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

Daniel Fershtman† Alessandro Pavan‡

November 2021

Abstract

We study the problem of a decision maker alternating between exploring existing alternatives in the consideration set and searching for new ones. We characterize the optimal policy and identify implications for search and exploration dynamics. When the search technology is stationary, or improves over time, search is equivalent to replacement. With deteriorating technologies, instead, alternatives are revisited after search is launched and each expansion is treated as if it were the last one. We show that the comparative statics with an endogenous consideration set can be qualitatively different from those with an exogenous one. For example, improvements in a category of alternatives may lead to a reduction in the category's usage as well as in the eventual selection of an alternative from that category. We apply the analysis to the administration of medical treatments, clinical trials toward regulatory approval, Weitzman (1979)'s Pandora's boxes problem, and online consumer search.

Keywords: Experimentation, Search, Learning, Endogenous Consideration Sets

*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, including University of Bonn, Cornell University, Hebrew University of Jerusalem, HEC Paris, Paris Dauphine, University of Pennsylvania, Princeton University, Stanford GSB, Tel-Aviv University, Toulouse School of Economics, Yale University, Warwick Economic Theory Workshop, 3rd Columbia Conference on Economic Theory (CCET), and the 30th Stony Brook International Conference on Game Theory. Dirk Bergemann, Eddie Dekel, David Dillenberger, Laura Doval, Kfir Eliaz, Stephan Lauermann, Benny Moldovanu, Philip Reny, Eran Shmaya, Andy Skrzypacz, Rani Spiegler, Bruno Strulovici, Asher Wolinsky, and Jidong Zhou provided excellent comments and suggestions. We also thank Matteo Camboni and Tuval Danenberg for outstanding research assistance. Fershtman gratefully acknowledges funding by the German Research Foundation (DFG) through CRC TR 224 (Project B01) and from ISF grant #1202/20. The usual disclaimer applies.

†Eitan Berglas School of Economics, Tel Aviv University. Email: danielfer@tauex.tau.ac.il

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

1 Introduction

Classic models of sequential experimentation or learning involve a decision maker (hereafter, DM) exploring a *fixed* set of alternatives with unknown returns. Yet, a ubiquitous feature of many dynamic decision problems is that the set of alternatives a DM can explore is expanded over time, in response to the information gathered by exploring the alternatives already in the consideration set (hereafter, CS).

In this paper, we study the tradeoff between the exploration of alternatives already in the CS and the expansion of the latter through search for additional alternatives. A key difference between exploration and expansion is the direct-vs-indirect nature of the two activities. When an alternative is in the CS, the DM can "point to it," that is, she can choose to explore that particular alternative instead of others. When, instead, an alternative is outside the CS, the DM cannot point to it, meaning that she cannot choose to explore that specific alternative instead of others.[1] This inability may reflect natural randomness in the search process, which may bring to the CS alternatives different from those the DM was looking for. Alternatively, search may bring more than a single alternative and such batching may have important implications for the decision to expand the CS in the first place. Finally, the DM may have limited knowledge about the alternatives outside her CS, and/or her ability to bring new alternatives to it.

To study the tradeoff between exploration of alternatives already in the CS and expansion of the latter, we consider a generalization of the classic multi-armed bandit problem in which the set of "arms" is endogenous. Exploring an alternative already in the CS (pulling an arm) yields a flow payoff and generates information (for example, about the distribution from which the flow payoff is drawn). Searching for new "arms" (that is, choosing to expand the CS) is costly and brings a random set of new alternatives (i.e., of arms).

We characterize the solution to the above problem and show that it takes the form of an "index" policy. Each alternative in the CS is assigned a history-dependent number that is a function only of the state of that alternative. This number (the arm's "index") is the same as in Gittins and Jones' (1974) original work on bandit problems with an exogenous set of arms. Search (that is, the decision to expand the CS) is also assigned an index, which depends only on the state of the search technology. Crucially, the search index does not depend on the information generated by the exploration of any of the alternatives already in the CS. It also differs from the value the DM attaches to the expansion of the CS but is linked to the indexes of the new alternatives the DM expects to find through current and future searches. The optimal policy consists in selecting at any period the alternative for which the index is the highest. We establish the optimality of such a policy through a proof based on a recursive characterization of the index

[1] Likewise, the DM cannot choose to bring a specific alternative from outside of the CS into the CS: If she could, there would be no distinction between exploring alternatives inside and outside the CS, making the latter irrelevant.

for search that also permits us to uncover various properties of the exploration and expansion dynamics under the optimal policy. We then show how such a characterization facilitates novel comparative statics relevant for applications.

At any point in time, the decision to expand the CS depends on the composition of the current CS only through (a) the state of the alternative with the highest index, and (b) the state of the search technology. This result holds despite the fact that the opportunity cost of searching for new alternatives (which is linked to the value of continuing with the current CS) depends on the entire composition of the current CS. Similarly, conditional on forgoing search in a given period, the decision of which alternative to explore in the current CS is independent of the search technology, despite the fact that search may bring alternatives that are more similar to certain alternatives currently in the CS than others.[2] If the search technology is stationary, or improving, in a sense made precise below, then alternatives in the CS at the time of its expansion never receive attention in the future, and hence are effectively discarded once the CS is expanded. Each search is then equivalent to replacement of the current CS with a new one. When, instead, the search technology deteriorates over time (e.g., because the DM becomes pessimistic about the possibility of finding attractive new alternatives), the alternatives in the current CS are put on hold and may be revisited after the CS is expanded. Furthermore, in this case, the decision to expand the CS is made as if there will be no further expansions after the current one.

The analysis can be applied to a broad class of experimentation and sequential learning problems. We consider a few in the paper: administration of medical treatments, clinical trials toward regulatory approval, Weitzman's (1979) "Pandora's boxes problem" with an endogenous set of boxes, and online consumer search.

Administration of medical treatments. One of the most prominent applications of the multi-armed bandit model is the administration of medical treatments.[3] A physician must choose in each period which medical treatment to administer, trading off the well-being of the patients receiving the treatments with the value of generating information about the treatments' efficacy useful for future patients. Each treatment belongs to a different category with treatments from the same category sharing some core characteristics. We enrich this problem by giving the physician the possibility to expand her CS by searching for new treatments to administer.

We use the model to study the effects of an improvement in a category of treatments on the administration of such treatments. The improvement may take the form of an increase in the

[2] These properties can be seen as a generalization of the IIA (independence of irrelevant alternatives) property of classic multi-armed bandit problems. What makes this problem different from the classic one enriched with a "meta" arm that comprises all the alternatives brought to the CS by search is that the evaluation of such a "meta" arm requires knowing how to subsequently explore the arms that search brings to the CS, which is what is investigated in the first place. Furthermore, dynamic problems with "meta" arms rarely admit an index solution. See also the discussion in footnote 22.

[3] See, e.g., Berry (2006), Berry and Fristedt (1985), Dickstein (2014), Katehakis and Derman (1986), Villar, Bowden and Wason (2015), and the FDA's "Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials".

ex-ante probability that the treatments are effective, or an increase in the physician's payoff from administering such treatments. Perhaps surprisingly, we show that, when the CS is endogenous, such improvements are not guaranteed to boost the administration of the improved treatments. In fact, they may even reduce it. The reason is that such improvements also increase the benefit of expanding the CS (as measured by the search index). When strong enough, this novel effect may induce the physician to administer fewer treatments from the improved category, not only relative to other treatments, but overall.

Clinical trials towards regulatory approval. Consider a pharmaceutical company seeking to persuade the FDA, or an equivalent regulatory authority, to approve one of its drugs. The safety of such drugs is unknown ex-ante, both to the firm and to the regulatory authority. At each period, the firm can conduct a public experiment on one of its products, or conduct basic research aimed at identifying new products to subsequently test for approval. The approval of a product requires a sufficiently large number of satisfactory outcomes. The firm can bring a product to the market only once it is approved. The firm's objective is to maximize the expected discounted profit from the sale of its products, net of all the experimentation and search costs.

The problem described above is a multi-product version of the one in Henry and Ottaviani (2019), with the additional feature that the set of products the firm can experiment with is endogenously expanded over time, based on the outcomes of past experiments.

We use the model to study the effects of changes in regulatory standards. We show that, when the approval standards for a category become more lenient, the firm also responds by investing more in basic research. The resulting expansion of the CS may crowd out a more thorough experimentation of the products whose standards have been lowered, possibly reducing their approval rate. To clarify, we are not claiming that such "crowding out" necessarily happens, but that the endogeneity of the CS may explain why this may happen.

Pandora's problem with an endogenous set of boxes. Our third application is an extension of Weitzman (1979)'s classic "Pandora's boxes problem" to a setting with an endogenous set of boxes. In Weitzman (1979)'s problem, the DM faces an exogenous set of boxes, each containing a prize of unknown value drawn from a known distribution. Opening a box reveals its prize, is costly (with the cost box-specific), and is necessary to collect its prize. The DM can collect only one prize among any of the opened boxes, and must choose the optimal sequence of inspections as well as when to stop. We extend this problem by allowing the DM to sequentially choose between inspecting existing boxes and searching for additional ones. Our solution generalizes Weitzman (1979)'s by introducing an appropriate reservation price for search.

Online consumer search and endogenous click-through-rates. We consider the problem of a consumer alternating between (a) reading new ads (bringing the corresponding products to the consumer's CS), (b) clicking on the ads of those products already in the CS (revealing the products' value to the consumer), and (c) finalizing a purchase with one of the visited vendors.

We use our results to endogenize the click-through rates (CTRs)—that is, the probability that an ad is clicked upon, as a function of the order by which the search engine displays the ads. The models that have been used to study such a problem typically assume that the consumer clicks on the ads in the order they are displayed, and that CTRs are either invariant to positions or depend on the latter in an exogenous fashion. Such assumptions do not seem to square well with empirical findings.[4] We characterize the CTRs and show that they need not be monotone in the ads' positions, even when the purchasing probabilities are. We establish this result by first deriving an "eventual-purchase theorem" in the spirit of Choi, Dai and Kim (2018) that relates the probability each product is purchased to the primitives of the problem (realized values, ads' positions, and search technology), but accounting for the endogeneity of the consumer's CS (the latter is exogenous in Choi, Dai and Kim, 2018).[5]

Finally, we use the model to illustrate why a firm that is already advertising on a platform's first page may experience a decline in its profits when the probability it displays an additional product's ad on the second page increases. This can happen even if the extra ad is unambiguously profitable when brought exogenously to the consumer's CS.

The rest of the paper is organized as follows. The remainder of this section briefly discusses the most pertinent literature. Section 2 introduces the model. Section 3 characterizes the optimal policy and identifies key properties for the dynamics of experimentation and expansion of the CS. Section 4 contains the applications mentioned above. Section 5 discusses a few extensions whereas Section 6 concludes. All proofs are either in the Appendix at the end of the document or in the Online Supplement.

1.1 Related literature

The paper is part of a fast-growing literature on CSs.[6] Eliaz and Spiegler (2011) study implications of different CSs on firms' behavior, assuming such sets are exogenous. Manzini and Mariotti (2014) and Masatlioglu, Nakajima, and Ozbay (2012), instead, identify CSs from choice behavior. Caplin, Dean, and Leahy (2018) provide necessary and sufficient conditions for rationally-inattentive agents to focus on a subset of all available choices, thus endogenizing the CSs. Simon (1955) considers a sequential search model, in which alternatives are examined until a "satisfying" alternative is found. Caplin, Dean, and Martin (2011) show that the rule in Simon (1955) can be viewed as resulting from an optimal procedure when there are information costs. Our analysis complements the one in this literature by providing a dynamic micro-foundation for endogenous CSs. Rather than committing to a CS up front and proceeding to evaluate its alternatives, the

[4] See, for example, Jeziorski and Segal (2015).

[5] See also Armstrong and Vickers, 2015 and Armstrong, 2017 for related models with an exogenous CS.

[6] For the earlier marketing literature, see, e.g., Hauser and Wernerfelt (1990) and Roberts and Lattin (1991). For a survey of recent developments, see Honka et al (2019).

DM gradually expands the CS, in response to the results obtained from the exploration of the alternatives in the set.

The paper is also related to the literature on experimentation and sequential learning. Most closely related are Austen-Smith and Martinelli (2018), Fudenberg, Strack and Strzalecki (2018), Gossner, Steiner and Stewart (2019), Ke and Villas-Boas (2019) and Ke, Shen, and Villas-Boas (2016), which study the problem of a DM that gradually acquires costly information about a set of options before stopping and choosing one of them. Related are also Che and Mierendorff (2019), who study the optimal sequential allocation of attention to two different signal sources biased towards alternative actions, and Liang, Mu, and Syrgkanis (2019), who study the dynamic acquisition of information about an unknown Gaussian state. In all of these papers, the set of alternatives is fixed ex-ante. In our model, instead, the DM expands the CS over time in response to the information she collects about the alternatives already in it.

Related is also Garfagnini and Strulovici (2016), which studies how successive (forward-looking) agents experiment with endogenous technologies. Trying "radically" new technologies reduces the cost of experimenting with similar technologies, which effectively expands the set of affordable technologies.[7] Schneider and Wolf (2019) study the time-risk tradeoff of an agent who wishes to solve a problem before a deadline, and allocates her time between implementing a given method and developing (and then implementing) a new one. While, at a high level, the problems examined in the last two papers bear a resemblance to ours in that they also consider environments in which the set of alternatives expands over time, both the models and the questions addressed are different.

In Fershtman and Pavan (2020), we study the effects of "soft" affirmative action on minority recruitment, in a setting in which the candidate pool is endogenous. While that paper studies the effects of changes in the search technology on the selection of minority candidates, the analysis in Subsections 4.1 and 4.2 in the present paper focuses on the effects of variations in a category's attractiveness on the usage (and ultimate selection) of its alternatives.

As anticipated above, the application to clinical trials and regulatory approval in Subsection 4.2 is related to Henry and Ottaviani (2019), whereas the application in Subsection 4.3 extends Weitzman (1979) to a setting with an endogenous set of boxes.[8] Finally, the application to online consumer search in Subsection 4.4 is related to independent work by Greminger (2021). While that paper focuses on the comparison between direct and indirect search, we endogenize the click-through-rates and study the effects of additional ad space on firms' profits.[9]

[7] Technologies are interdependent in their environment. In particular, a radically new technology is informative about the value of similar technologies.

[8] Despite its many applications, relatively few extensions of Weitzman's problem have been studied in the literature. Notable exceptions include Olszewski and Weber (2015), Choi and Smith (2016), and Doval (2018). In these papers, though, the set of boxes is fixed.

[9] The model in that paper is a special version of the one in Subsection 4.4 in which payoffs are additively separable in an observable and an unobservable component.

Our characterization of the optimality of an index policy is related to the branching-arm literature (e.g., Weiss, 1988, Weber, 1992, and Keller and Oldale, 2003). This literature studies decision problems in which arms branch into new ones. Under certain conditions, the problem in the present paper can be viewed as a special case of the one considered in that literature. To the best of our knowledge, our proof of indexability is new and uncovers a novel recursive representation of the index of search which is key to the exploration and expansion dynamics that we identify as well as to the comparative statics that we study.

2 Model

In each period $t = 0, 1, 2, ...$, the DM chooses between exploring one of the alternatives within her CS and expanding the CS by searching for additional alternatives. Exploring an alternative yields a flow payoff and generates information about it. 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 $n_t \in \mathbb{N}$. 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 normalized 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 observable to the DM when the alternative is brought to the CS. 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 $m - 1 \in \mathbb{N}$ denote the number of past explorations, and $\vartheta^{m-1} \equiv (\vartheta_s)_{s=0}^{m-1}$ the history of past signal realizations, with $\vartheta_0 \equiv \emptyset$. When the DM explores the alternative for the m-th time, she receives an additional signal ϑ_m 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 $L_\xi(m; \mu)$ that does not depend on the calendar time t.

The search (expansion) technology. When 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. 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, with $(c_0, E_0) \equiv \emptyset$. The dependence of J on the history of past search outcomes allows us to capture, for example, learning about the effectiveness of search, as well as changes in the DM's ability to find new alternatives (e.g., learning by doing and/or fatigue).

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 $t \to \infty$.[10]

The model above describes an infinite-horizon experimentation problem (augmented by search) in which payoffs are accumulated alongside learning. In Sections 4.2-4.4 we also consider applications in which the DM sequentially decides between learning about alternatives in her CS and expanding the CS, until a final choice is made among the alternatives in the CS, ending the decision problem.

2.1 Example: Administration of medical treatments

As an illustration of the type of problems that the above formalism captures, consider the following extension of the classic problem of sequentially administering medical treatments. In each period, a physician must choose between administering a medical treatment among those in her CS, or expanding the latter by searching for new treatments (see, e.g., "How Physicians Can Keep Up with the Knowledge Explosion in Medicine", 2016, *Harvard Business Review*). Whenever the physician administers a treatment, she observes the outcome on the patient that receives it. The outcome yields a payoff to the physician – which may be linked to the well-being of the patient receiving the treatment – and is informative about the treatment's efficacy, which may be valuable for the good of future patients.

For simplicity, suppose there are two possible categories of treatments, indexed by $\xi \in \Xi = \{\alpha, \beta\}$. Ex-ante, treatments from the same category are identical. In keeping with the classic framework (e.g., Berry and Fristedt, 1985), the efficacy $\mu \in \{0, 1\}$ of a treatment is unknown ex-ante, with $\mu = 1$ in case the treatment is effective and $\mu = 0$ otherwise. Let $p^{\xi}(\emptyset)$ denote

[10] 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.

the ex-ante probability that a ξ-treatment is effective, with each μ drawn independently across treatments, conditional on their category (hence, in this example, the distribution Γ_ξ is Bernoulli with parameter $p^\xi(\emptyset)$). Using a treatment generates information about the treatment's efficacy. Specifically, when a treatment is administered for the m-th time, an outcome $\vartheta_m \in \{G, B\}$ is observed, with $\vartheta_m = G$ denoting a "good" outcome, and $\vartheta_m = B$ a "bad" outcome. If the treatment is effective (i.e., if $\mu = 1$), the outcome is good with probability $q_1^\xi \in (0, 1]$. If the treatment is ineffective (i.e., if $\mu = 0$), the outcome is bad with probability $q_0^\xi \in (0, 1]$, with $1 - q_0^\xi < q_1^\xi$. Hence, in this example, $G_\xi(\vartheta^m; \mu)$ is also Bernoulli with parameter q_μ^ξ that does not depend on the history ϑ^m of past signal realizations.

The physician's flow payoff from administering a ξ-treatment is v^ξ if the outcome is good and 0 otherwise. Given a history ϑ^{m-1} of past treatment's outcomes, denote by $p^\xi(\vartheta^{m-1})$ the posterior probability that the specific ξ-treatment is effective. The distribution $L_\xi(m; \mu)$ from which the physician's payoff from administering the ξ-treatment for the m-th time is drawn is thus binary with support $\{0, v^\xi\}$ and parameter $p^\xi(\vartheta^{m-1})$.

We enrich this classic problem by considering the possibility that the physician can search for new treatments. Specifically, each search costs the physician $c \geq 0$ and results in the discovery of a new ξ-treatment with probability ρ^ξ, with $\rho^\alpha + \rho^\beta = 1$ (as is typically the case in scientific discovery, the outcome of search is unknown ex-ante). Hence, in this example, for each m, $J((c_k, E_k)_{k=0}^{m-1})$ is a distribution that assigns probability ρ^α to the event that $c_m = c$ and $E_k \equiv (n_k(\alpha), n_k(\beta)) = (1, 0)$, and probability ρ^β to the event that $c_m = c$ and $E_k = (0, 1)$.

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 $\vartheta^m \equiv (\vartheta_s)_{s=1}^m$ 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.[11,12] While the category ξ is fixed, the history θ of past signal realizations changes over time as the result of the information that 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, $\omega^S = (c_k, E_k)_{k=0}^{m-1}$ 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

[11]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.

[12]As shown below, when the CS is endogenous, the outcome of each CS's expansion may depend on the categories of the alternatives added to the CS through previous expansions. If the CS were exogenous, without loss of generality, one could always take each alternative's category to coincide with its "name".

of an alternative $\omega^P \in \Omega^P$, the number of alternatives in the CS in that state. Let $\Omega \equiv \Omega^P \cup \Omega^S$.[13] 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 alternatives' states $\omega^P = (\xi, \theta)$ 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.

3.1 Optimal Policy

We now characterize the optimal policy and discuss its implications for experimentation and search dynamics. Recall that a policy χ for the decision problem above specifies, 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.[14]

For each state ω^P of an alternative, let[15]

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

denote the "index" of an alternative in the CS currently in state ω^P, where τ denotes a stopping time (that is, a rule prescribing when to stop, as a function of the observed signal realizations), and where u_s denotes the flow payoff from the alternative's s-th exploration. The definition in (1) is equivalent to the definition in Gittins and Jones (1974).[16] As is well known, the optimal stopping rule in the definition of the index is the first time at which the index falls 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, denote the maximal index among the alternatives within the CS by $\mathcal{I}^*(\mathcal{S}^P)$.[17]

We now define an index for search (i.e., expansion of the CS). This index is independent of the state of each alternative in the CS, conditional on the state of the search technology ω^S.[18] Analogously to the indexes defined above, the index for search is defined as the maximal

[13] Note that $\Omega^P \cap \Omega^S = \emptyset$.

[14] 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.

[15] The expectations in (1) are under the process obtained by selecting the given alternative in all periods.

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

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

[18] That is, the index depends on the state of each alternative in the CS only through the information that the

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. 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.[19]

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, and where the expectations are under the process generated by the rule π.

Definition 1. 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 $\mathcal{I}^S(\omega^S) < \mathcal{I}^*(\mathcal{S}^P)$.[20]

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. 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).[21]

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

Theorem 1. *(i) The index policy χ^* is optimal in the sequential experimentation problem with endogenous CS.*

(ii) The index for search, as defined in (2), admits the following recursive representation. For

latter state contains for the state ω^S of the search technology.

[19] 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.

[20] Recall that $\mathcal{I}^*(\mathcal{S}^P)$ is the largest index among the alternatives in the CS.

[21] 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 $\kappa(v)$.

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* $s \geq 1$ *at which* $\mathcal{I}^S$ *and all the indexes of the alternatives brought to the CS by search fall weakly below the value* $\mathcal{I}^S(\omega^S)$ *of the search index when search was launched, and where the expectations are under the process induced by the index policy* χ^*.

(iii) 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 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 new alternatives brought to the CS at each expansion only depend on the number of alternatives from each category brought to the CS in the past. Under such assumptions, the theorem establishes a generalization of the Gittins-index Theorem, according to which selecting in each period the alternative, or search, with the highest index is optimal.[22] Part (ii) further characterizes the stopping time in the index of search. Such recursive representation facilitates an explicit characterization of the index in applications, permits us to identify various properties of the dynamics of experimentation and search, and can be used for comparative statics, as illustrated in the next section. Finally, part (iii) 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).

3.2 Implications for Exploration and Expansion Dynamics

We now highlight several properties of the dynamics of exploration and CS expansion, under the optimal policy.

Corollary 1 (Invariance of expansion to CS composition)**.** *At any period, the decision to expand*

[22] The reason why indexability of the optimal policy is not obvious is that search is a "meta-arm" bringing alternatives whose returns are correlated at the time search is launched (through the alternatives' categories) and that one needs to process optimally. Problems in which alternatives correspond to "meta arms", i.e., to sub-problems with their own sub-decisions, typically do not admit an index solution, even if each sub-problem is independent from the others, and even if one knows the solution to each independent sub-problem. In the same vein, dependence, or correlation, between alternatives typically precludes indexability. This is so even if each subset of dependent alternatives evolves independently of all other subsets, and even if one knows how to optimally choose among the dependent alternatives in each subset in isolation. We provide an example illustrating such difficulties in the Online Supplement.

the CS is invariant to the composition of the CS, conditional on the value $\mathcal{I}^*(\mathcal{S}^P)$ of the alternative with the highest index, and the state ω^S of the search technology.

The corollary is an immediate implication of the optimal policy being an index policy. The result is not trivial, because the *opportunity cost* of expanding the CS (i.e., the value of continuing with the current CS) may well depend on the entire composition of the CS, beyond the information contained in $\mathcal{I}^*(\mathcal{S}^P)$ and ω^S.

Corollary 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.*

Corollary 2 is also an immediate implication of 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 what 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.

Corollary 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.*

The result 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 (ii) 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 search fall weakly below the value of the search index at the time search was 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.

Definition 2. (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 $J(\omega^S)$.[23]

[23] That is, the search technology is deteriorating if, regardless of the outcome of past searches, for any k and

Corollary 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, $\mathcal{V}^*(\mathcal{S}) = \mathcal{V}^*(\mathcal{S}')$.*

The corollary says that the continuation value when search is launched is invariant to the state of the CS. The result follows from the fact that, without loss of optimality, the DM never comes back to any alternative in the CS after search is launched. The same property holds in case of improving search technologies.

Corollary 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.*

Since 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. In this case, search corresponds to disposal of all alternatives in the current CS. Each time the DM searches, she starts fresh.

Corollary 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.*

The result follows again from the recursive characterization of the stopping time in the index of search, as per part (ii) 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 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.

Corollary 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*

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

The result in Corollary 7, which follows directly from part (iii) in Theorem 1, can be used to price changes to 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

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 in the corollaries below.

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 search technology from ω^S to $\hat{\omega}^S$.

4 Applications

Having characterized the optimal policy and its key implications for the dynamics of exploration and CS expansion, we now put the results to work in a few applications of interest.

4.1 Administration of Medical Treatments

Consider the environment described in Subsection 2.1. To simplify, further assume that, when a treatment is ineffective (i.e., $\mu = 0$), the outcome is bad with certainty, that is, $q_0^\xi = 1$.[24] Using the results in the previous section, we can arrive at the following characterization of the indexes for the optimal policy (see the Appendix for the derivations). For any $\omega^P = (\xi, \theta)$, the index of a treatment in state ω^P is equal to

$$\mathcal{I}^P(\omega^P) = \frac{\left(1 - \delta + \delta q_1^\xi\right) p^\xi(\theta) q_1^\xi v^\xi}{1 - \delta + \delta p^\xi(\theta) q_1^\xi}. \tag{5}$$

The index of search is invariant to ω^S and equal to[25]

$$\mathcal{I}^S = \frac{(1-\delta)\left\{\sum_{\xi\in\{\alpha,\beta\}} \rho^\xi \mathbb{E}\left[\sum_{s=0}^{\tau^{\xi*}-1} \delta^s u_s \middle| (\xi, \emptyset)\right] - c\right\}}{1 - \sum_{\xi\in\{\alpha,\beta\}} \rho^\xi \mathbb{E}\left[\delta^{\tau^{\xi*}} \middle| (\xi, \emptyset)\right]}, \tag{6}$$

where $\tau^{\xi*}$ is the first time at which the value of the index of the new ξ-treatment brought to the CS by search drops weakly below $\mathcal{I}^S$ ($\tau^{\xi*} = \infty$ if this event never occurs), and where u_s denotes the flow payoff from the s-th administration of the treatment.

We now use the above results to illustrate some of the novel comparative statics that may arise when accounting for the endogeneity of the CS. Consider an improvement in a category of treatments. Such an improvement may occur because of technological advancement, changes in pricing, or changes in the incentive schemes offered to the physician for the administration of the various treatments.

With an exogenous CS, an improvement in a category of treatments always leads to an

[24]The results below extend to environments where, occasionally, ineffective treatments also yield good outcomes with positive probability, provided that such a probability is smaller than the one for effective treatments.

[25]The expectations in the formula in (6) are under a rule selecting in each period the ξ-treatment brought to the CS by search.

increase in their administration at the expense of other categories. While this may happen also with an endogenous CS, the opposite may also occur.

Corollary 8. *Consider the environment described above. The ex-ante expected discounted number of times the α-treatments are administered may decrease both with the ex-ante probability $p^\alpha(\emptyset)$ such treatments are effective and with the value v^α the physician assigns to good outcomes delivered through such treatments.*

Importantly, the phenomenon in the corollary is not knife-edge (see the Appendix for conditions under which the phenomenon occurs). Roughly, when the CS is endogenous, an improvement in the α-treatments also increases the value of expanding the CS. This is because such expansions are expected to bring more attractive α-treatments. Because the increase in the index of the α-treatments is not homogeneous across histories, there can be histories at which the increase in the index of search is larger than the increase in the index of the α-treatments. At such histories, the physician may choose to expand the CS instead of further exploring one of the α-treatments. This may happen despite the fact that the only reason why search is more attractive is that it brings more attractive α-treatments. Once the expansion is carried out, it may change the composition of the CS in favor of the β-treatments.

Interestingly, this phenomenon is not monotone in the parameters of the model (and, as a result, does not involve extreme parameter values). To see why this is the case, consider the role of ρ^β, the probability that search brings a β-treatment. When ρ^β is low, because search delivers primarily α-treatments, the changes in the composition of the CS due to search being carried out more often contribute to a boost in the administration of the α-treatments. In this case, there is no crowding out. When, instead, ρ^β is high, search is unlikely to bring α-treatments. As a result, improvements in the α-treatments have a small impact on the search index as the latter is driven primarily by the properties of the β-treatments (this is because the search index "averages" over the properties of the two categories, as can be seen from the formula in (6)). In this case, the improvement in the search index is small compared to the improvement in the indexes of the α-treatments that delivered bad outcomes. The instances where search is carried out instead of further exploration of the α-treatments are then rare. The direct effect of the improvement in the attractiveness of the α-treatments then prevails over the indirect effect of the improvement in the attractiveness of search. Hence, in this case, too, there is no crowding out. The phenomenon in the corollary thus occurs only for intermediate values of ρ^β. Similar non-monotonicities apply to the other parameters of the model.

Phenomena analogous to those discussed in this subsection are likely to be relevant also in other scientific research environments, where improvements in a product category may spur investments in basic research which in turn may induce more experimentation with products from other categories.

4.2 Clinical Trials towards Regulatory Approval

In many problems of interest, a player produces research output to persuade another player to take a decision. For example, a pharmaceutical company may undertake clinical trials experimenting with different drugs or vaccines with the intent to induce a regulatory authority (e.g., the FDA) to approve one of its products. Our results can be used to shed light on such problems.

Consider a firm that needs regulatory approval to sell its products. The products differ in their profitability to the firm, but also in their safety. We capture such a heterogeneity by assuming that each product belongs to a category $\xi \in \Xi$, where Ξ is a finite set. Each ξ-product can either be safe ($\mu = 1$) or unsafe ($\mu = 0$), and this event is unknown both to the firm and to the regulator at the outset. The products' safety is independent across products, conditional on their category. Let $p^\xi(\emptyset)$ denote the prior probability that a category-ξ product is safe. Each ξ-product, when sold to the market, brings the firm a flow payoff equal to $(1-\delta)v^\xi > 0$. There is no value to the firm in selling more than one product per period (e.g., because the products are seen as close substitutes by the consumers).

At each period, the firm can either conduct an experiment on one of the products in its current CS, expand the CS by conducting basic research aimed at identifying new products (e.g., testing new molecule compounds), or sell one of its approved products.[26] Each experiment on a ξ-product generates a binary outcome $\vartheta \in \{G, B\}$, with $\vartheta = G$ denoting a "good" outcome and $\vartheta = B$ a "bad" one. If the ξ-product is safe, a good outcome $\vartheta = G$ is generated with probability $q_1^\xi = \Pr(\vartheta = G|\mu = 1) \in (0, 1]$. If, instead, the ξ-product is unsafe, a bad outcome $\vartheta = B$ is generated with probability $q_0^\xi = \Pr(\vartheta = B|\mu = 0)$, with $q_1^\xi \geq 1 - q_0^\xi$. Given a history $\theta = (\vartheta_s)_{s=1}^m$ of past outcomes, denote by $p^\xi(\theta)$ the posterior probability that a ξ-product is safe. The history θ is public.

Experimentation with a ξ-product entails a cost $\lambda^\xi(\theta)$ to the firm, the magnitude of which may depend on the outcomes of past experiments. Expansion of the CS entails a constant cost $c \geq 0$ and brings a new product from category $\xi \in \Xi$ with probability ρ^ξ, with $\sum_{\xi\in\Xi} \rho^\xi = 1$.[27]

The firm's goal is to maximize the expected discounted profit from selling its approved products, net of all experimentation and search costs.

As in Henry and Ottaviani (2019), the approval process is modeled as follows. For each category ξ, there exists a threshold $\Psi^\xi \in (0, 1]$ such that a ξ-product is approved once the posterior probability that the product is safe exceeds the threshold Ψ^ξ. To avoid trivialities, assume that $p^\xi(\emptyset) < \Psi^\xi$, so that each product must be tested at least once to be approved. Contrary to Henry and Ottaviani (2019), here we take the approval thresholds Ψ^ξ as given

[26] Such basic research is typically interpreted to be "undirected," that is, conducive to innovations and products that are not under the firm's direct control.

[27] That $\sum_{\xi\in\Xi} \rho^\xi = 1$ is without loss of generality. The case where search brings no product with positive probability can always be captured by letting one of the categories replicate the arrival of no new product.

and investigate how the latter influence the firm's investment in basic research and ultimately the approval of products from different categories. In future work, it would be interesting to endogenize such thresholds by studying the game between the regulatory authority and the experimenting firm.

In the context of this application, the indexes of Theorem 1 take the following form (see the Appendix for the derivations). Given $\omega^P = (\xi, \theta)$, the index of a product in state ω^P that has not been approved yet is equal to

$$\mathcal{I}^P(\omega^P) = \frac{(1-\delta)\mathbb{E}\left[-\sum_{s=0}^{min\{\tau^*,\phi\}} \delta^s \lambda^\xi(\theta) + \delta^\phi \mathbf{1}_{\{\phi<\tau^*\}} v^\xi | \omega^P\right]}{1-\mathbb{E}\left[\delta^{\tau^*}|\omega^P\right]}, \tag{7}$$

where the expectations in (7) are under the process obtained by selecting the product under consideration in each period, ϕ is the first time at which the posterior belief that the product is safe exceeds the threshold Ψ^ξ ($\phi = \infty$ if this event never occurs), and τ^* is either the first period at which the posterior belief that the product is safe is below $p^\xi(\theta)$, when such an event occurs before the product is approved (i.e., before period ϕ), or is equal to $\tau^* = \infty$ otherwise. The index for search is given by

$$\mathcal{I}^S = \frac{(1-\delta)\left(-c + \delta \sum_{\xi \in \Xi} \rho^\xi \mathbb{E}^{\chi^*}\left[-\sum_{s=0}^{min\{\tau^{\xi*},\phi^\xi\}} \delta^s \lambda^\xi(\theta) + \delta^{\phi^\xi} \mathbf{1}_{\{\phi^\xi<\tau^{\xi*}\}} v^\xi | \xi, \emptyset\right]\right)}{1-\sum_{\xi\in\Xi} \rho^\xi \mathbb{E}^{\chi^*}\left[\delta^{\tau^{\xi*}}|\xi,\emptyset\right]}, \tag{8}$$

where ϕ^ξ is the first time at which the posterior belief that the new ξ-product brought to the CS by search is safe exceeds the approval threshold, whereas $\tau^{\xi*}$ is either the first time at which the value of the index of the new ξ-product drops weakly below $\mathcal{I}^S$, when this event occurs before ϕ^ξ, or is equal to $\tau^{\xi*} = \infty$ otherwise. The index of a ξ-product that received regulatory approval is constant and equal to $(1-\delta)v^\xi$. As we show in the Appendix, as soon as one of the firm's products is approved, the firm brings to an end its experimentation process and sells the approved product in each of the subsequent periods.

Now suppose that the regulator lowers the approval standard for one of the products' category, possibly with the intent of favoring the approval of products from that category. As we show below, such regulatory changes are not guaranteed to deliver the desired results, when accounting for their effects on the firm's investment in basic research. For simplicity, suppose that, as in the previous application, there are two products' categories, α and β.

Corollary 9. *A reduction in the approval standard for the α-products from Ψ^α to $\Psi^\alpha - \varepsilon$, $\varepsilon > 0$, may reduce the ex-ante probability that an α-product is approved.*

The arguments are related to those establishing Corollary 8 (the proof is in the Online Supplement). The reduction in the approval standard for the α-products unambiguously contributes to an increase in the value the firm attaches to experimenting with such products. However, it also increases the value the firm assigns to basic research, as the new α-products that the firm expects to discover are more likely to be approved. Search may thus crowd out further experimentation with those α-products whose earlier tests yielded negative outcomes. Because search also brings β-products, the changes in the composition of the CS due to search may then favor the β-products more than the α-ones, possibly leading to a reduction in the ex-ante probability of approval of the α-products. The proof in the Online Supplement identifies conditions for which this happens. As in the previous application, such conditions tend to involve intermediate values of the relevant parameters (the reason for the non-monotonicities is similar to the one discussed above).

More generally, the result in the corollary may warn against simplistic assessments of policy interventions when such interventions also affect the incentives for search.

4.3 Pandora's Problem with Endogenous Set of Boxes

Consider the following variant of Weitzman's (1979) "Pandora's boxes problem" in which the set of boxes is endogenously expanded over time. Each alternative is a "box" and belongs to a category $\xi \in \Xi$. To each category corresponds 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., of opening) the box. As in Weitzman's (1979) original setting, each box's prize v is revealed upon the first inspection. At each period, the DM can either (a) search for additional boxes to add to the CS, (b) open one of the boxes in the CS to learn its prize, or (c) stop and either recall the prize of one of the previously opened boxes, or take the outside option, with either one of the last two choices ending the decision problem. For simplicity, assume that each search $m \in \mathbb{N}$ brings exactly one box, whose category ξ is drawn from Ξ according to a time-homogeneous distribution $\rho \in \Delta(\Xi)$ independently across searches, with ρ^{ξ} denoting the probability that search brings a ξ-box, and $\sum_{\xi \in \Xi} \rho^{\xi} = 1$.[28] The boxes' prizes are drawn independently, conditional on the boxes' categories. The cost of expanding the set of boxes depends on the number of past searches, with $c(m)$ denoting the cost of the m-th search, where $c(\cdot)$ is a positive and increasing function. The DM discounts the future according to δ.

The solution to Pandora's boxes problem with an endogenous CS is the index policy of Theorem 1. In the context of this application, the indexes take the following form (see the Appendix for the derivations). For any $\omega^P = (\xi, \emptyset)$, the index of a ξ-box that has not been

[28] All the results extend to the case where Ξ is infinite. Likewise, that the distribution ρ is invariant to the number of past searches is not essential. The results below extend to the case where such a distribution depends on m provided that the indexes for search decline (weakly) with the number of past searches.

opened yet is given by[29]

$$\mathcal{I}^P(\omega^P) = \frac{-\lambda^\xi + \delta \int_{\frac{\mathcal{I}^P(\omega^P)}{1-\delta}}^{\infty} v \mathrm{d}F^\xi(v)}{1 + \frac{\delta}{1-\delta}\left(1 - F^\xi\left(\frac{\mathcal{I}^P(\omega^P)}{1-\delta}\right)\right)}. \tag{9}$$

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 price exceeds l. One can then use our recursive characterization in Theorem 1 to verify that, in this problem, the index of search is given by[30]

$$\mathcal{I}^S(m) = \frac{-c(m) + \delta \sum_{\xi \in \Xi(\mathcal{I}^S(m))} \rho^\xi \left(-\lambda^\xi + \delta \int_{\frac{\mathcal{I}^S(m)}{1-\delta}}^{\infty} v \mathrm{d}F^\xi(u)\right)}{1 + \sum_{\xi \in \Xi(\mathcal{I}^S(m))} \rho^\xi \left[\delta + \frac{\delta^2}{1-\delta}\left(1 - F^\xi\left(\frac{\mathcal{I}^S(m)}{1-\delta}\right)\right)\right]}. \tag{10}$$

As in Weitzman (1979), the reservation prices $\mathcal{I}^P(\omega^P)$ have the following interpretation. Suppose there are only two alternatives. One is an unopened ξ-box and the other is a hypothetical box, with a known value K. The reservation price is the value of K, multiplied by $(1-\delta)$, for which the DM is indifferent between taking the hypothetical box and inspecting the ξ-box while maintaining the option to recall the hypothetical box once the prize v of the ξ-box is discovered. The reservation price $\mathcal{I}^S(m)$ of search extends this interpretation as follows. Suppose there are two options: the hypothetical box with known value K described above, and the option of expanding the CS. The reservation price of search is the value K for which the DM is indifferent between taking the hypothetical box right away, and expanding the CS, maintaining the option to take the hypothetical box either (a) once the category ξ of the newly discovered box is discovered and $\mathcal{I}^P(\xi, \emptyset) \leq K$, or (b), in case $\mathcal{I}^P(\xi, \emptyset) > K$, after the prize v of the newly discovered ξ-box is learned and $v \leq K$.

We highlight various implications of endogenizing the set of boxes in the next subsection, after showing how the boxes problem described above can be adapted to study online consumer search.

4.4 Online consumer search: endogenous click-through-rates

Consider the problem of a consumer searching online for a product to purchase. The Pandora's boxes problem with an endogenous set of boxes described in the previous subsection can be used to endogenize the relationship between the positions of the products' ads on the platform and the

[29]This index corresponds to what Weitzman refers to as a box's "reservation price". Weitzman defines the reservation price $\hat{\mathcal{I}}^P(\omega^P)$ 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
$\hat{\mathcal{I}}^P(\omega^P) = [-\lambda^\xi + \delta \int_{\hat{\mathcal{I}}^P(\omega^P)}^{\infty} v \mathrm{d}F^\xi(v)]/[1 - \delta + \delta[1 - F^\xi(\hat{\mathcal{I}}^P(\omega^P))]]$. The indexes in (9) are thus equal to the reservation prices in Weitzman (1979) multiplied by $(1-\delta)$, that is, $\mathcal{I}^P(\omega^P) = (1-\delta)\hat{\mathcal{I}}^P(\omega^P)$.

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

corresponding click-through-rates (hereafter CTRs). To see this, note that, in this environment, reading an ad corresponds to expanding the CS, clicking on an ad corresponds to opening a box, and purchasing a product from one of the visited vendors corresponds to selecting an opened box.[31]

Formally, suppose that each category $\xi \in \Xi$ corresponds to a different firm and that each position $m \in \mathbb{N}$ is occupied by the ad of one and only one firm, with the same firm possibly displaying ads for different products at multiple positions. Reading the m-th ad reveals to the consumer the identity $\xi(m) \in \Xi$ of the firm occupying the m-th position. The consumer believes that each $\xi(m)$ is drawn from $\rho \in \Delta(\Xi)$, independently across m.[32] By clicking on the m-th ad, the consumer is directed to firm $\xi(m)$'s website, where she incurs a cost $\lambda^{\xi(m)}$ to learn her value v for the firm's product, with v drawn from an absolutely continuous distribution $F^{\xi(m)}$.[33] The consumer expects the values v to be drawn independently.[34] Let $c(m)$ denote the cost of reading the m-th ad. We then have that the index for the decision to read the m-th ad is equal to $\mathcal{I}^S(m)$, with $\mathcal{I}^S(m)$ as in (10), whereas the index for the decision to click on the m-th ad, after discovering the identity $\xi(m)$ of the firm advertising at the m-th position, is equal to $\mathcal{I}_m \equiv \mathcal{I}^P(\xi(m), \emptyset)$, with $\mathcal{I}^P(\xi(m), \emptyset)$ as in (9).

One can then use the model to endogenize the probability with which the consumer reads the ads, clicks on them, and finalizes her purchases. In a similar setting, but with an exogenous CS, Choi, Dai and Kim (2018) – and, independently, Armstrong, 2017 – derive a static condition characterizing eventual purchasing decisions based on a comparison of "effective values." Proposition 1 below extends the characterization to an endogenous CS. Let v_m denote the value to the consumer for the product sold by the firm advertising at the m-th position. For all $m \geq 1$, let $w_m \equiv \min\{\mathcal{I}_m, v_m(1-\delta)\}$ be the "effective value" of the product advertised at the m-th position (for brevity, product m) when the product is already in the consumer's CS, as in Choi, Dai and Kim (2018), and $d_m \equiv \min\{w_m, \mathcal{I}^S(m)\}$ the product's "discovery value," when the product must be brought to the consumer's CS before it can be explored. Let product $m = 0$ correspond to the consumer's outside option, with $w_0 = d_0 = 0$.

Proposition 1. *The consumer purchases product m if, for all $l \in \mathbb{N} \cup \{0\}$, $l \neq m$, $d_l < d_m$ (and only if $d_l \leq d_m$, for all $l \neq m$).*

[31]This formulation assumes that consumers read the ads in the order they are displayed but, after reading the ads, click on the links of the ads they have read in the order of their choice. This seems consistent with normal practices.

[32]This assumption simplifies the exposition but is not essential. The results below extend to other search technologies that are weakly deteriorating. For example, the consumer may expect lower positions to be occupied by firms providing, on average, lower-value products.

[33]The assumption that each F^ξ is absolutely continuous is made only to avoid the need to keep track of possible indifferences in the consumer's optimal behavior which affect the formulas but not the qualitative results.

[34]This also means that, in case the consumer encounters the same firm at different positions, she expects her value for each of the firm's products to be drawn independently across products.

As in Choi, Dai and Kim (2018), purchasing decisions are determined by a static comparison of the products' values, as in canonical discrete-choice models. Contrary to Choi, Dai and Kim (2018), however, such values account for the order by which the various products are brought to the CS. In particular, Proposition 1 implies that, provided that the reading cost $c(\cdot)$ is non-decreasing, all other things equal, the further down a product is on the list, the lower the ex-ante probability the product is purchased (and hence its ex-ante demand), a property typically assumed, but not micro-founded, in existing search models.

The result in Proposition 1 follows from the fact that the optimal policy is an index rule along with the fact that the search indexes $\mathcal{I}^S(m)$ decline with m. Heuristically, if a consumer reads the m-th ad, it must be that the reservation prices $\mathcal{I}_l$ of all products $l < m$ already in her CS, as well as the discovered values $v_l(1-\delta)$ of those products $l < m$ that have been inspected already, are no greater than $\mathcal{I}^S(m)$. Furthermore, because the search technology is non-improving, $\mathcal{I}^S(m+1) \leq \mathcal{I}^S(m)$. Hence, if after reading the m'th ad, $\mathcal{I}_m \geq \mathcal{I}^S(m)$, the consumer necessarily clicks on the m'th ad, thus learning product m's value v_m. Once v_m is learned, if $(1-\delta)v_m \geq \mathcal{I}_m$, the consumer then stops the search and purchases product m. The formal proof in the Appendix shows how the above monotonicity properties imply the result in the proposition.

The result in the previous proposition can be used to endogenize the CTRs (the fraction of ads at each position that are clicked upon, among those that are brought to the consumer's CS). In the setting described above, a product is brought to the consumer's CS after its ad has been displayed to and read by the consumer. We thus have that, for each position m, the corresponding CTR is equal to[35]

$$CTR(m) \equiv \Pr\left(m\text{'s ad is clicked} | m\text{'s ad is read}\right).$$

The following proposition characterizes CTRs in terms of effective and discovery values.

Proposition 2. *The CTR for each position $m \geq 1$ is given by*[36]

$$CTR(m) = Pr\left(\mathcal{I}_m \geq max\{max_{l<m}\{w_l\}, max_{l>m}\{d_l\}\} \mid \mathcal{I}^S(m) \geq max_{l<m}\{w_l\}\right).$$

In order for product m to be read, it must be that $\mathcal{I}^S(m) \geq \max_{l<m}\{w_l\}$, for otherwise the consumer selects one of the products advertised in one of the preceding positions before reading

[35]Note that the probability in the definition of $CTR(m)$ is computed ex-ante by integrating over the different products that are advertised at the different positions. It is not the probability that a specific product advertised at a given position is clicked upon. In other words, the CTRs are position-specific and not product-specific, consistently with the definition used in practice.

[36]For simplicity, the formula in the proposition assumes that, in case of indifference, the consumer favors position m (both when it comes to reading and clicking it). This is what justifies the weak inequalities in the formula. The proof discusses how alternative ways of breaking the indifferences must be accounted for if one were to compute bounds for such probabilities.

the ad displayed in the m'th position. Once product m is read, in order for it to be clicked upon, it must be that its index $\mathcal{I}_m$ exceeds the effective value of each product brought to the consumer's CS prior to m, but also the discovery value of all products advertised further down the list, for otherwise the consumer selects one of the other products before clicking on m.[37]

The result in Proposition 2 also suggests that, while the ex-ante demands are naturally decreasing in the positions, the CTRs need not be monotone in m. To see this, note that $\Pr\left(\mathcal{I}_m \geq \max_{l<m}\{w_l\}\right)$ is decreasing in m, which contributes to CTRs declining in m. However, for product m to be read, it must be that $\max_{l<m}\{w_l\} \leq \mathcal{I}^S(m)$. Because $\mathcal{I}^S(m)$ is decreasing in m, $\Pr\left(\max_{l<m}\{w_l\} \leq \mathcal{I}^S(m)\right)$ is also decreasing in m, thus contributing to the possibility that CTRs are non-monotone in positions.[38]

One can also use the model to investigate the effects of additional ad space on firms' profits. Typically, a firm receiving additional ad space expects larger profits. This, however, is not guaranteed when consumers' CS are endogenous. To see this, consider the following situation. The consumer's initial CS contains three products, one from each firm $\xi \in \Xi = \{A, B, C\}$. By searching online, the consumer is presented with a fourth product drawn from Ξ according to $\rho \in \Delta(\Xi)$. As above, each product yields the consumer a net value v drawn from a distribution F^ξ, independently across products.[39] The cost to the consumer of learning her value for each ξ's products is λ^ξ.[40] The consumer has unit demand and each firm's profit is the same for each of its products.

The following result illustrates how the increase in the probability that search brings an additional product by firm ξ may reduce the index of search, inducing the consumer to visit the website of one of firm ξ's competitors before searching for new products. When strong enough, such an effect may reduce the probability that one of firm ξ's product is selected, and hence firm ξ's profits.

Corollary 10. *Consider the environment described above. An increase in the probability ρ^ξ that search brings an additional product from firm ξ may reduce firm ξ's ex-ante expected profits.*

See the Online Supplement for details.

[37] Note that the assumption that $\mathcal{I}^S(l)$ is weakly decreasing in l is important here. It implies that, if for some position $l > m$, $d_l > \mathcal{I}_m$, then for all $j = m+1, ..., l$, $\mathcal{I}^S(j) > \mathcal{I}_m$, meaning that the consumer will necessarily read the ad of any product displayed between position m and position l before clicking on m. If for any of such product the discovery value exceeds $\mathcal{I}_m$, the consumer purchases one of these products before clicking on m.

[38] That $\Pr\left(\mathcal{I}_m \geq \max_{l>m}\{d_l\}\right)$ is increasing in m also contributes to the possibility of CTRs increasing in m.

[39] If the extra product the consumer is presented when searching is from firm ξ, the value the consumer derives from such a product is also drawn from F^ξ, independently from the value derived from the three products already in the CS.

[40] Again, if the extra product discovered is by firm ξ, because the value the consumer derives from such a product is independent from the other ξ-product already in the CS, the total cost the consumer incurs to learn her value for the two ξ-products is $2\lambda^\xi$.

5 Extensions

The results accommodate for a few extensions that may be relevant for applications.

Multiple expansion possibilities. In certain problems of interest, the decision to search also involves an intensive margin, as when the DM chooses "how much" to invest in search. As we show in the Online Supplement, in general, such problems do not admit an index solution because of the correlation in the search outcomes. Instead, the analysis readily extends to an environment in which there are multiple search possibilities with independent outcomes, by allowing for multiple "search arms".

No discounting. All 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 $\delta \to 1$ (e.g., Dumitriu, Tetali, and Winkler, 2003).

Irreversible Choice. In many decision problems, in addition to learning about existing options and searching for new ones, the DM can irreversibly commit to one of the alternatives, bringing to an end the exploration process. In general, such problems do not admit an index solution. In the Online Supplement, we derive a sufficient condition under which the optimality of an index rule extends to such problems. We assume the DM must explore each alternative of category ξ at least $M_\xi \geq 0$ times before she can irreversibly commit to it (for example, a consumer must visit a vendor's webpage at least once to finalize a transaction with that vendor, as in the consumer search problem of Section 4.4). The condition guarantees that, once an alternative reaches a state in which the DM can irreversibly commit to it, its "retirement value" (that is, the value of irreversibly committing to it) either drops below the value of the outside option, or improves over time. This property is related to a similar condition in Glazebrook (1979), who establishes the optimality of an index policy in a class of bandit problems with stoppable processes. Our proof, however, is different and accounts for the fact that the set of alternatives evolves endogenously over time.

Relative length of expansion. In order to allow for frictions in the search for new alternatives, we assume that, whenever the DM searches, she cannot explore any of the alternatives in the CS, with search occupying the same amount of time as the exploration of any of the alternatives in the CS. All the results extend to a setting in which both the time that each search occupies and the time that each exploration takes vary stochastically with the state.[41] Furthermore, because the time that each exploration takes can be arbitrary, by rescaling the payoffs and adjusting the discount factor appropriately, one can make the length of time during which the exploration of the existing alternatives is paused because of search arbitrarily small. The results

[41] More generally, all of the results can be extended to a semi-Markov environment, where time is not slotted.

therefore also apply to problems in which search and learning occur "almost" in parallel.

6 Concluding remarks

We introduce a model of experimentation in which the decision maker alternates between exploring alternatives already in the consideration set and searching for new ones to explore in the future. The consideration set is thus constructed gradually over time in response to the information the decision maker collects. We characterize the optimal policy and study how the tradeoff between the exploration of existing alternatives and the expansion of the consideration set depends on the search technology. The evolution of this tradeoff is driven by a a comparison of independent indexes, where the index for search is computed in recursive form, accounting for future optimal decisions.

The analysis may also be of interest to certain dynamic problems in which the decision maker is unable to consider all feasible alternatives from the outset, either because of limited attention, or because of the sequential provision of information by interested third parties such as online platforms and search engines.

References

Armstrong, M. (2017). "Ordered consumer search," *Journal of the European Economic Association, 15*(5), 989-1024.

Armstrong, M. and J. Vickers (2015). "Which demand systems can be generated by discrete choice?" *Journal of Economic Theory*, 158, 293-307.

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

Bergemann, Dirk 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.

Berry, D. A. (2006). "Bayesian clinical trials," *Nature Reviews Drug discovery, 5*(1), 27-36.

Berry, D. A., and B. Fristedt (1985). "Bandit problems: sequential allocation of experiments" (Monographs on statistics and applied probability). *London: Chapman and Hall,* 5(71-87), 7-7.

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., Dai, A. Y. and K. Kim (2018). "Consumer search and price competition." *Econometrica*, 86(4), 1257-1281.

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

Dickstein, M. J. (2014). "Efficient provision of experience goods: Evidence from antidepressant choice." mimeo Stanford University.

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 (2020). "Soft Affirmative Action and Minority Recruitment," mimeo, Northwestern and Tel-Aviv University.

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. C. (1979). "Bandit processes and dynamic allocation indexes," *Journal of the Royal Statistical Society. Series B (Methodological)* 41 (2), 148–177.

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 (2019). "Attention, please!," mimeo.

Greminger, R. (2021). "Optimal search and discovery," mimeo, Tilburg University.

Hauser, J. R. and B. Wernerfelt (1990). "An evaluation cost model of consideration sets," *Journal of Consumer Research*, 16(4), 393-408.

Henry, E., and M. Ottaviani (2019). "Research and the approval process: the organization of persuasion," *American Economic Review, 109*(3), 911-55.

Honka, E., Hortacsu, A. and M. Wildebeest (2019). "Empirical search and consideration sets," *Handbook of the Economics of Marketing*, Jean-Pierre Dube and Peter Rossi (ed.), Elsevier, 193-257.

Jeziorski, P. and I. Segal (2015). "What makes them click: Empirical analysis of consumer demand for search advertising," *American Economic Journal: Microeconomics*, 7(3), 24-53.

Katehakis, M., and C. Derman (1986). "Computing optimal sequential allocation rules in clinical trials." *Lecture Notes-Monograph Series*, 8, 29-39.

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.

Liang, A., Mu, X., and V. Syrgkanis (2019). "Dynamically aggregating diverse information," mimeo.

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.

Roberts, J. H. and J. M. Lattin (1991). "Development and testing of a model of consideration set composition," *Journal of Marketing Research*, 28(4), 429-440.

Schneider, J. and C. Wolf (2019). "Switching forth and back: Problem solving under time pressure," mimeo.

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

Villar, S. S., Bowden, J., and J. Wason (2015). "Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges." *Statistical Science: A Review Journal of the Institute of Mathematical Statistics, 30*(2), 199.

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.

7 Appendix

Preliminaries. In the analysis below, it will be 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, it will be convenient to 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.[42] Note that 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 $H_{\omega^S} \in \Delta(\Omega^S)$. Note that 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.

Proof of Theorem 1. The proof is in three steps. Step 1 first establishes the result in part (ii) and then uses the recursive representation of the index of search 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 (iii). Steps 2 and 3 then use the representation in (4) to show that the DM's payoff under the proposed index rule satisfies the Bellman equation for the dynamic program under consideration, thus proving the optimality of the index policy in part (i).

Step 1. 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 following 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 following 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 definition of $\hat{\tau}$ in $\mathcal{I}^S(\omega^S)$.[43] 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 following 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 following the search under consideration are weakly below $\mathcal{I}^S(\omega^S)$, if $\hat{\tau} > \tau^*$, the

[42] Clearly, 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.

[43] 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).

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 representation in part (ii) then follows from the arguments above.

Next, consider part (iii). 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*. Starting with the initial 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 $v^1 \equiv \max\{\mathcal{I}^*(\mathcal{S}_{t(v^0)}^P), \mathcal{I}^S(\omega_{t(v^0)}^S)\}$ be the value of the largest index at $t(v^0)$, where $\mathcal{S}_{t(v^0)} = (\mathcal{S}_{t(v^0)}^P, \omega_{t(v^0)}^S)$ 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 $\omega_{t(v^0)}^S = \omega_0^S$. 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 (1) and (2), observe that

$$v^i = \frac{(1-\delta)\mathbb{E}^{\chi^*}\left[\eta^i | \mathcal{S}_{\kappa(v^i)}\right]}{\mathbb{E}^{\chi^*}\left[1 - \delta^{\kappa(v^{i+1})-\kappa(v^i)} | \mathcal{S}_{\kappa(v^i)}\right]}. \tag{11}$$

To see why (11) holds, recall that, at period $\kappa(v^i)$, given the state of the decision problem $\mathcal{S}_{\kappa(v^i)}$, the optimal stopping time in the definition of the index v^i is the first time at which the index of the alternative corresponding to v^i (if v^i corresponds to a physical alternative), or the index of search and of all alternatives introduced through future searches (in case v^i corresponds to the search index), drop below v^i.[44]

[44]Note that if, at period $\kappa(v^i)$, there are multiple options ("physical" alternatives and search) with index v^i, the average sum of the discounted net payoffs until the indexes of all options drop below v^i per unit of average discounted time is the same as the average sum of the discounted net payoffs of each individual option with index v^i normalized by the average discounted time until the index of that alternative falls below v^i. This follows from

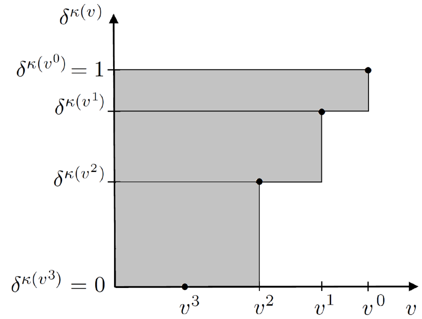

Figure 1: An illustration of the function $\delta^{\kappa(v)}$ and the region $\sum_{i=0}^{\infty} v^i \left(\delta^{\kappa(v^i)} - \delta^{\kappa(v^{i+1})}\right) = \int_0^\infty v \mathrm{d}\delta^{\kappa(v)}$, for a particular path with $\kappa(v^3) = \infty$.

Rearranging, multiplying both sides of (11) by $\delta^{\kappa(v^i)}$, 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 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

$$\mathcal{V}(\mathcal{S}_0) = \mathbb{E}^{\chi^*}\left[\sum_{i=0}^{\infty} v^i\left(\delta^{\kappa(v^i)} - \delta^{\kappa(v^{i+1})}\right)|\mathcal{S}_0\right]. \tag{12}$$

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 (12) is equivalent to

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

The construction of the integral function (13) is illustrated in Figure 1.

Step 2. We use the representation of the DM's payoff under the index rule in (4) 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. This will permit us to establish in Step 3 the optimality of χ^* through dynamic programming.

Abusing notation, in this step, we find it useful to denote the state of the decision problem by

the independence of the processes. Hence, Condition (11) holds irrespectively of whether, at $\kappa(v^i)$, there is a single or multiple options with index v^i.

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.[45] 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).[46]

Lemma 1. *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 1. 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' 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. $\square$

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

$$V^P(\omega^P|\mathcal{S}_0) \equiv (1-\delta)\mathbb{E}\left[u|\omega^P\right] + \delta\mathbb{E}^{\chi^*}\left[\mathcal{V}\left(\mathcal{S}_0 \backslash e(\omega^P) \vee e(\tilde{\omega}^P)\right)|\omega^P\right] \tag{14}$$

[45] Clearly, 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 $\omega^S \in \Omega^S$, $\mathcal{S}(\omega^S) = 0$.

[46] 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.

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

$$V^S(\omega^S|\mathcal{S}_0) \equiv -(1-\delta)\mathbb{E}\left[c|\omega^S\right] + \delta\mathbb{E}^{\chi^*}\left[\mathcal{V}\left(S_0 \backslash e(\omega^S) \vee e(\tilde{\omega}^S) \vee W^P(\tilde{\omega}^S)\right)|\omega^S\right] \tag{15}$$

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}.[47]

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^A_M, and enlarge Ω^P to include ω^A_M. Similarly, let $e(\omega^A_M)$ 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 $\kappa(v|\mathcal{S}_0 \vee e(\omega^A_M)) = \kappa(v|\mathcal{S}_0)$, whereas if $v < M$, then $\kappa(v|\mathcal{S}_0 \vee e(\omega^A_M)) = \infty$. Hence, the representation in (4), adapted to the fictitious environment that includes the auxiliary option, implies that

$$\begin{aligned}\mathcal{V}(\mathcal{S}_0 \vee e(\omega^A_M)) &= \int_0^\infty \left(1 - \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|\mathcal{S}_0 \vee e(\omega^A_M)\right]\right)\mathrm{d}v = M + \int_M^\infty \left(1 - \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|\mathcal{S}_0\right]\right)\mathrm{d}v \\ &= \mathcal{V}(\mathcal{S}_0) + \int_0^M \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|\mathcal{S}_0\right]\mathrm{d}v. \end{aligned} \tag{16}$$

The definition of χ^*, along with Conditions (14) and (15), then imply the following:

Lemma 2. *For any* (ω^S, ω^P, M),

$$\mathcal{V}(e(\omega^S) \vee e(\omega^A_M)) = \begin{cases} V^S(\omega^S|e(\omega^S) \vee e(\omega^A_M)) & \text{if } M \leq \mathcal{I}^S(\omega^S) \\ M > V^S(\omega^S|e(\omega^S) \vee e(\omega^A_M)) & \text{if } M > \mathcal{I}^S(\omega^S) \end{cases} \tag{17}$$

$$\mathcal{V}(e(\omega^P) \vee e(\omega^A_M)) = \begin{cases} V^P(\omega^P|e(\omega^P) \vee e(\omega^A_M)) & \text{if } M \leq \mathcal{I}^P(\omega^P) \\ M > V^P(\omega^P|e(\omega^P) \vee e(\omega^A_M)) & \text{if } M > \mathcal{I}^P(\omega^P). \end{cases} \tag{18}$$

Proof of Lemma 2. First note that the index corresponding to the auxiliary option is equal to M. Hence, if $M \leq \mathcal{I}^S(\omega^S)$, given $e(\omega^S) \vee e(\omega^A_M)$, χ^* prescribes to start with search, implying that $\mathcal{V}(e(\omega^S) \vee e(\omega^A_M)) = V^S(\omega^S|e(\omega^S) \vee e(\omega^A_M))$. 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, $M > V^S(\omega^S|e(\omega^S) \vee e(\omega^A_M))$, observe that the payoff $V^S(\omega^S|e(\omega^S) \vee e(\omega^A_M))$ from starting

[47] 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.

with search and then following χ^* in each subsequent period is equal to $V^S(\omega^S|e(\omega^S)\vee e(\omega^A_M)) = \mathbb{E}^{\chi^*}_{>1}\left[(1-\delta)\sum_{s=0}^{\bar{\tau}-1}\delta^s U_s + \delta^{\bar{\tau}}M|\omega^S\right]$, 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 $e(\omega^S)\vee e(\omega^A_M)$ by searching in the first period and then following the index policy in each subsequent period (the notation $\mathbb{E}^{\chi^*}_{>1}[\cdot]$ 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 $\mathcal{I}^S(\omega^S)$,

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

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

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

Similar arguments establish Condition (18). □

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.[48]

Lemma 3. *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*[49]

$$D^S(\omega^S|\mathcal{S}_0) = \int_0^{\mathcal{I}^*(\mathcal{S}_0^P)} D^S(\omega^S|e(\omega^S)\vee e(\omega^A_v))d\mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|\mathcal{S}_0\setminus e(\omega^S)\right] + \mathbb{E}^{\chi^*}\left[\delta^{\kappa(0)}|S_0\backslash e(\omega^S))\right]D^S(\omega^S|e(\omega^S)\vee e(\omega^A_0)). \tag{19}$$

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

$$D^P(\omega^P|\mathcal{S}_0) = \int_0^{max\{\mathcal{I}^*(\mathcal{S}_0^P\backslash e(\omega^P)),\mathcal{I}^S(\omega^S)\}} D^P(\omega^P|e(\omega^P)\vee e(\omega^A_v))d\mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|\mathcal{S}_0\setminus e(\omega^P)\right] + \mathbb{E}^{\chi^*}\left[\delta^{\kappa(0)}|S_0\backslash e(\omega^P))\right]D^P(\omega^P|e(\omega^P)\vee e(\omega^A_0)). \tag{20}$$

[48] 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.

[49] Recall that $\mathcal{I}^*(\mathcal{S}_0^P)$ is the largest index of the alternatives in the CS under the state $\mathcal{S}_0$.

Proof of Lemma 3. Using Condition (16), we have that, given the state $\mathcal{S}_0 \vee e(\omega_M^A)$ of the decision problem, and $\omega^S \in \Omega^S$,

$$D^S(\omega^S|\mathcal{S}_0 \vee e(\omega_M^A)) = \mathcal{V}(\mathcal{S}_0) + \int_0^M \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|\mathcal{S}_0\right] \mathrm{d}v + (1-\delta)\mathbb{E}\left[c|\omega^S\right] \tag{21}$$
$$- \delta\mathbb{E}^{\chi^*}\left[\mathcal{V}(S_0 \backslash e(\omega^S) \vee e(\tilde{\omega}^S) \vee W^P(\tilde{\omega}^S)) + \int_0^M \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|S_0 \backslash e(\omega^S) \vee e(\tilde{\omega}^S) \vee W^P(\tilde{\omega}^S))\right] \mathrm{d}v|\omega^S\right],$$

where the equality follows from combining (15) with (16). Similarly,

$$\begin{aligned} D^S(\omega^S|e(\omega^S) \vee e(\omega_M^A)) &= \mathcal{V}(e(\omega^S)) + \textstyle\int_0^M \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|e(\omega^S)\right] \mathrm{d}v + (1-\delta)\mathbb{E}\left[c|\omega^S\right] \\ &-\delta\mathbb{E}^{\chi^*}\left[\mathcal{V}(e(\tilde{\omega}^S) \vee W^P(\tilde{\omega}^S)) + \textstyle\int_0^M \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|e(\tilde{\omega}^S) \vee W^P(\tilde{\omega}^S))\right] \mathrm{d}v|\omega^S\right]. \end{aligned} \tag{22}$$

Differentiating (21) and (22) with respect to M, using the independence across alternatives and search and Lemma 1, we have that

$$\frac{\partial}{\partial M} D^S(\omega^S|\mathcal{S}_0 \vee e(\omega_M^A)) = \mathbb{E}^{\chi^*}\left[\delta^{\kappa(M)}|S_0 \backslash e(\omega^S)\right] \frac{\partial}{\partial M} D^S(\omega^S|e(\omega^S) \vee e(\omega_M^A)). \tag{23}$$

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, $D^S(\omega^S|e(\omega^S) \vee e(\omega_M^A))$, 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 $\omega^P \in \{\hat{\omega}^P \in \Omega^P : \mathcal{S}_0(\hat{\omega}^P) > 0\}$,

$$\frac{\partial}{\partial M} D^P(\omega^P|\mathcal{S}_0 \vee e(\omega_M^A)) = \mathbb{E}^{\chi^*}\left[\delta^{\kappa(M)}|S_0 \backslash e(\omega^P)\right] \frac{\partial}{\partial M} D^P(\omega^P|e(\omega^P) \vee e(\omega_M^A)). \tag{24}$$

Let $M^* \equiv \max\{\mathcal{I}^*(\mathcal{S}_0^P), \mathcal{I}^S(\omega^S)\}$. Integrating (23) over the interval $(0, M^*)$ of possible values for the auxiliary option and rearranging, we have that

$$\begin{aligned} D^S(\omega^S|\mathcal{S}_0 \vee e(\omega_0^A)) &= D^S(\omega^S|\mathcal{S}_0 \vee e(\omega_{M^*}^A)) - \int_0^{M^*} \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|S_0 \backslash e(\omega^S))\right] \frac{\partial}{\partial v} D^S(\omega^S|e(\omega^S) \vee e(\omega_v^A))\mathrm{d}v \\ &= D^S(\omega^S|\mathcal{S}_0 \vee e(\omega_{M^*}^A)) - D^S(\omega^S|e(\omega^S) \vee e(\omega_{M^*}^A)) \\ &\quad + \mathbb{E}^{\chi^*}\left[\delta^{\kappa(0)}|S_0 \backslash e(\omega^S))\right] D^S(\omega^S|e(\omega^S) \vee e(\omega_0^A)) \\ &\quad + \int_0^{M^*} D^S(\omega^S|e(\omega^S) \vee e(\omega_v^A))\mathrm{d}\mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|S_0 \backslash e(\omega^S))\right], \end{aligned}$$

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 $D^S(\omega^S|\mathcal{S}_0 \vee e(\omega_0^A)) = D^S(\omega^S|\mathcal{S}_0)$. It is also easily verified that $D^S(\omega^S|\mathcal{S}_0 \vee e(\omega_{M^*}^A)) = D^S(\omega^S|e(\omega^S) \vee e(\omega_{M^*}^A))$.[50] There-

[50]This follows immediately from the observation that $\mathcal{V}(\mathcal{S}_0 \vee e(\omega_{M^*}^A)) = \mathcal{V}(e(\omega^S) \vee e(\omega_{M^*}^A)) = M^*$, and similarly

fore, we have that

$$D^S(\omega^S|\mathcal{S}_0) = \int_0^{M^*} D^S(\omega^S|e(\omega^S) \vee e(\omega_v^A))\mathrm{d}\mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|S_0\backslash e(\omega^S)\right] \\ + \mathbb{E}^{\chi^*}\left[\delta^{\kappa(0)}|S_0\backslash e(\omega^S))\right] D^S(\omega^S|e(\omega^S) \vee e(\omega_0^A)). \tag{25}$$

Similar arguments imply that

$$D^P(\omega^P|\mathcal{S}_0) = \int_0^{M^*} D^P(\omega^P|e(\omega^P) \vee e(\omega_v^A))\mathrm{d}\mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)}|S_0\backslash e(\omega^P)\right] \\ + \mathbb{E}^{\chi^*}\left[\delta^{\kappa(0)}|S_0\backslash e(\omega^P))\right] D^P(\omega^P|e(\omega^P) \vee e(\omega_0^A)). \tag{26}$$

To complete the proof of Lemma 3, 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 (19) in the lemma follows directly from (25). Thus consider Condition (20). 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 (26) 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 (20) 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 (20) follows directly from (26).

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 (20) in the lemma follows directly from (26). That Condition (19) 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 (25) is equal to zero over the entire interval $\left[0, \max\{\mathcal{I}^*(\mathcal{S}_0^P\backslash e(\omega^P)), \mathcal{I}^S(\omega^S)\}\right]$. $\square$

Step 3. Using the characterization of the payoff differentials in Lemma 3, 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 4. *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 $\mathcal{I}^P(\omega^P) = \mathcal{I}^*(\mathcal{S}_0^P) \geq \mathcal{I}^S(\omega^S)$.*

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

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

Proof of Lemma 4. *Part 1*. First, use (17) 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 (19), $D^S(\omega^S|\mathcal{S}_0) \geq 0$ – 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 (18) 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 (20), $D^P(\omega^P|\mathcal{S}_0) \geq 0$ 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 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 $\mathcal{V}(\mathcal{S}_0) = \mathcal{V}^*(\mathcal{S}_0)$, and hence the optimality of χ^*. □

This completes the proof of the theorem. ■

Derivations of the indexes in (5) and (6). Recall that the optimal stopping time in the index definition in (1) is the first time at which the index drops below its initial value (i.e., its value at the time the index is calculated). This event occurs at the first time at which the posterior belief that the treatment is effective drops below its value $p^\xi(\theta)$ at the time the index is computed. The formula in (5) then uses the fact that a good outcome perfectly reveals that the treatment is effective, in which case $\tau^* = \infty$ in (1), whereas a single bad outcome suffices to reduce the posterior belief that the treatment is effective below the value at the time the index was computed, implying that $\tau^* = 1$.

Next, consider the index of search. Using the recursive representation in part (ii) of Theorem 1, together with Corollary 6, we have that the index for search is invariant to ω^S and equal to (6). To see this, recall that the search technology is stationary (and, hence, weakly deteriorating) in this problem. Corollary 6 then implies that the index is the same as in a fictitious environment with a single opportunity to search. Part (ii) of Theorem 1 in turn implies that the optimal stopping time in (2) is the first time at which the posterior belief about the newly added treatment's efficacy is such that the treatment's index drops below the index of search when the latter was launched. ■

Proof of Corollary 8. Hereafter, we first identify conditions under which the exploration and expansions dynamics take a particularly simple form. We then compute the ex-ante expected discounted number of times an α-treatment is administered prior to the improvement and after. Finally, we show that the administration of the α-treatments may be higher before the improvement. The conditions we identify are sufficient but not necessary for the result. While a complete characterization of the conditions for which the result holds is not easily attainable, the conditions below make clear that the result is not knife-edge.

Suppose that initially there are two treatments in the physician's CS, one of each category, and suppose that the α-treatments improve. Such an improvement may take the form of (1) an increase in the ex-ante probability that each α-treatment is effective from $p^{\alpha}(\emptyset)$ to $\hat{p}^{\alpha}(\emptyset) = p^{\alpha}(\emptyset) + \varepsilon_p$, $\varepsilon_p > 0$, or (2) an increase in the payoff the physician derives from a good outcome delivered through the administration of an α-treatment from v^{α} to $\hat{v}^{\alpha} = v^{\alpha} + \varepsilon_v$, with $\varepsilon_v \geq 0$. Let $\Lambda^{\xi}(\theta) \equiv p^{\xi}(\theta) q_1^{\xi}$ denote the posterior probability that a treatment yields a good outcome. Because the posterior belief $p^{\xi}(\theta)$ that a ξ-treatment is effective is equal to one if $\theta \neq \emptyset$ contains at least one good outcome and else depends on $\theta \neq \emptyset$ only through the number s of bad outcomes recorded in θ, with an abuse of notation, hereafter we simplify the formulas for $p^{\xi}(\theta)$, $\Lambda^{\xi}(\theta)$, and $\mathcal{I}^P(\xi, \theta)$, by replacing any vector $\theta = (B, B, ..., B)$ containing only bad outcomes with the number s of bad outcomes in the vector. Clearly, if θ contains one or more good outcomes, then $p^{\xi}(\theta) = 1$, in which case $\mathcal{I}^P(\omega^P) = q_1^{\xi} v^{\xi}$. We continue to denote by $p^{\xi}(\emptyset)$ and $\mathcal{I}^P(\xi, \emptyset)$ the prior belief a ξ-treatment is effective and the index of a ξ-treatment that has never been tested, respectively. Likewise, we let $\Lambda^{\xi}(\emptyset) \equiv p^{\xi}(\emptyset) q_1^{\xi}$.

Suppose that the following order applies

$$
\begin{aligned}
&\mathcal{I}^P(\alpha, \emptyset) > \mathcal{I}^P(\beta, \emptyset) > \mathcal{I}^P(\alpha, 1) \\
&> \frac{-c(1-\delta) + \delta \left\{ \rho^{\alpha} \Lambda^{\alpha}(\emptyset) \left[1 - \delta^2 (1 - q_1^{\alpha})^2\right] v^{\alpha} + \rho^{\beta} \Lambda^{\beta}(\emptyset) \left[1 - \delta(1 - q_1^{\beta})\right] v^{\beta} \right\}}{1 - \delta^2 \left\{ \rho^{\alpha} \delta \left[1 - 2\Lambda^{\alpha}(\emptyset) + q_1^{\alpha} \Lambda^{\alpha}(\emptyset)\right] + \rho^{\beta} \left[1 - \Lambda^{\beta}(\emptyset)\right] \right\}} \\
&> \max\{\mathcal{I}^P(\beta, 1), \mathcal{I}^P(\alpha, 2)\}.
\end{aligned}
\tag{27}
$$

We then argue that the formula for the index of search in (6) simplifies to

$$
\mathcal{I}^S = \frac{-c(1-\delta) + \delta \left\{ \rho^{\alpha} \Lambda^{\alpha}(\emptyset) \left[1 - \delta^2 (1 - q_1^{\alpha})^2\right] v^{\alpha} + \rho^{\beta} \Lambda^{\beta}(\emptyset) \left[1 - \delta(1 - q_1^{\beta})\right] v^{\beta} \right\}}{1 - \delta^2 \left\{ \rho^{\alpha} \delta \left[1 - 2\Lambda^{\alpha}(\emptyset) + q_1^{\alpha} \Lambda^{\alpha}(\emptyset)\right] + \rho^{\beta} \left[1 - \Lambda^{\beta}(\emptyset)\right] \right\}}.
$$

Once again, the result follows from part (ii) of Theorem 1, along with Corollary 5. In particular, the order in (27) implies that the optimal stopping time in (6) is equal to: (a) $\tau^{\xi *} = \infty$ if either the new treatment is an α-treatment and a good outcome is observed in one of the treatment's first two administrations, or the new treatment is a β-treatment and a good outcome is observed after the treatment's first administration; (b) $\tau^{\xi *} = 2$ if the new treatment is a β-treatment and the outcome of its first administration is bad; (c) $\tau^{\xi *} = 3$ if the new treatment is an α-treatment and each of its first two administrations yielded a bad outcome.[51]

Now suppose that the α-treatments improve. Let $\hat{\Lambda}^{\alpha}(\emptyset) \equiv \hat{p}^{\alpha}(\emptyset) q_1^{\alpha}$ and suppose that

$$\hat{\mathcal{I}}^P(\alpha, \emptyset) > \mathcal{I}^P(\beta, \emptyset) > \frac{-c(1-\delta) + \delta \left\{ \rho^{\alpha} \hat{\Lambda}^{\alpha}(\emptyset) \left[1 - \delta(1 - q_1^{\alpha})\right] \hat{v}^{\alpha} + \rho^{\beta} \Lambda^{\beta}(\emptyset) \left[1 - \delta(1 - q_1^{\beta})\right] v^{\beta} \right\}}{1 - \delta^2 \left[1 - \rho^{\alpha} \hat{\Lambda}^{\alpha}(\emptyset) - \rho^{\beta} \Lambda^{\beta}(\emptyset)\right]}$$
$$> \max\{\hat{\mathcal{I}}^P(\alpha, 1), \mathcal{I}^P(\beta, 1)\}, \tag{28}$$

where the hat on the indexes $\hat{\mathcal{I}}$ indicates that the indexes are computed after the improvement in the α-treatments. We then argue that the index for search after the improvement is equal to

$$\hat{\mathcal{I}}^S = \frac{-c(1-\delta) + \delta \left\{ \rho^{\alpha} \hat{\Lambda}^{\alpha}(\emptyset) \left[1 - \delta(1 - q_1^{\alpha})\right] \hat{v}^{\alpha} + \rho^{\beta} \Lambda^{\beta}(\emptyset) \left[1 - \delta(1 - q_1^{\beta})\right] v^{\beta} \right\}}{1 - \delta^2 \left[1 - \rho^{\alpha} \hat{\Lambda}^{\alpha}(\emptyset) - \rho^{\beta} \Lambda^{\beta}(\emptyset)\right]}.$$

The result follows again from part (ii) of Theorem 1 along with Corollary 5. Given (28), the optimal stopping time in the definition of the index of search is $\tau^{\xi *} = 2$ if the new treatment brought to the CS by search yields a bad outcome after its first administration, and $\tau^{\xi *} = \infty$ otherwise.

Next, compare the ex-ante expected discounted number of times the physician administers an α-treatment before the improvement and after. The ordering in (27) implies that, before the improvement, the physician starts by administering the α-treatment in the CS. If such a treatment yields a bad outcome, she then administers the β-treatment in the CS. If the latter also yields a bad outcome, the physician administers again the α-treatment in the CS that yielded the initial bad outcome. If this latter treatment yields a second bad outcome, the physician then searches for new treatments. If, at any point, the administered treatment yields a good outcome, because the treatment is revealed effective, the physician then administers it in all subsequent periods, thus bringing the experimentation de facto to a halt.

Because the search technology is stationary, by virtue of Corollary 5, all treatments in the CS are effectively discarded once each search for new treatments is carried out. Therefore, the expected discounted number of times the physician administers an α treatment after each search is carried out is given by

$$A_S = \rho^{\alpha} \left[1 + \delta + \frac{\Lambda^{\alpha}(\emptyset)(2 - q_1^{\alpha})\delta^2}{1-\delta} + (1 - \Lambda^{\alpha}(\emptyset)(2 - q_1^{\alpha}))\, \delta^3 A_S \right] + \rho^{\beta} \left(1 - \Lambda^{\beta}(\emptyset)\right) \delta^2 A_S.$$

Solving for A_S, we have that

$$A_S = \frac{\frac{\rho^{\alpha}}{1-\delta} \left(1 - \delta^2 + \delta^2 \Lambda^{\alpha}(\emptyset)\left(2 - q_1^{\alpha}\right)\right)}{1 - \delta^2 \left(\rho^{\alpha} \left(1 - \Lambda^{\alpha}(\emptyset)\left(2 - q_1^{\alpha}\right)\right) \delta + \rho^{\beta} \left(1 - \Lambda^{\beta}(\emptyset)\right)\right)}.$$

[51] Recall that search itself occupies one period.

From an ex-ante standpoint, the overall expected discounted number of times an α treatment is administered is therefore equal to

$$A = 1 + \frac{\delta\Lambda^{\alpha}(\emptyset)}{1-\delta} + (1-\Lambda^{\alpha}(\emptyset))(1-\Lambda^{\beta}(\emptyset))\delta^2\left[1 + \frac{\delta\Lambda^{\alpha}(1)}{1-\delta} + (1-\Lambda^{\alpha}(1))\delta^2 A_S\right],$$

where $\Lambda^{\xi}(1) = p^{\xi}(1)q_1^{\xi} = (1-q_1^{\xi})\Lambda^{\xi}(\emptyset)/\left(1-\Lambda^{\xi}(\emptyset)\right)$ is the probability of a good outcome from a ξ-treatment that yielded a bad outcome at its first administration.

Now let $\hat{A}_S$ and $\hat{A}$ be the analogs of A_S and A, respectively, after the improvement in the α-treatments. Under the order in (28), the physician first administers the α-treatment in the CS. If the latter yields a bad outcome, the physician then administers the β-treatment in the CS. If the latter also yields a bad outcome, the physician then searches for new treatments.[52] Then

$$\hat{A}_S = \rho^{\alpha}\left(1 + \frac{\delta\hat{\Lambda}^{\alpha}(\emptyset)}{1-\delta} + \delta^2\hat{A}_S(1-\hat{\Lambda}^{\alpha}(\emptyset))\right) + \rho^{\beta}\left(1-\Lambda^{\beta}(\emptyset)\right)\delta^2\hat{A}_S.$$

Solving for $\hat{A}_S$, we have that

$$\hat{A}_S = \frac{\rho^{\alpha}\left(1+\frac{\delta\hat{\Lambda}^{\alpha}(\emptyset)}{1-\delta}\right)}{1-\delta^2+\delta^2\left(\rho^{\alpha}\hat{\Lambda}^{\alpha}(\emptyset)+\rho^{\beta}\Lambda^{\beta}(\emptyset)\right)}.$$

Therefore, the ex-ante expected discounted number of times an α-treatment is administered when the α-treatments are improved is equal to

$$\hat{A} = 1 + \frac{\delta\hat{\Lambda}^{\alpha}(\emptyset)}{1-\delta} + (1-\hat{\Lambda}^{\alpha}(\emptyset))(1-\Lambda^{\beta}(\emptyset))\delta^3\hat{A}_S.$$

Finally, it can be verified that Conditions (27) and (28) are consistent with

$$A > \hat{A} \tag{29}$$

over an open set of parameter values such that $\varepsilon_p \geq 0$ and $\varepsilon_v \geq 0$, with at least one inequality strict. ■

Characterization of optimal policy in clinical trials application. The characterization of the indexes follows from arguments similar to those establishing the formulas of the indexes in (5) and (6), and is therefore omitted. Next, note that because experimenting with a product that has been approved already is dominated by selling the approved product, the index of a ξ-product that received regulatory approval is constant and equal to $(1-\delta)v^{\xi}$. The optimal policy being an index policy (which follows from Theorem 1) then implies that, as soon as one of

[52] Again, if at any point a treatment yields a good outcome, because it is revealed effective, it is then administered in each subsequent period. Furthermore, because the search technology is stationary, all treatments in the current CS are effectively discarded when search is carried out.

the firm's products is approved, the firm brings to an end its experimentation process and sells the approved product in each of the subsequent periods.[53] ■

Characterization of optimal policy in Pandora's problem with an endogenous set of boxes. Consider a relaxed problem in which the DM gets a flow payoff equal to $(1-\delta)v$ each time she selects an opened box with value v, and can revert her decision at any period. The solution to such a problem is the index policy of Theorem 1 and has the property that, once an opened box is selected, it continues to be selected in all subsequent periods. The index policy for such a problem is thus feasible (and hence optimal) also in the primitive problem.

To see that the index of a ξ-box that has not been opened yet is given by (9), note that the index of an opened box is equal to $(1-\delta)v$. Because the optimal stopping time τ^* in the definition of the index $\mathcal{I}^P(\omega^P)$ in (1) is the first time at which the value of the index drops below its value $\mathcal{I}^P(\omega^P)$ at the time the index is computed, we then have that $\tau^* = 1$ if $(1-\delta)v \leq \mathcal{I}^P(\omega^P)$ and $\tau^* = \infty$ otherwise.

Turning to the index for search, by Corollary 6, because the search technology is deteriorating, the optimal stopping-time τ^* in (2) is equal to (a) $\tau^* = \infty$ if the box identified at the m-th search has a reservation price $\mathcal{I}^P(\omega^P) > \mathcal{I}^S(m)$ and its realized flow payoff satisfies $v(1-\delta) > \mathcal{I}^S(m)$, (b) $\tau^* = 1$ if $\mathcal{I}^P(\omega^P) \leq \mathcal{I}^S(m)$, and (c) $\tau^* = 2$ if $\mathcal{I}^P(\omega^P) > \mathcal{I}^S(m)$ and $v(1-\delta) \leq \mathcal{I}^S(m)$. ■

Proof of Proposition 1. Since product 0 corresponds to the outside option, one of the products is always purchased. Let $l \neq m$ be such that $d_l < d_m$. We show that product l will not be purchased.

Case 1: $l > m$ (i.e., l is read after m is read). First, suppose that $d_l = \mathcal{I}^S(l)$. Because $\mathcal{I}^S(l) \leq \mathcal{I}^S(m)$ and because $\min\{\mathcal{I}_m, (1-\delta)v_m\} \geq d_m > \mathcal{I}^S(l)$, under the index policy of Theorem 1, product l is read only after product m is clicked upon. Once m is clicked, however, because $(1-\delta)v_m > \mathcal{I}^S(l)$, l is never read. Hence, l will not be purchased. Next suppose that $d_l = \mathcal{I}_l$. Then $\min\{\mathcal{I}_m, (1-\delta)v_m\} \geq d_m > \mathcal{I}_l$. Thus, product l is clicked only after m is clicked. But again, once m is clicked, because $(1-\delta)v_m > \mathcal{I}_l$, l is never clicked, implying that l is not purchased. Finally, suppose $d_l = (1-\delta)v_l$. Then because $\min\{\mathcal{I}_m, (1-\delta)v_m\} \geq d_m > (1-\delta)v_l$, m must be clicked before l is purchased. Because $v_m > v_l$, l is not purchased after m's value is learned.

Case 2: $l < m$ (i.e., l is read before m is read). Because $\mathcal{I}^S(m) \geq d_m > d_l = \min\{\mathcal{I}_l, (1-\delta)v_l, \mathcal{I}^S(l)\}$, and because $\mathcal{I}^S(m) \leq \mathcal{I}^S(l)$, it must be that $d_l = \min\{\mathcal{I}_l, (1-\delta)v_l\}$ and hence

$$\min\{\mathcal{I}_l, (1-\delta)v_l\} < d_m \leq \min\{\mathcal{I}_m, (1-\delta)v_m\}. \tag{30}$$

Furthermore, because the search technology is non-improving, $\mathcal{I}^S(l+1) \geq ... \geq \mathcal{I}^S(m-1) \geq$

[53] This follows from the fact that the index of a ξ-product that has been approved already is higher than the index of any ξ-product that has not been approved yet.

$\mathcal{I}^S(m)$. Along with the fact that $d_l = \min\{\mathcal{I}_l, (1-\delta)v_l\} < d_m \leq \mathcal{I}^S(m)$, this implies that $\min\{\mathcal{I}_l, (1-\delta)v_l\} < \mathcal{I}^S(k)$ for all $(l+1) \leq k \leq m$. This last property in turn implies that either clicking on l, or purchasing l, is dominated by reading any product k, with $(l+1) \leq k \leq m$. If m is read, then (30) implies that l will not be purchased (the arguments are similar to those for case 1). If, instead, m is not read, it must be that another product $k \neq l, m$ is purchased. In either case, product l is not purchased. ■

Proof of Proposition 2. The proof is in two steps. Step 1 shows that $\mathcal{I}^S(m) \geq \max_{l<m}\{w_l\}$ is necessary for product m to be read and that $\mathcal{I}^S(m) > \max_{l<m}\{w_l\}$ implies that product m is necessarily read. Step 2 shows that product m is read and clicked only if

$$\mathcal{I}^S(m) \geq \max_{l<m}\{w_l\} \ \text{ and } \ \mathcal{I}_m \geq \max\left\{\max_{l>m}\{d_l\}, \max_{l<m}\{w_l\}\right\} \tag{31}$$

and that, when both of the above inequalities are strict, product m is necessarily read and clicked. The result in the proposition then follows directly from the above properties along with the definition of $CTR(m)$.

Step 1. To see that $\mathcal{I}^S(m) \geq \max_{l<m}\{w_l\}$ is necessary for product m to be read, suppose that, for some $l < m$, $w_l > \mathcal{I}^S(m)$. That is, both the index corresponding to clicking on product l, $\mathcal{I}_l$, and the one corresponding to purchasing product l, $(1-\delta)v_l$, are strictly greater than $\mathcal{I}^S(m)$. Because product l is read before product m is read, by Theorem 1 in the main text, m is never read.

Next, we show that, when $\mathcal{I}^S(m) > \max_{l<m}\{w_l\}$, product m is always read. To see this, note that since the search cost $c(\cdot)$ is increasing, $\mathcal{I}^S(1) \geq ... \geq \mathcal{I}^S(m-1) \geq \mathcal{I}^S(m)$. Therefore, $\mathcal{I}^S(m) > \max_{l<m}\{w_l\}$ implies that, for any $1 \leq l \leq m$, $\mathcal{I}^S(l) > w_{l-1} = \min\{\mathcal{I}_{l-1}, (1-\delta)v_{l-1}\}$. Hence, by Theorem 1, for any $1 \leq l \leq m$, it cannot be that product $l-1$ is purchased before the product l is read. Repeatedly applying this argument for all $1 \leq l \leq m$ implies product m must be read before any product $l < m$ is purchased.

Step 2. To see that both inequalities in (31) must hold for product m to be read and clicked, first observe that we already established in Step 1 that the first inequality in (31) is necessary for product m to be read. Thus assume that such inequality holds. To see that the second inequality in (31) must also hold, suppose that $\mathcal{I}_m < \max\left\{\max_{l>m}\{d_l\}, \max_{l<m}\{w_l\}\right\}$. Then either there exists a product $l < m$ such that $w_l > \mathcal{I}_m$, or a product $l > m$ such that $d_l > \mathcal{I}_m$, or both. Suppose there is a product $l < m$ such that $w_l > \mathcal{I}_m$. Then product m cannot be clicked, because product l is necessarily read before m and, because both $\mathcal{I}_l$ and $(1-\delta)v_l$ are strictly greater than $\mathcal{I}_m$, product l is purchased before m is clicked. Next, suppose that there exists a product $l > m$ such that $d_l = \min\{\mathcal{I}^S(l), \mathcal{I}_l, (1-\delta)v_l\} > \mathcal{I}_m$. By the monotonicity of the search indexes, $\mathcal{I}^S(m) \geq \mathcal{I}^S(m+1) \geq ... \geq \mathcal{I}^S(l)$. That $\mathcal{I}^S(l) > \mathcal{I}_m$, then implies that $\mathcal{I}^S(k) > \mathcal{I}_m$ for any $k = m, m+1, ..., l$. In turn, this last property implies that clicking on m is dominated by reading product k, for any $k = m+1, ..., l$. If product l is read, because both $\mathcal{I}_l$ and $(1-\delta)v_l$ are strictly greater than $\mathcal{I}_m$, product m is not clicked. If, instead, product l is not read, it must be that another product $k \neq l, m$, with $k \in \{m+1, ..., l-1\}$, is purchased. In either case, product m is not clicked. Hence, both inequalities in (31) are necessary for product m to be read and clicked.

Next, we show that when both inequalities in (31) are strict, product m is necessarily read and clicked. We already established in Step 1 that, when the first inequality in (31) is strict, product m is read. Now suppose that the second inequality is also strict. That $\mathcal{I}_m > \max_{l<m}\{w_l\}$ implies that for each product $l < m$, either $\mathcal{I}_l$ or $(1-\delta)v_l$ are strictly smaller than $\mathcal{I}_m$. Because product m is read, by Theorem 1, it cannot be that any product $l < m$ is purchased before product m is clicked. Similarly, that $\mathcal{I}_m > \max_{l>m}\{d_l\}$ implies that, for each $l > m$, either $\mathcal{I}^S(l)$, or $\mathcal{I}_l$, or $(1-\delta)v_l$ are strictly smaller than $\mathcal{I}_m$, which again guarantees that no product $l > m$ can be purchased before product m is clicked. Since one of the products is necessarily purchased (product 0 representing the outside option), it must be that product m is clicked. Hence, we conclude that when both inequalities in (31) are strict, product m is necessarily read and clicked.

■