)$ is the first time at which all indexes are strictly below v^i. Because the CS in period $\kappa(v^i)$ contains only one alternative with index equal to v^i (the physical one under consideration), $t(v^i)$ also coincides with the first period at which the index of the specific alternative under consideration drops strictly below v^i. Recall that v^{i+1} is the largest index at period $t(v^i)$ and that $t(v^i) = \kappa(v^{i+1})$. The definition of the index in (1), along with the optimality of stopping at the first time the

index drops strictly below its initial value, and the definition of η^i, then imply that

$$v^i = \frac{\mathbb{E}^{\chi^*}\left[\sum_{s=\kappa(v^i)}^{\kappa(v^{i+1})-1} \delta^{s-\kappa(v^i)} U_s | \mathcal{S}_{\kappa(v^i)}\right]}{\mathbb{E}^{\chi^*}\left[\sum_{s=\kappa(v^i)}^{\kappa(v^{i+1})-1} \delta^{s-\kappa(v^i)} | \mathcal{S}_{\kappa(v^i)}\right]} = \frac{\mathbb{E}^{\chi^*}\left[\eta^i | \mathcal{S}_{\kappa(v^i)}\right]}{\mathbb{E}^{\chi^*}\left[\frac{1-\delta^{\kappa(v^{i+1})-\kappa(v^i)}}{1-\delta} | \mathcal{S}_{\kappa(v^i)}\right]}$$

which corresponds to the formula in (8).

Next, suppose that the alternative with the highest index at period $\kappa(v^i)$ is search, and that all physical alternatives in the CS in period $\kappa(v^i)$ have an index strictly smaller than v^i. As shown in the proof of Part (1) of Theorem 1 above, the optimal stopping time in the definition of the index of search in (2) is the first period (strictly above $\kappa(v^i)$) at which the index of search and of all the alternatives introduced through search, fall weakly below v^i. Equivalently, as discussed above, the optimal stopping time can also be taken to be the first period at which the index of search and of all the alternatives introduced through search fall strictly below v^i. Because all physical alternatives in the CS at period $\kappa(v^i)$ have an index strictly below v^i, such a period coincides with $t(v^i)$, that is, with the first period at which the index of search and of all alternatives in the CS are strictly below v^i. Using the above property of the optimal stopping time in the definition of the search index in (2), along with the fact that $t(v^i) = \kappa(v^{i+1})$ and the definition of η^i, we then have that the search index evaluated at period $\kappa(v^i)$ also satisfies the condition in (8).

Finally, suppose that, at period $\kappa(v^i)$, there are multiple options ("physical" alternatives and/or search) with index v^i. Then observe that the average sum $\mathbb{E}^{\chi^*}\left[\sum_{s=\kappa(v^i)}^{\kappa(v^{i+1})-1} \delta^{s-\kappa(v^i)} U_s | \mathcal{S}_{\kappa(v^i)}\right]$ of the discounted net payoffs from utilizing all options whose period-$\kappa(v^i)$ index is equal to v^i till the first period $t(v^i) = \kappa(v^{i+1})$ at which the indexes of all options are strictly below v^i, normalized by the average per unit discounted time $\mathbb{E}^{\chi^*}\left[\sum_{s=\kappa(v^i)}^{\kappa(v^{i+1})-1} \delta^{s-\kappa(v^i)} | \mathcal{S}_{\kappa(v^i)}\right]$ is the same as the average sum $\mathbb{E}^{\chi^*}\left[\sum_{s=\kappa(v^i)}^{T-1} \delta^{s-\kappa(v^i)} U_s | \mathcal{S}_{\kappa(v^i)}\right]$ of the discounted net payoffs from utilizing each individual option with index (at period $\kappa(v^i)$) equal to v^i till the first time T at which that option's index (and, in case the option is search, also the indexes of all alternatives brought to the CS by the search initiated at $\kappa(v^i)$) fall strictly below v^i, normalized by the average discounted time $\mathbb{E}^{\chi^*}\left[\sum_{s=\kappa(v^i)}^{T-1} \delta^{s-\kappa(v^i)} | \mathcal{S}_{\kappa(v^i)}\right]$. This follows from the independence of the processes. Hence, Condition (8) also holds when, at $\kappa(v^i)$, there are multiple options with index v^i.

Multiplying both sides of (8) by $\delta^{\kappa(v^i)}$, rearranging terms, and using the fact that $\delta^{\kappa(v^i)}$ is known at $\kappa(v^i)$, we have that

$$(1-\delta)\mathbb{E}^{\chi^*}\left[\delta^{\kappa(v^i)} \eta^i | \mathcal{S}_{\kappa(v^i)}\right] = v^i \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v^i)} - \delta^{\kappa(v^{i+1})} | \mathcal{S}_{\kappa(v^i)}\right].$$

Taking expectations of both sides of the previous equality given the initial state $\mathcal{S}_0$, and using

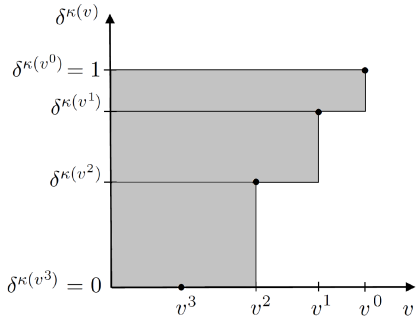

Figure 1: An illustration of the function $\delta^{\kappa(v)}$ and the region $\sum_{i=0}^{\infty} v^i \left(\delta^{\kappa(v^i)} - \delta^{\kappa(v^{i+1})}\right) = \int_0^{\infty} v \mathrm{d}\delta^{\kappa(v)}$, for a particular path with $\kappa(v^3) = \infty$.

the law of iterated expectations, we have that

$$(1-\delta)\mathbb{E}^{\chi^*}\left[\delta^{\kappa(v^i)}\eta^i|\mathcal{S}_0\right] = \mathbb{E}^{\chi^*}\left[v^i\left(\delta^{\kappa(v^i)} - \delta^{\kappa(v^{i+1})}\right)|\mathcal{S}_0\right].$$

If follows that

$$\mathcal{V}(\mathcal{S}_0) = \mathbb{E}^{\chi^*}\left[\sum_{i=0}^{\infty} v^i\left(\delta^{\kappa(v^i)} - \delta^{\kappa(v^{i+1})}\right)|\mathcal{S}_0\right]. \tag{9}$$

Next, note that $\delta^{\kappa(v^i)} = 0$ whenever $\kappa(v^i) = \infty$, and that, for any $i = 0, 1, ...$, $\kappa(v) = \kappa(v^{i+1})$ for all $v^{i+1} < v < v^i$. It follows that (9) is equivalent to

$$\mathcal{V}(\mathcal{S}_0) = \mathbb{E}^{\chi^*}\left[\int_0^{\infty} v \mathrm{d}\delta^{\kappa(v)}|\mathcal{S}_0\right] = \mathbb{E}^{\chi^*}\left[\int_0^{\infty}\left(1-\delta^{\kappa(v)}\right)\mathrm{d}v|\mathcal{S}_0\right] = \int_0^{\infty}\left(1-\mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|\mathcal{S}_0\right]\right)\mathrm{d}v. \tag{10}$$

The construction of the integral function (10) is illustrated in Figure 1.

Part (1). The proof of Part (1) exploits the recursive representation of the search index established in Part (2) of Theorem 1, along with the representation of the DM's payoff under the index rule established in Part (3) of Theorem 1 and an appropriate description of the state space, to verify that the DM's payoff under the index policy satisfies the Bellman equation of the corresponding dynamic program. The proof is in two steps. Step 1 uses the representation of the DM's payoff under the index rule established in Part (3) of Theorem 1 in the main text to characterize how much the DM obtains from following the index policy χ^* from the outset rather than being forced to make a different decision in the first period and then reverting to χ^* from the next period onward. Step 2 then uses the results in step 1 to establish the optimality of χ^* through dynamic programming.

Step 1. In the analysis below, we find it useful to describe changes in the composition of the CS, the evolution of the search technology, as well as all information acquired about the alternatives, entirely in terms of transitions between states. Rather than keeping track of the

collection of kernels $G_\xi(\vartheta^m;\mu)$ describing the conditional distributions from which the marginal signals ϑ_{m+1} are drawn, we describe directly the evolution of each alternative's state ω^P as follows. When the DM explores an alternative currently in state ω^P, its new state $\tilde{\omega}^P$ is drawn from a distribution $H_{\omega^P} \in \Delta(\Omega^P)$ that is invariant to time.[20] When the DM explores a different alternative, or expands the CS, the alternative currently in state ω^P remains in the same state with certainty at the beginning of the next period. Similarly, each time search is conducted, given the current state of the search technology ω^S, the new state of the search technology $\tilde{\omega}^S$ is drawn from a distribution $H_{\omega^S} \in \Delta(\Omega^S)$.The distributions H_{ω^s} are time-homogeneous (i.e., the evolution of the search technology depends on past search outcomes but is invariant in calendar time), and the outcome of each new search is drawn from H_{ω^S} independently from the idiosyncratic and time-varying component θ of each alternative in the CS.

Abusing notation, then denote the state of the decision problem by a function $\mathcal{S} : \Omega \to \mathbb{N}$ that specifies, for each $\omega \in \Omega$, including $\omega \in \Omega^S$, the number of alternatives, including the search technology, that are in state ω.[21] Given this notation, for any pair of states $\mathcal{S}'$ and $\mathcal{S}''$ then define $\mathcal{S}' \vee \mathcal{S}'' \equiv (\mathcal{S}'(\omega) + \mathcal{S}''(\omega) : \omega \in \Omega)$ and $\mathcal{S}' \backslash \mathcal{S}'' \equiv (\max\{\mathcal{S}'(\omega) - \mathcal{S}''(\omega), 0\} : \omega \in \Omega)$. Any feasible state of the decision problem must specify one, and only one, state of the search technology (i.e., one state $\hat{\omega}^S$ for which $\mathcal{S}(\hat{\omega}^S) = 1$ and such that $\mathcal{S}(\omega^S) = 0$ for all $\omega^S \neq \hat{\omega}^s$). However, it will be convenient to consider fictitious (infeasible) states where search is not possible, as well as fictitious states with multiple search possibilities. If the state of the decision problem is such that either (i) the CS is empty, or (ii) there is a single alternative in the CS and the latter cannot be expanded, we will denote such a state by $e(\omega)$, where $\omega \in \Omega$ is the state of the search technology in case (i) and of the single physical alternative in case (ii). Throughout the analysis below, we maintain the assumption that an outside option with value equal to zero is available to the DM. However, to avoid possible confusion, here we do not explicitly treat the outside option as a separate alternative.

Lemma 5. *For any $v \in \mathbb{R}$ and states $\mathcal{S}'$ and $\mathcal{S}''$, $\kappa(v|\mathcal{S}' \vee \mathcal{S}'') = \kappa(v|\mathcal{S}') + \kappa(v|\mathcal{S}'')$.*

Proof of Lemma 5. The result follows from the fact that the state of each alternative that is not explored in a given period remains unchanged, along with the fact that the time-varying components θ of the various alternatives evolve independently of one another and of the state of the search technology, given the alternatives' categories ξ. Similarly, the state of the search technology remains unchanged in periods in which search is not conducted, and evolves independently of the time-varying component θ in the state of each existing alternative, given the alternatives'

[20]Because each alternative's category ξ is fixed, given the current state $\omega^P = (\xi, \theta)$, the distribution H_{ω^P} assigns probability one to states whose category is ξ and whose signal history $\vartheta^{m+1} = (\vartheta^m, \vartheta_{m+1})$ is a "follower" of ϑ^m, meaning that it is obtained by adding a new signal realization ϑ_{m+1} to the history ϑ^m.

[21]With this representation, there is a unique $\hat{\omega}^s \in \Omega^S$ such that $\mathcal{S}(\omega^S) = 1$ if $\omega^s = \hat{\omega}^s$ and $\mathcal{S}(\omega^S) = 0$ if $\omega^s \neq \hat{\omega}^s$. The special case where the DM does not have the option to search corresponds to the case where for all $\omega^S \in \Omega^S$, $\mathcal{S}(\omega^S) = 0$.

categories ξ. Furthermore, the index of each alternative is a function only of the alternative's state, and the index of search is a function only of the state of the search technology. Therefore, all indexes evolve independently of one another (conditional on the alternatives' categories), and evolve only when their corresponding decision (search or exploration of an alternative) is chosen. Since the decisions are taken under the index policy χ^*, the result follows from the fact that, starting from any state $\mathcal{S}$, the *total* time it takes to bring *all* indexes (that is, those of the alternatives in the CS as well as the index of search) below any value v is the sum (across alternatives in the CS and search) of the *individual* times necessary to bring each index below v in isolation. □

Given the initial state $\mathcal{S}_0$, for any $\omega^P \in \{\hat{\omega}^P \in \Omega^P : \mathcal{S}_0^P(\hat{\omega}^P) > 0\}$, denote by $\mathbb{E}\left[u|\omega^P\right]$ the immediate expected payoff from exploring an alternative in state ω^P and by $\tilde{\omega}^P$ the new state of that alternative triggered by its exploration (drawn from H_{ω^P}). Let

$$V^P(\omega^P|\mathcal{S}_0) \equiv (1-\delta)\mathbb{E}\left[u|\omega^P\right] + \delta\mathbb{E}^{\chi^*}\left[\mathcal{V}\left(\mathcal{S}_0 \backslash e(\omega^P) \vee e(\tilde{\omega}^P)\right)|\omega^P\right] \tag{11}$$

denote the DM's payoff from starting with exploring an alternative in state ω^P and then following the index policy χ^* from the next period onward. Similarly, let

$$V^S(\omega^S|\mathcal{S}_0) \equiv -(1-\delta)\mathbb{E}\left[c|\omega^S\right] + \delta\mathbb{E}^{\chi^*}\left[\mathcal{V}\left(S_0 \backslash e(\omega^S) \vee e(\tilde{\omega}^S) \vee W^P(\tilde{\omega}^S)\right)|\omega^S\right] \tag{12}$$

denote the DM's payoff from expanding the CS when the state of search is ω^S, and then following the index policy χ^* from the next period onward, where $\mathbb{E}\left[c|\omega^S\right]$ is the immediate expected cost from searching (when the state of the search technology is ω^S), $\tilde{\omega}^S$ is the new state of the search technology, and $W^P(\tilde{\omega}^S)$ is the state of the new alternatives brought to the CS by the current search, with c and $W^P(\tilde{\omega}^S)$ jointly drawn from the distribution H_{ω^S}. Note that $W^P(\tilde{\omega}^S)$ is a deterministic function of the new state $\tilde{\omega}^S$ of the search technology. To see this, recall that, for any $m \in \mathbb{N}$, the function E_m in the definition of the state of the search technology counts how many alternatives of each possible state ω^P have been added to the CS, as a result of the m-th search.

We introduce a fictitious "auxiliary option" which is available at all periods and yields a constant reward $M < \infty$ when chosen. Denote the state corresponding to this fictitious auxiliary option by ω_M^A, and enlarge Ω^P to include ω_M^A. Similarly, let $e(\omega_M^A)$ denote the state of the problem in which only the auxiliary option with fixed reward M is available. Since the payoff from the auxiliary option is constant at M, if $v \geq M$, then $\kappa(v|\mathcal{S}_0 \vee e(\omega_M^A)) = \kappa(v|\mathcal{S}_0)$, whereas if $v < M$, then $\kappa(v|\mathcal{S}_0 \vee e(\omega_M^A)) = \infty$. Hence, the representation of the DM's payoff under the index policy in Part (3) of Theorem 1 in the main text, adapted to the fictitious environment that includes

the auxiliary option, implies that

$$\mathcal{V}(\mathcal{S}_0 \vee e(\omega_M^A)) = \int_0^\infty \left(1 - \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|\mathcal{S}_0 \vee e(\omega_M^A)\right]\right) \mathrm{d}v = M + \int_M^\infty \left(1 - \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|\mathcal{S}_0\right]\right) \mathrm{d}v$$
$$= \mathcal{V}(\mathcal{S}_0) + \int_0^M \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|\mathcal{S}_0\right] \mathrm{d}v. \tag{13}$$

The definition of χ^*, along with Conditions (11) and (12), then imply the following:

Lemma 6. *For any* (ω^S, ω^P, M),

$$\mathcal{V}(e(\omega^S) \vee e(\omega_M^A)) = \begin{cases} V^S(\omega^S|e(\omega^S) \vee e(\omega_M^A)) & \text{if } M \leq \mathcal{I}^S(\omega^S) \\ M > V^S(\omega^S|e(\omega^S) \vee e(\omega_M^A)) & \text{if } M > \mathcal{I}^S(\omega^S) \end{cases} \tag{14}$$

$$\mathcal{V}(e(\omega^P) \vee e(\omega_M^A)) = \begin{cases} V^P(\omega^P|e(\omega^P) \vee e(\omega_M^A)) & \text{if } M \leq \mathcal{I}^P(\omega^P) \\ M > V^P(\omega^P|e(\omega^P) \vee e(\omega_M^A)) & \text{if } M > \mathcal{I}^P(\omega^P). \end{cases} \tag{15}$$

Proof of Lemma 6. First note that the index corresponding to the auxiliary option is equal to M. Hence, if $M \leq \mathcal{I}^S(\omega^S)$, given $e(\omega^S) \vee e(\omega_M^A)$, χ^* prescribes to start with search, implying that $\mathcal{V}(e(\omega^S) \vee e(\omega_M^A)) = V^S(\omega^S|e(\omega^S) \vee e(\omega_M^A))$. If, instead, $M > \mathcal{I}^S(\omega^S)$, χ^* prescribes to select the auxiliary option forever, with an expected (per period) payoff of M. To see why, in this case, $M > V^S(\omega^S|e(\omega^S) \vee e(\omega_M^A))$, observe that the payoff $V^S(\omega^S|e(\omega^S) \vee e(\omega_M^A))$ from starting with search and then following χ^* in each subsequent period is equal to $V^S(\omega^S|e(\omega^S) \vee e(\omega_M^A)) = \mathbb{E}_{>1}^{\chi^*}\left[(1-\delta)\sum_{s=0}^{\bar{\tau}-1} \delta^s U_s + \delta^{\bar{\tau}} M|\omega^S\right]$, where $\bar{\tau}$ is the first time at which the index of search and of all the alternatives brought to the CS by search fall weakly below M, and where the expectation is under the process that obtains starting from $e(\omega^S) \vee e(\omega_M^A)$ by searching in the first period and then following the index policy in each subsequent period (the notation $\mathbb{E}_{>1}^{\chi^*}[\cdot]$ is meant to highlight that the expectation is under such a process). This follows from the fact that, once the DM, under χ^*, opts for the auxiliary option, he will continue to select that option in all subsequent periods. By definition of $\mathcal{I}^S(\omega^S)$,

$$M > \mathcal{I}^S(\omega^S) \equiv \sup_{\pi,\tau} \frac{\mathbb{E}^{\pi}\left[\sum_{s=0}^{\tau-1} \delta^s U_s|\omega^S\right]}{\mathbb{E}^{\pi}\left[\sum_{s=0}^{\tau-1} \delta^s|\omega^S\right]} \geq \frac{\mathbb{E}_{>1}^{\chi^*}\left[\sum_{s=0}^{\bar{\tau}-1} \delta^s U_s|\omega^S\right]}{\mathbb{E}_{>1}^{\chi^*}\left[\sum_{s=0}^{\bar{\tau}-1} \delta^s|\omega^S\right]}.$$

Rearranging, $M\mathbb{E}_{>1}^{\chi^*}\left[\sum_{s=0}^{\bar{\tau}-1} \delta^s|\omega^s\right] > \mathbb{E}_{>1}^{\chi^*}\left[\sum_{s=0}^{\bar{\tau}-1} \delta^s U_s|\omega^S\right]$. Therefore,

$$\mathbb{E}_{>1}^{\chi^*}\left[(1-\delta)\sum_{s=0}^{\bar{\tau}-1} \delta^s U_s + \delta^{\bar{\tau}} M|\omega^S\right] < M\mathbb{E}_{>1}^{\chi^*}\left[(1-\delta)\sum_{s=0}^{\bar{\tau}-1} \delta^s + \delta^{\bar{\tau}}|\omega^S\right] = M.$$

Similar arguments establish Condition (15). □

Next, for any initial state $\mathcal{S}_0$ of the decision problem, and any state $\omega^P \in \{\hat{\omega}^P \in \Omega^P : \mathcal{S}_0(\hat{\omega}^P) > 0\}$ of the alternatives in the CS corresponding to $\mathcal{S}_0$, let $D^P(\omega^P|\mathcal{S}_0) \equiv \mathcal{V}(\mathcal{S}_0) - V^P(\omega^P|\mathcal{S}_0)$ denote

the payoff differential between (a) starting by following the index rule χ^* right away and (b) exploring first one of the alternatives in state ω^P and then following χ^* thereafter. Similarly, let $D^S(\omega^S|\mathcal{S}_0) \equiv \mathcal{V}(\mathcal{S}_0) - V^S(\omega^S|\mathcal{S}_0)$ denote the payoff differential between (c) starting with χ^* and (d) starting with search in state ω^S and then following χ^*. The next lemma relates these payoff differentials to the corresponding ones in a fictitious environment with the auxiliary option. In the statement of the lemma, $\mathcal{S}_0 \setminus e(\omega^S)$ is the state of a fictitious problem where search is not possible, whereas $\mathcal{S}_0^P \backslash e(\omega^P)$ is the state of the CS obtained from $\mathcal{S}_0^P$ by subtracting an alternative in state ω^P.

Lemma 7. *Let $\mathcal{S}_0$ be the initial state of the decision problem, with $\omega^S \in \Omega^S$ denoting the state of the search technology, as specified in $\mathcal{S}_0$. We have that*

$$\begin{aligned} D^S(\omega^S|\mathcal{S}_0) = \int_0^{\mathcal{I}^*(\mathcal{S}_0^P)} D^S(\omega^S|e(\omega^S) \vee e(\omega_v^A))\, d\mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|\mathcal{S}_0 \setminus e(\omega^S)\right] \\ + \mathbb{E}^{\chi^*}\left[\delta^{\kappa(0)}|S_0 \backslash e(\omega^S))\right] D^S(\omega^S|e(\omega^S) \vee e(\omega_0^A)). \end{aligned} \tag{16}$$

Similarly, for any alternative in the CS in state $\omega^P \in \{\hat{\omega}^P \in \Omega^P : \mathcal{S}_0^P(\hat{\omega}^P) > 0\}$,

$$\begin{aligned} D^P(\omega^P|\mathcal{S}_0) = \int_0^{max\{\mathcal{I}^*(\mathcal{S}_0^P \backslash e(\omega^P)), \mathcal{I}^S(\omega^S)\}} D^P(\omega^P|e(\omega^P) \vee e(\omega_v^A))\, d\mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|\mathcal{S}_0 \setminus e(\omega^P)\right] \\ + \mathbb{E}^{\chi^*}\left[\delta^{\kappa(0)}|S_0 \backslash e(\omega^P))\right] D^P(\omega^P|e(\omega^P) \vee e(\omega_0^A)). \end{aligned} \tag{17}$$

Proof of Lemma 7. Using Condition (13), we have that, given the state $\mathcal{S}_0 \vee e(\omega_M^A)$ of the decision problem, and $\omega^S \in \Omega^S$,

$$\begin{aligned} D^S(\omega^S|\mathcal{S}_0 \vee e(\omega_M^A)) = \mathcal{V}(\mathcal{S}_0) + \int_0^M \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|\mathcal{S}_0\right] \mathrm{d}v + (1-\delta)\mathbb{E}\left[c|\omega^S\right] \\ - \delta\mathbb{E}^{\chi^*}\left[\mathcal{V}(S_0 \backslash e(\omega^S) \vee e(\tilde{\omega}^S) \vee W^P(\tilde{\omega}^S)) + \int_0^M \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|S_0 \backslash e(\omega^S) \vee e(\tilde{\omega}^S) \vee W^P(\tilde{\omega}^S))\right] \mathrm{d}v|\omega^S\right], \end{aligned} \tag{18}$$

where the equality follows from combining (12) with (13). Similarly,

$$\begin{aligned} D^S(\omega^S|e(\omega^S) \vee e(\omega_M^A)) = \mathcal{V}(e(\omega^S)) + \textstyle\int_0^M \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|e(\omega^S)\right] \mathrm{d}v + (1-\delta)\mathbb{E}\left[c|\omega^S\right] \\ -\delta\mathbb{E}^{\chi^*}\left[\mathcal{V}(e(\tilde{\omega}^S) \vee W^P(\tilde{\omega}^S)) + \textstyle\int_0^M \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|e(\tilde{\omega}^S) \vee W^P(\tilde{\omega}^S))\right] \mathrm{d}v|\omega^S\right]. \end{aligned} \tag{19}$$

Differentiating (18) and (19) with respect to M, using the independence across alternatives and search and Lemma 5, we have that

$$\frac{\partial}{\partial M} D^S(\omega^S|\mathcal{S}_0 \vee e(\omega_M^A)) = \mathbb{E}^{\chi^*}\left[\delta^{\kappa(M)}|S_0 \backslash e(\omega^S)\right] \frac{\partial}{\partial M} D^S(\omega^S|e(\omega^S) \vee e(\omega_M^A)). \tag{20}$$

That is, the improvement in $D^S(\omega^S|\mathcal{S}_0 \vee e(\omega_M^A))$ that originates from a slight increase in the value of the auxiliary option M is the same as in a setting with only search and the auxiliary

option, $D^S(\omega^S|e(\omega^S) \vee e(\omega^A_M))$, discounted by the expected time it takes (under the index rule χ^*) until there are no indexes with value strictly higher than M, in an environment without search where the CS is the same as the one specified in $\mathcal{S}_0$. Similar arguments imply that, for any $\omega^P \in \{\hat{\omega}^P \in \Omega^P : \mathcal{S}_0(\hat{\omega}^P) > 0\}$,

$$\frac{\partial}{\partial M} D^P(\omega^P|\mathcal{S}_0 \vee e(\omega^A_M)) = \mathbb{E}^{\chi^*}\left[\delta^{\kappa(M)}|S_0\backslash e(\omega^P)\right] \frac{\partial}{\partial M} D^P(\omega^P|e(\omega^P) \vee e(\omega^A_M)). \tag{21}$$

Let $M^* \equiv \max\{\mathcal{I}^*(\mathcal{S}^P_0), \mathcal{I}^S(\omega^S)\}$. Integrating (20) over the interval $(0, M^*)$ of possible values for the auxiliary option and rearranging, we have that

$$\begin{aligned} D^S(\omega^S|\mathcal{S}_0 \vee e(\omega^A_0)) &= D^S(\omega^S|\mathcal{S}_0 \vee e(\omega^A_{M^*})) - \int_0^{M^*} \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|S_0\backslash e(\omega^S))\right] \frac{\partial}{\partial v} D^S(\omega^S|e(\omega^S) \vee e(\omega^A_v))\mathrm{d}v \\ &= D^S(\omega^S|\mathcal{S}_0 \vee e(\omega^A_{M^*})) - D^S(\omega^S|e(\omega^S) \vee e(\omega^A_{M^*})) \\ &\quad + \mathbb{E}^{\chi^*}\left[\delta^{\kappa(0)}|S_0\backslash e(\omega^S))\right] D^S(\omega^S|e(\omega^S) \vee e(\omega^A_0)) \\ &\quad + \int_0^{M^*} D^S(\omega^S|e(\omega^S) \vee e(\omega^A_v))\mathrm{d}\mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|S_0\backslash e(\omega^S))\right], \end{aligned}$$

where the second equality follows from integration by parts and from the fact that

$$\mathbb{E}^{\chi^*}\left[\delta^{\kappa(M^*)}|S_0\backslash e(\omega^S)\right] = 1.$$

That the outside option has value normalized to zero also implies that $D^S(\omega^S|\mathcal{S}_0 \vee e(\omega^A_0)) = D^S(\omega^S|\mathcal{S}_0)$. It is also easily verified that $D^S(\omega^S|\mathcal{S}_0 \vee e(\omega^A_{M^*})) = D^S(\omega^S|e(\omega^S) \vee e(\omega^A_{M^*}))$. This follows immediately from the observation that $\mathcal{V}(\mathcal{S}_0 \vee e(\omega^A_{M^*})) = \mathcal{V}(e(\omega^S) \vee e(\omega^A_{M^*})) = M^*$, and similarly

$$\mathbb{E}^{\chi^*}\left[\mathcal{V}\left(\mathcal{S}_0\backslash e(\omega^S) \vee e(\tilde{\omega}^S) \vee W^P(\tilde{\omega}^S) \vee e(\omega^A_{M^*})\right)|\omega^S\right] = \mathbb{E}^{\chi^*}\left[\mathcal{V}\left(e(\tilde{\omega}^S) \vee W^P(\tilde{\omega}^S) \vee e(\omega^A_{M^*})\right)|\omega^S\right].$$

Intuitively, under the index policy, any alternative with index strictly below M^* is never explored given the presence of an auxiliary alternative with payoff M^*. Therefore, we have that

$$\begin{aligned} D^S(\omega^S|\mathcal{S}_0) &= \int_0^{M^*} D^S(\omega^S|e(\omega^S) \vee e(\omega^A_v))\mathrm{d}\mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|S_0\backslash e(\omega^S)\right] \\ &\quad + \mathbb{E}^{\chi^*}\left[\delta^{\kappa(0)}|S_0\backslash e(\omega^S))\right] D^S(\omega^S|e(\omega^S) \vee e(\omega^A_0)). \end{aligned} \tag{22}$$

Similar arguments imply that

$$\begin{aligned} D^P(\omega^P|\mathcal{S}_0) &= \int_0^{M^*} D^P(\omega^P|e(\omega^P) \vee e(\omega^A_v))\mathrm{d}\mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|S_0\backslash e(\omega^P)\right] \\ &\quad + \mathbb{E}^{\chi^*}\left[\delta^{\kappa(0)}|S_0\backslash e(\omega^P))\right] D^P(\omega^P|e(\omega^P) \vee e(\omega^A_0)). \end{aligned} \tag{23}$$

To complete the proof of Lemma 7, we consider separately two cases. Case (1): given $\mathcal{S}_0$, χ^* specifies starting by exploring a physical alternative (i.e., $M^* = \mathcal{I}^*(\mathcal{S}_0^P)$). Then Condition (16) in the lemma follows directly from (22). Thus consider Condition (17). First observe that, for any state $\omega^P \in \Omega^P$ such that $M^* > \max\{\mathcal{I}^*(\mathcal{S}_0^P \backslash e(\omega^P)), \mathcal{I}^S(\omega^S)\}$, we have that $M^* = \mathcal{I}^P(\omega^P)$, in which case $D^P(\omega^P|\mathcal{S}_0) = D^P(\omega^P|e(\omega^P) \vee e(\omega_0^A)) = 0$ and the integrand $D^P(\omega^P|e(\omega^P) \vee e(\omega_v^A))$ in (23) is equal to zero over the interval $[0, \mathcal{I}^P(\omega^P)]$ and hence also over the interval $[0, \max\{\mathcal{I}^*(\mathcal{S}_0^P \backslash e(\omega^P)), \mathcal{I}^S(\omega^S)\}]$. We thus have that, in this case, Condition (17) clearly holds. Next observe that, for any state $\omega^P \in \Omega^P$ such that $M^* = \max\{\mathcal{I}^*(\mathcal{S}_0^P \backslash e(\omega^P)), \mathcal{I}^S(\omega^S)\}$, Condition (17) follows directly from (23).

Case (2): given $\mathcal{S}_0$, χ^* specifies starting with search (i.e., $M^* = \mathcal{I}^S(\omega^S)$). Then, for any $\omega^P \in \Omega^P$, $\max\{\mathcal{I}^*(\mathcal{S}_0^P \backslash e(\omega^P)), \mathcal{I}^S(\omega^S)\} = M^*$, in which case Condition (17) in the lemma follows directly from (23). That Condition (16) also holds follows from the fact that, in this case, $D^S(\omega^S|\mathcal{S}_0) = D^S(\omega^S|e(\omega^S) \vee e(\omega_0^A)) = 0$ and the integrand $D^S(\omega^S|e(\omega^S) \vee e(\omega_v^A))$ in (22) is equal to zero over the entire interval $\left[0, \max\{\mathcal{I}^*(\mathcal{S}_0^P \backslash e(\omega^P)), \mathcal{I}^S(\omega^S)\}\right]$. □

Step 2. Using the characterization of the payoff differentials in Lemma 7, we now establish that the average per-period payoff under χ^* solves the Bellman equation for our dynamic optimization problem. Let $\mathcal{V}^*(\mathcal{S}_0) \equiv (1-\delta)\sup_{\chi \in \mathcal{X}} \mathbb{E}^\chi \left[\sum_{t=0}^\infty \delta^t U_t | \mathcal{S}_0\right]$ denote the value function for the dynamic optimization problem.

Lemma 8. *For any state of the decision problem $\mathcal{S}_0$, with ω^S denoting the state of the search technology as specified under $\mathcal{S}_0$,*

1. *$\mathcal{V}(\mathcal{S}_0) \geq V^S(\omega^S|\mathcal{S}_0)$, and $\mathcal{V}(\mathcal{S}_0) = V^S(\omega^S|\mathcal{S}_0)$ if and only if $\mathcal{I}^S(\omega^S) \geq \mathcal{I}^*(\mathcal{S}_0^P)$;*
2. *for any $\omega^P \in \{\hat{\omega}^P \in \Omega^P : \mathcal{S}_0(\hat{\omega}^P) > 0\}$, $\mathcal{V}(\mathcal{S}_0) \geq V^P(\omega^P|\mathcal{S}_0)$, and $\mathcal{V}(\mathcal{S}_0) = V^P(\omega^P|\mathcal{S}_0)$ if and only if $\mathcal{I}^P(\omega^P) = \mathcal{I}^*(\mathcal{S}_0^P) \geq \mathcal{I}^S(\omega^S)$.*

Hence, for any $\mathcal{S}_0$, $\mathcal{V}(\mathcal{S}_0) = \mathcal{V}^*(\mathcal{S}_0)$, and χ^* is optimal.

Proof of Lemma 8. *Part (1)*. First, use (14) to note that, for all $v \geq 0$, $D^S(\omega^S|e(\omega^S) \vee e(\omega_v^A)) \geq 0$, with the inequality holding as an equality if and only $v \leq \mathcal{I}^S(\omega^S)$. Therefore, from (16), $D^S(\omega^S|\mathcal{S}_0) \geq 0$ – and hence $\mathcal{V}(\mathcal{S}_0) \geq V^S(\omega^S|\mathcal{S}_0)$ – with the inequality holding as an equality if and only if $\mathcal{I}^*(\mathcal{S}_0^P) \leq \mathcal{I}^S(\omega^S)$.

Part (2). Similarly, use (15) to observe that for any $\omega^P \in \{\hat{\omega}^P \in \Omega^P : \mathcal{S}_0^P(\hat{\omega}^P) > 0\}$ and any $v \geq 0$, $D^P(\omega^P|e(\omega^P) \vee e(\omega_v^A)) \geq 0$, with the inequality holding as an equality if and only if $0 \leq v \leq \mathcal{I}^P(\omega^P)$. Therefore, from (17), $D^P(\omega^P|\mathcal{S}_0) \geq 0$ with the inequality holding as equality if and only if $\mathcal{I}^P(\omega^P) \geq \max\{\mathcal{I}^*(\mathcal{S}_0^P \backslash e(\omega^P)), \mathcal{I}^S(\omega^S)\}$. The result in Part (2) of the lemma then follows from the fact that the last inequality holds if and only if $\mathcal{I}^P(\omega^P) = \mathcal{I}^*(\mathcal{S}_0^P) \geq \mathcal{I}^S(\omega^S)$.

Next, note that, jointly, Conditions 1 and 2 in the lemma imply that

$$\mathcal{V}(\mathcal{S}_0) = \max\left\{V^S(\omega^S|\mathcal{S}_0), \max_{\omega^P \in \{\hat{\omega}^P \in \Omega^P : \mathcal{S}_0^P(\hat{\omega}^P) > 0\}} V^P(\omega^P|\mathcal{S}_0)\right\}.$$

Hence $\mathcal{V}$ solves the Bellman equation. That $\delta^T \mathbb{E}^\chi\left[\sum_{s=T}^\infty \delta^s U_s|\mathcal{S}\right] \to 0$ as $T \to \infty$ guarantees that $\mathcal{V}(\mathcal{S}_0) = \mathcal{V}^*(\mathcal{S}_0)$, and hence the optimality of the index policyχ^*. □

This completes the proof of Part (1), and hence Theorem 1. ■

Proof of Lemma 1. Observe that, under each of the two objectives (i.e., for $x = W, PR$), for any state of the mechanism $\mathcal{S} = (\omega^S, \mathcal{S}^P)$, a firm with cost c joining the mechanism with a bid equal to b obtains an expected payoff equal to

$$(b - c)Q_x(b; \mathcal{S}^\wedge(b)) + T_x(b; \mathcal{S}^\wedge(b))$$

when, as a consequence of the firm's joining, the state of the mechanism is $\mathcal{S}^\wedge(b)$. Because $Q_x(b; \mathcal{S}^\wedge(b))$ is decreasing in b, standard results in mechanism design imply that, irrespective of $\mathcal{S}$, joining the mechanism and bidding $b = c$ is weakly dominant for the firm.

That the procurement mechanism with transfers T_W is welfare maximizing follows directly from Theorem 1. That the procurement mechanism with transfers T_{PR} is profit maximizing follows from the following observations. Under any (interim) IC and IR mechanism, expected profits are equal to expected dynamic virtual surplus. The latter is equal to expected total welfare when the buyer's cost of procuring from each firm with cost c is equal to $\gamma(c)$. The result then follows from the same arguments as for welfare maximization. ■

Proof of Lemma 2. Denote by F_{ω^S} the cdf of the random variable describing the total number of firms in the market that have not been identified yet, when the state of the search technology is ω^S, with the latter describing, for each past search and each bid b, the total number of firms joining as the result of that search with a bid equal to b. Let f_{ω^S} denote the corresponding probability distribution function. That is, for any $l \in \mathbb{N}$, $f_{\omega^S}(l)$ is the probability that the number of firms in the market that remain to be identified is l. Note that $f_{\omega^S}(l)$ depends on ω^s only through the total number m of past searches and the total number n of firms identified through these searches. Hereafter, we denote by $f^{n,m}(l)$ the probability that the number of remaining firms is l given (n, m) and observe that

$$f^{n,m}(l) \propto f(l + n)\binom{l + n}{n}(1 - \rho)^{lm}. \tag{24}$$

We denote by $F^{n,m}$ the corresponding cdf.

Part (1) of the lemma follows from the above observation along with the fact that each

expansion identifies each firm in the market that has not been identified yet independently of the outcome of any other search and of the firms' costs.

Before turning to Parts (2) and (3), we establish the following claim.

Claim 1. If the prior F is UL-concave (respectively, UL-convex), then, for any (n,m), $F^{n,m}$ is UL-concave (respectively, UL-convex).

Proof of Claim 1. For any n and m, any l, let

$$L^{n,m}(l) \equiv \frac{(l+1)f^{n,m}(l+1)}{f^{n,m}(l)}$$

and observe that

$$L^{n,m}(l) = \frac{(l+1)f(l+1+n)\binom{l+1+n}{n}(1-\rho)^{(l+1)m}}{f(l+n)\binom{l+n}{n}(1-\rho)^{lm}} = L(l+n)(1-\rho)^m,$$

Hence, $L^{n,m}(l)$ is increasing (respectively, decreasing) in l if and only if L is increasing (respectively, decreasing). □

To establish Part 2, we show that for any prior F, fixing the number of firms n in the CS, $F^{n,m}$ is decreasing in m in the MLR order (and hence also in the FOSD order). To see that $F^{n,m} \succ F^{n,m+1}$, use (24) to observe that

$$\frac{f^{n,m}(l)}{f^{n,m+1}(l)} = h(n,m) \cdot \frac{f(l+n)\binom{l+n}{n}(1-\rho)^{lm}}{f(l+n)\binom{l+n}{n}(1-\rho)^{l(m+1)}} = h(n,m) \cdot \frac{1}{(1-\rho)^l},$$

where $h(n,m)$ is a function independent of l. Hence, $f^{n,m}(l)/f^{n,m+1}(l)$ is increasing in l, as claimed. Finally, the fact that the posterior $F^{n,m}$ becomes worse (in the MLR and hence FOSD sense) as m increases implies that $\mathcal{I}^S(m,n)$ is decreasing in m. This follows from the recursive structure of the search index (Theorem 1, Part (2)). In particular, recall that the search index maximizes the expected discounted payoff per unit of expected discounted time, which is clearly higher the better the distribution $F^{n,m}$ (in the sense of MLR, and hence, FOSD), i.e., the higher the number of firms that the next search is expected to identify.

To establish Part (3), we now show that the search index $\mathcal{I}^S(m,n)$ is decreasing (respectively, increasing) in n if F is UL-concave (respectively, UL-convex). Fix $m \in \mathbb{N}$. Suppose F is UL-concave. Then $F^{n,m} \succ F^{n+1,m}$ because

$$\frac{f^{n,m}(l)}{f^{n+1,m}(l)} = \tilde{h}(n,m) \cdot \frac{f(l+n)\binom{l+n}{n}(1-\rho)^{lm}}{f(l+n+1)\binom{l+n+1}{n+1}(1-\rho)^{lm}} = \tilde{h}(n,m) \cdot \frac{n+1}{L(l+n)},$$

where $\tilde{h}(n,m)$ is a function independent of l. Since F is UL-concave, $L(l+n)$ is decreasing in l, which means $f^{n,m}(l)/f^{n+1,m}(l)$ is increasing in l. Hence, $F^{n,m} \succ F^{n+1,m}$. This also means

that, for any $m \in \mathbb{N}$ and any $n, n' \in \mathbb{N}$ with $n' > n$, $F^{n,m} \succ F^{n',m}$. Arguments similar to those establishing Part (2) then imply that the search index is decreasing in n. An analogous argument establishes that, when F is UL-convex, for any $m \in \mathbb{N}$ and $n, n' \in \mathbb{N}$, with $n' > n$, $F^{n',m} \succ F^{n,m}$. Hence, in this case, the search index is increasing in n. ■

Proof of Lemma 3. Observe that, given the state $\omega^P = (c, \vartheta^{s-1})$ of each firm in the mechanism with cost c (revealed through bidding), the firm's index is

$$\mathcal{I}_x^P(c, \vartheta^{s-1}) = \begin{cases} q - \gamma_x(c) & \text{if } \vartheta^{s-1} \text{ s.t. } \vartheta_l = 1 \text{ for some } l \leq s-1 \\ q_x(c; s) & \text{otherwise,} \end{cases}$$

where, for any c, $\gamma_x(c) = c$ if $x = W$ and $\gamma_x(c) = \gamma(c)$ if $x = PR$, and where, for any c and any $a \in \mathbb{N}$, $q_x(c; a)$ is a function decreasing in both the firm's cost c and the number of times a the firm has been used without delivering a positive flow payoff. Clearly, $A_x(c; m, n) = \sup\{a : q_x(c; a) \geq \mathcal{I}_x^S(m, n)\}$.

Part (1) follows from the fact that $\mathcal{I}_x^S(m, n)$ is decreasing in m and $\mathcal{I}_x^P(c, \vartheta^{s-1})$ is decreasing in c and, when ϑ^{s-1} is such that $\vartheta_l = 0$ for all $l \leq s-1$, $\mathcal{I}_x^P(c, \vartheta^{s-1})$ is decreasing in s. Part (2) follows from the same properties together with the fact that $\mathcal{I}_x^S(m, n)$ is decreasing (respectively, increasing) in n when F is UL-concave (respectively, UL-convex). ■

Proof of Proposition 2. The result follows from Lemma 3. ■

Proof of Proposition 3. Because $\mathring{\mathcal{S}} = (\mathring{\omega}^S, \mathring{\mathcal{S}}^P)$ and $\mathcal{S} = (\omega^S, \mathcal{S}^P)$ are such that $\mathring{n} = n$ but $\mathring{m} > m$, by virtue of Lemma 2, $\mathcal{I}_x^S(\mathring{m}, \mathring{n}) < \mathcal{I}_x^S(m, n)$. On the other hand, for any $\omega^P = (c, \vartheta^{s-1})$, $\mathring{\mathcal{S}}^P(\omega^P) = \mathcal{S}^P(\omega^P)$, meaning that the composition of the CS is the same under both states $\mathring{\mathcal{S}}$ and $\mathcal{S}$. The latter property implies that the index $\mathcal{I}_x^P$ of each firm in the CS is the same under $\mathring{\mathcal{S}}$ and $\mathcal{S}$. Lastly, observe that the arguments in the proof of Lemma 2 imply that the distribution describing the outcome of future expansions is less favorable starting from $\mathring{\omega}^S$ than starting from ω^S (future expansions are expected to identify fewer firms, in the MLR order, whereas the cost of each joining firm is drawn independently of the number of firms identified). Jointly, the above properties, together with the fact that, for both $x = W$ and $x = PR$, the buyer's policy is the index one of Theorem 1, imply that $Q_x(c; \mathring{\mathcal{S}}) \geq Q_x(c; \mathcal{S})$. That is, the expected discounted number of times the firm supplies is greater when the state of the procurement mechanism at the end of the period the firm joins is $\mathring{\mathcal{S}}$ than when it is $\mathcal{S}$. That the firm's expected profits are also higher under $\mathring{\mathcal{S}}$ than under $\mathcal{S}$ follows from this property together with the fact that the firm's profits are $T_x(c; \mathring{\mathcal{S}})$ in the former case and $T_x(c; \mathcal{S})$ in the latter case, for $x = W, PR$. ■

Proof of Lemma 4. As established in the proof of Lemma 3,

$$A_x(c; m, n) = \sup\{a : q_x(a, c) \geq \mathcal{I}_x^S(m, n)\},$$

where $q_x(a, c)$ is the index of a firm with cost c that, after being utilized a times, never delivered a satisfactory service.

Observe that $q_W(a, c) - q_{PR}(a, c) = G(c)/g(c)$. Furthermore, $\mathcal{I}_{PR}^S(m, n) = \mathcal{I}_W^S(m, n) - y(G, m, n)$, where $y(G, m, n)$ is a function of the distribution G and of (m, n).

For small c, $q_W(a, c) - q_{PR}(a, c)$ is small, and hence $q_W(a, c) - \mathcal{I}_W^S(m, n) > 0$ implies $q_{PR}(a, c) - \mathcal{I}_{PR}^S(m, n) > 0$. Because q_W and q_{PR} are decreasing in a, $A_{PR}(c; m, n) > A_W(c; m, n)$. Furthermore, because $q_W(a, c) - q_{PR}(a, c)$ is increasing in c while $\mathcal{I}_W^S(m, n) - \mathcal{I}_{PR}^S(m, n)$ is invariant in c, there exists a $\hat{c}$ such that

$$\big(q_W(a, c) - q_{PR}(a, c) - \big(\mathcal{I}_W^S(m, n) - \mathcal{I}_{PR}^S(m, n)\big)\big) \cdot (c - \hat{c}(m, n)) \geq 0.$$

Note that, by the recursive properties of the search index in Part (2) of Theorem 1,

$$0 < \mathcal{I}_W^S(m, n) - \mathcal{I}_{PR}^S(m, n) < \frac{G(\bar{c})}{g(\bar{c})}.$$

Hence, the equation

$$q_W(a, c) - q_{PR}(a, c) - \big(\mathcal{I}_W^S(m, n) - \mathcal{I}_{PR}^S(m, n)\big) = 0$$

admits a solution $\hat{c}(m, n) \in (\underline{c}, \bar{c})$.

Suppose $c < \hat{c}(m, n)$. Then $q_W(a, c) - q_{PR}(a, c) < \mathcal{I}_W^S(m, n) - \mathcal{I}_{PR}^S(m, n)$, which means that if $q_W(a, c) - \mathcal{I}_W^S(m, n) \geq 0$ then $q_{PR}(a, c) - \mathcal{I}_{PR}^S(m, n) \geq 0$. This means that $\Delta(c; m, n) \leq 0$.

Next, suppose $c > \hat{c}(m, n)$. Then $q_W(a, c) - q_{PR}(a, c) > \mathcal{I}_W^S(m, n) - \mathcal{I}_{PR}^S(m, n)$, which means that if $q_{PR}(a, c) - \mathcal{I}_{PR}^S(m, n) \geq 0$ then $q_W(a, c) - \mathcal{I}_W^S(m, n) \geq 0$. Hence $\Delta(c; m, n) \geq 0$. ■

Proof of Proposition 4. The result follows from Lemma 4. ■