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

Daniel Fershtman† Alessandro Pavan‡

September 2024

Abstract

A decision-maker alternates between exploring alternatives in the consideration set and searching for new ones. The problem admits an index solution with a novel recursive index for the expansion of the consideration set. When the expansion technology is stationary, or improving, alternatives are replaced at each expansion. When, instead, it deteriorates, alternatives are revisited, and each expansion is treated as the last one. Applied to recruitment, the model shows how hard affirmative action might be detrimental to minority hiring. Applied to the administration of medical treatments, it shows why improvements in treatments may trigger a decline in the administration of such treatments and favor the discovery of new ones. Applied to online consumer search, it rationalizes non-sequential-non-cascading clicking and explains why the generalized second-price auction may lead to inefficient ad assignments even under its ascending-clock implementation.

Keywords: Search, Experimentation, 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. Nageeb Ali, Arjada Bardhi, Dirk Bergemann, Eddie Dekel, David Dillenberger, Laura Doval, Kfir Eliaz, Teddy Kim, Stephan Lauermann, Charles Manski, Benny Moldovanu, Xiaosheng Mu, Derek Neal, Michael Ostrovsky, David Pearce, Philip Reny, Andrew Rhodes, Eran Shmaya, Andy Skrzypacz, Rani Spiegler, Bruno Strulovici, Asher Wolinsky, and Jidong Zhou provided very useful comments and suggestions. We also thank Matteo Camboni and Tuval Danenberg for outstanding research assistance. Fershtman gratefully acknowledges funding from ISF grant #1202/20. The usual disclaimer applies.

†Emory University and Tel Aviv University. Email: danfershtman@gmail.com

‡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 implications for the decision to expand the CS in the first place. Finally, the DM may have limited knowledge about the alternatives outside the CS, and/or her ability to bring new alternatives to it, and may revise her beliefs about the process governing the "search technology" based on the results of past searches.

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). The model allows for gradual resolution of uncertainty, with each alternative explored (possibly infinitely) many times before its characteristics are fully discovered.[2] 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 show that the solution to the above problem 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 on the state of the search technology. The latter summarizes the results of past searches as well as the DM's beliefs about the process governing the discovery of new alternatives. The search index does not depend on the information

[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.

[2] As we clarify in due time, allowing for gradual resolution of uncertainty is important for many of the predictions of the model.

generated by the exploration of any of the alternatives already in the CS. Importantly, the search index 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. We provide a recursive characterization of such an index which we then use for many of our results. The optimal policy consists in selecting at any period the alternative for which the index is the highest.

Our environment can be viewed as a special case of the branching problem in the operations-research literature, where certain arms, after being activated, branch into new ones and then disappear (see, e.g., Weiss, 1988, Weber, 1992, and Keller and Oldale, 2003). In our problem, the decision to expand the CS corresponds to the activation of a branching arm that yields negative rewards (in the form of search costs) and brings a stochastic set of new arms according to a distribution that depends on the results of past searches. Our proof of indexability, however, is based on a novel recursive characterization of the search index which facilitates its computation and permits us to uncover various properties of the dynamics of exploration and CS expansion that are relevant for economic applications. (1) At any point in time, the decision to search for new alternatives depends on the current CS only through (a) the value of the alternative with the highest index, and (b) the state of the search technology. This property 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.[3] (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. (3) 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/or sequential learning

[3]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, and how to alternate between the exploration of such arms and further expansions of the CS, which is what is investigated in the first place. In other words, the selection of the "meta" search arm also involves the choice of "how to use it" and not merely "for how long to use it". Dynamic problems with "meta" arms rarely admit an index solution. See also the discussion in footnote 20.

problems. We consider three in the paper. The first one is recruitment, where the pool of candidates is expanded over time based on the evaluations of those candidates already in the pool. We use the model to investigate the effects of affirmative action aimed at facilitating the hiring of minority candidates. In Fershtman and Pavan (2021), we consider a special version of the general problem examined here to study "soft" affirmative action, that is, changes in the search technology such as the decision or mandate to focus on districts populated primarily by minorities, motivated by the intent to facilitate the hiring of minority candidates.[4] The application to recruitment considered in the present paper, instead, uses the general model developed here to study "hard" affirmative action, namely the reduction in the qualification standards for minority candidates. We show that, when the resolution of uncertainty is immediate (that is, when it takes a single evaluation to find out a candidate's type, or more generally to discover whatever can be learned about that candidate, so that further evaluations provide no extra information), then, no matter the search technology, affirmative action is guaranteed to increase the chances that the available positions are given to minority candidates. When, instead, the resolution of uncertainty is gradual, meaning that the discovery of a candidate's type may require multiple (possibly infinitely many) evaluations, these policies may backfire, leading to a reduction in the probability that the positions are assigned to minorities. This is because, when the resolution of uncertainty is gradual, these policies may induce the recruiter to favor the expansion of the candidate pool over a more thorough evaluation of candidates whose earlier examination yielded primarily negative outcomes. When this happens, minority candidates may not be given "a second chance," and this may happen even when hiring these candidates is more valuable than hiring majority ones. When, instead, the resolution of uncertainty is immediate, a reduction in the qualification standards for minorities increases both the index of unexplored minority candidates and the search index. However, whenever the latter is larger than the former under the new standards, the same is true under the old standards. In other words, such policies never induce the recruiter to search for new candidates when she would have evaluated minority candidates under the old standards. In this case, hard affirmative action is guaranteed to favor the hiring of minorities.

Our second application is online consumer search. We consider the problem of a consumer alternating between (a) reading new ads (bringing the corresponding products to the CS), (b) clicking on the ads of those products already in the CS (revealing the products' value to the consumer, net of the purchasing price), and (c) finalizing the purchase with one of the visited vendors. Contrary to recruitment, in this problem, it seems natural to assume that the resolution of uncertainty is immediate, meaning that a product's value to the consumer is learned after the first exploration (namely, after visiting a vendor's website). Under such an assumption,

[4]See also Bardhi, Guo, and Strulovici (2024) for the effects of initial asymmetries across alternatives on the alternatives' long-run utilization and their implications for minority hiring, in a setting with an exogenous pool of candidates.

the consumer's problem is a generalization of Weitzman (1979)'s problem to a setting with an endogenous set of boxes (we explain the connection in detail in the Supplement). Other models of online consumer search typically assume that positions' CTRs are either exogenous (e.g., Edelman et al., 2007) or are such that consumers click on the ads in the order in which the ads are displayed, which implies that firms advertising on a platform experience externalities only from the ads displayed at higher positions than the one their ad occupies (see, e.g., Athey and Ellison, 2011). Neither assumption squares well with empirical findings. For example, Jeziorski and Segal (2015) show that (1) approximately half of the users do not click on ads sequentially in the order they are displayed (*non-sequential clicking),* (2) over half of the users who click more than once, click on a higher position's ad after clicking on a lower position's ad (*non-cascading clicking*), and (3) the rate at which an ad at a given position is clicked upon depends not only on the ads displayed at higher positions but also on the ads displayed at lower positions (*externalities from lower positions*). We show that these phenomena are consistent with the dynamics in our model, and, perhaps surprisingly, that this is the case even when the consumer expects lower positions, on average, to be occupied by ads that are less attractive (in a sense made precise below) than those displayed at higher position, and the cost of reading increases with the number of ads brought to the CS.[5] We also show that these results have implications for bidding in sponsored-search auctions and for the efficiency of the allocations sustained in equilibrium. The analysis uncovers that the generalized second-price auction may lead to inefficient assignments even under its ascending-clock implementation considered in the literature (see, e.g., Edelman et al., 2007 and Gomes and Sweeney, 2014). The model also explains why a firm advertising on a platform may experience a decline in its profits when the probability it is given slots for additional products advertised by a platform at downstream positions increases. Such a decline can occur even if the extra ad is unambiguously profitable when brought exogenously to the consumer's CS.

Our third application is the administration of medical treatments, a classic problem that has received a lot of attention in the statistical medical literature. In this problem, the DM (e.g., a physician) sequentially administers treatments of different categories trading off patients' wellbeing with the information about the treatments' efficacy which is potentially useful when treating other patients. We enrich this canonical problem by giving the physician the possibility to search for new treatments when those previously administered led to unsatisfactory results. This pos-

[5]Note that the endogeneity of the CS is important for these results. A sequential search model in which the set of products is known to the consumer from the outset (such as Weitzman (1979)'s Pandora's boxes model with an exogenous set of boxes) fails to deliver any structural relationship between the positions at which the ads are displayed and the corresponding CTRs. In the Supplement, we also explain why CTRs need not be monotone in the ads' positions, even when the purchasing probabilities are, and why the "values per click" (VPCs) are naturally position-specific. These results are obtained by establishing a structural relationship relating the probability each product is purchased to the primitives of the problem (consumers' realized values, ads' positions, search costs, and consumer's beliefs over the attractiveness of the products displayed at the various positions).

sibility is natural and has been discussed at length in the statistical medicine literature (see also "How Physicians Can Keep Up with the Knowledge Explosion in Medicine", 2016, *Harvard Business Review*). However, to the best of our knowledge, it has not been captured in formal models. In this application, as in the recruitment one, the resolution of uncertainty is gradual, with the efficacy of each treatment learned through multiple (possibly infinitely many) administrations. We show that such a property has important implications for how improvements in a category of treatments may lead to the discovery of treatments from other categories, eventually crowding out the category that experienced the improvement, a property broadly consistent with what is discussed in the medical literature but which, as our results show, never obtains under an exogenous set of treatments, or when the resolution of uncertainty is immediate.

Our environment allows for general search (i.e., CS expansion) and exploration technologies. In particular, it allows for gradual resolution of uncertainty about the alternatives already in the CS (as in Gittins and Jones, 1974) while accommodating for a general process governing the discovery of new alternatives. This generality is important, because it permits us to study applications in which the exploration of each alternative unfolds over multiple periods, with the DM going back and forth over multiple alternatives, and in which the DM's beliefs over what can be obtained by expanding the CS evolve endogenously over time.

Outline. The rest of the paper is organized as follows. The remainder of this section discusses the pertinent literature. Section 2 introduces the model. Section 3 characterizes the optimal policy and identifies key properties for the dynamics of experimentation and CS expansion. Section 4 contains the three applications mentioned above. Section 5 concludes with a brief discussion of possible enrichments of the baseline model. All proofs are either in the Appendix at the end of the document or in the Supplement. The latter also contains additional material. In particular, Section S.3 in the Supplement shows how to use Theorem 1 and Proposition 1 to arrive at a solution to the extension of Weitzman (1979)'s Pandora's boxes problem with an endogenous CS, under general search technology and arbitrary categories of boxes. Section S.4 contains various additional results for the application to online consumer search. Section S.5 shows how the indexability of the optimal policy extends to certain problems with irreversible choice, whereas Section S.6 discusses why an index policy need not be optimal in the presence of "meta" arms with associated super-processes.

1.1 Related literature

To the best of our knowledge, the problem studied in the present paper (where the DM alternates between exploring "arms" already in the CS and stochastically expanding the latter) is new. As mentioned above, this problem can be viewed as a special case of the branching problem in the operations-research literature (see, e.g., Weiss, 1988, Weber, 1992, and Keller and Oldale, 2003).

Our recursive characterization of the index for CS expansion is new and exploits a novel classification of the alternatives into "categories," with the latter summarizing all characteristics of the arms that are relevant for the dynamics of the technology governing the expansion of the CS. This characterization permits us to arrive at a novel representation of the DM's payoff under the index policy, which we use to establish the optimality of such a policy. Importantly, the same recursive characterization is instrumental to all our subsequent results. It favors the computation of the index in applications, permits us to identify various properties of the dynamics of exploration and CS expansion, and uncovers new economic insights, such as those in Propositions 2-5. The paper is not a contribution to the operations-research bandit literature. It contributes to the economics literature by identifying the implications of a novel trade-off (between CS expansion and further examination of alternatives already in the CS), as well as in modeling the evolution of the search technology (by means of categories) in a way that isolates what is necessary for indexability and favors tractability and the uncovering of novel predictions.

The paper is also related to work studying experimentation and sequential learning with an exogenous CS,[6] e.g., Weitzman (1979), Austen-Smith and Martinelli (2018), Fudenberg, Strack and Strzalecki (2018), Gossner, Steiner and Stewart (2021), Ke and Villas-Boas (2019), and Ke, Shen, and Villas-Boas (2016). The problem studied in these papers involves a DM acquiring costly information about a set of options before stopping and choosing one of them. Related are also Che and Mierendorff (2019) and Liang, Mu, and Syrgkanis (2022). Che and Mierendorff (2019) studies the optimal sequential allocation of attention to two different signal sources biased towards alternative actions. Liang, Mu, and Syrgkanis (2022) studies 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 this set over time in response to the information she collects about the alternatives already in it. Related are also Garfagnini and Strulovici (2016) and Carnehl and Schneider (2023). The first paper considers a setting where 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] The second paper studies 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 these papers are related 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.

Pandora's boxes problem, as studied in Weitzman (1979), is a special multi-armed bandit

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

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

problem with immediate resolution of uncertainty. As such, it can be considered a special case of the general class of bandit problems with gradual resolution of uncertainty studied by Gittins and Jones (1974). 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. In the marketing literature, Greminger (2022) studies a consumer search problem with an endogenous set of products, focusing on the comparison between directed and undirected search. In the consumer search application, instead, we use the model to explain the phenomenon of non-sequential-non-cascading clicking observed in the empirical literature, and to draw novel implications for the design of sponsored search auctions. Au and Whitmeyer (2023) endogenize the collection of prize distributions from which a consumer samples in a model of oligopolistic competition with search frictions, where competing firms with products of unknown quality advertise how much information a consumer's visit will reveal.

The analysis in Section 4.2 is related to a broad literature on sponsored search (in addition to the papers by Edelman et al., 2007, Athey and Ellison, 2011, and Gomes and Sweeney, 2014, cited above, see also Edelman and Schwarz, 2010 and the references therein). Our contribution is in showing how to endogenize the click-through-rates and the value-per-click (with the latter naturally varying with the position occupied by the ads), and then using the characterization to explain dynamics of clicking and purchases that are consistent with the empirical literature, as reported in Jeziorski and Segal (2015), but not with standard search models.

Finally, the paper is part of a fast-growing literature on CS.[8] Eliaz and Spiegler (2011) study implications of different CS on firms' behavior, assuming such sets are exogenous. Manzini and Mariotti (2014) and Masatlioglu, Nakajima, and Ozbay (2012), instead, identify CS from choice behavior. Caplin, Dean, and Leahy (2019) provide necessary and sufficient conditions for rationally-inattentive agents to focus on a subset of all available choices, thus endogenizing the CS. 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 CS. Rather than committing to a CS up front and proceeding to evaluate its alternatives, the DM gradually expands the CS, in response to the results obtained from the exploration of the alternatives in the set.

[8]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).

2 Model

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

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

Alternatives, categories, learning, and payoffs. Each alternative belongs to a fixed *category* $\xi \in \Xi$ that is observed by the DM when the alternative is brought to the CS.[9] 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 of an alternative, and $\vartheta^{m-1} \equiv (\vartheta_s)_{s=0}^{m-1}$ its 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 about it, drawn from some distribution $G_\xi(\vartheta^{m-1}; \mu)$ and updates her beliefs about μ using Bayes' rule. Importantly, signal realizations are drawn independently across alternatives, given the alternatives' categories. The flow payoff u that the DM obtains from exploring an alternative from category ξ with parameter μ for the m-th time is drawn from a distribution $L_\xi(\vartheta^{m-1}; \mu)$.[10]

The search (i.e., CS expansion) technology. When the DM searches for the k-th time, she incurs a cost c_k and discovers alternatives of different categories. Let $E_k = (n_k(\xi) : \xi \in \Xi)$ denote the complete description of the alternatives identified through the k-th search, with $n_k(\xi) \in \mathbb{N}$ representing the number of category-ξ alternatives discovered. Let $(c_k, E_k)_{k=0}^{m-1}$ denote the history of the past $m-1$ search outcomes. 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

[9] The set of categories, Ξ, is measurable and need not be finite.

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

(e.g., learning by doing and/or fatigue).

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

Objective. A *policy* χ for the decision problem described above is a rule specifying, for each period t, whether to experiment with one of the alternatives in the CS C_t or expand the latter through search. A policy χ is *optimal* if, after each period t, it maximizes the expected discounted sum $\mathbb{E}^{\chi}\left[\sum_{s=t}^{\infty} \delta^s U_s | \mathcal{S}_t\right]$ of the flow payoffs, where $\delta \in (0,1)$ denotes the discount factor, U_s denotes the flow period-s payoff (with the latter equal to the search cost in case search is conducted in period s), $\mathcal{S}_t$ denotes the state of the problem in period t (the latter specifies, for each alternative in the CS, the history of signals, along with the history of all past search outcomes; see Section 3 for the formal definition) and $\mathbb{E}^{\chi}\left[\cdot | \mathcal{S}_t\right]$ denotes the expectation under the endogenous process for the flow payoffs obtained by starting from the state $\mathcal{S}_t$ and following the policy χ at each period $s \geq t$. To guarantee that the process of the expected payoffs is well behaved, we assume that, for any t, any $\mathcal{S}_t$ and any χ, $\delta^t \mathbb{E}^{\chi}\left[\sum_{s=t}^{\infty} \delta^s U_s | \mathcal{S}_t\right] \to 0$ as $t \to \infty$.[11]

Remark. The model above describes an infinite-horizon experimentation problem (with endogenous set of alternatives) in which payoffs are accumulated alongside learning. However, flow payoffs and learning need not be intertwined. In Sections 4.1-4.2, we consider settings in which the DM sequentially decides between learning about alternatives in the CS and expanding the CS, until a final choice is made among the alternatives in the CS, ending the decision problem. In the Supplement, we discuss how the results extend to a broader family of problems where the DM needs to irreversibly stop learning in order to be able to accumulate rewards.

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.[12] While the category ξ is fixed, the history θ of past

[11]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.

[12]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.

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

3.1 Optimal policy

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

[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.

equivalent to the one in Gittins and Jones (1974). As is well known, the optimal stopping rule in the definition of the index is the first period (after the one at which the index is computed) at which the index falls weakly below the value at the time the index was computed (see, e.g., Mandelbaum, 1986).

Given each state $\mathcal{S} = (\omega^S, \mathcal{S}^P)$ of the decision problem, denote the maximal index among the alternatives within the CS by $\mathcal{I}^*(\mathcal{S}^P)$.[16]

We now define an index for search (i.e., expansion of the CS). This index depends on the state $\omega^P(\xi, \theta)$ of each alternative in the CS only through the alternative's category ξ, which is the only component of ω^P that is relevant for the state ω^S of the search technology. Analogously to the indexes defined above, the index for search is defined as the maximal expected average discounted net payoff, per unit of expected discounted time, obtained between the current period and an optimal stopping time. Contrary to the standard indexes, however, the maximization is not just over the stopping time, but also over the rule governing the selection among the new alternatives brought to the CS by the current and further searches as well as the decision of when to further expand the CS. Denote by τ a stopping time, and by π a rule prescribing, for any period s between the current one and the stopping time τ, either the selection of one of the new alternatives brought to the CS by search or further search. Importantly, π selects only among search and alternatives that are not already in the CS when the decision to search is made.[17]

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

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

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

Definition 1 (**Index policy**). *The* ***index policy*** χ^* *selects at each period t the option with the greatest index given the overall state $\mathcal{S}_t = (\omega^S, \mathcal{S}^P)$ of the decision problem: search if $\mathcal{I}^S(\omega^S) \geq \mathcal{I}^*(\mathcal{S}^P)$, and an arbitrary alternative with index $\mathcal{I}^*(\mathcal{S}^P)$ if $\mathcal{I}^S(\omega^S) < \mathcal{I}^*(\mathcal{S}^P)$.*[18]

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

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

[17] 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.

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

optimal policy, we first introduce the following notation. For any $v \in \mathbb{R}$,let $\kappa(v) \in \mathbb{N} \cup \{\infty\}$ denote the first time at which, when the DM follows the index policy χ^*, (a) the search technology reaches a state in which its index is no greater than v, and (b) *all* alternatives in the CS – regardless of when they were introduced into it – have an index no greater than v. That is, $\kappa(v)$ is the minimal number of periods until all indexes are weakly below v ($\kappa(v) = \infty$ if this event never occurs).[19]

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

Theorem 1 (**Optimal policy**).

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

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

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

where τ^* is the first time (strictly after the one at which the index is computed) at which $\mathcal{I}^S$ and the indexes of all the alternatives brought to the CS by the current and future searches fall weakly below the value $\mathcal{I}^S(\omega^S)$ of the search index at the time the latter is computed, and where all the expectations are under the process induced by the index policy χ^*.

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

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

As in the classic multi-armed bandit problem with an exogenous CS, independence across alternatives is the key assumption behind the optimality of the index policy. That is, the payoffs (and the signals) from the various alternatives are drawn independently across the alternatives, given the latter's categories, and the set of new alternatives brought to the CS at each expansion only depends on the state of the current CS through the number of alternatives from each category in the CS. Under such assumptions, the theorem establishes a generalization of the

[19] 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 $\kappa(v)$.

index Theorem, according to which selecting in each period the alternative, or search, with the highest index is optimal.[20]

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

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

3.2 Implications for dynamics of exploration and CS expansion

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

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

[20] The reason why indexability of the optimal policy is not obvious is that search is a "meta" arm that can be used in different ways. In particular, the selection of such an arm also requires the specification of how to use it (i.e., whether to explore one of the physical alternatives brought to the CS by search, or continue searching). While our results imply that search can effectively be treated as a meta arm with its own index, the result is not a priori warranted. Indeed, problems in which alternatives correspond to meta arms, i.e., to sub-problems with their own sub-decisions (sometimes referred to as super-processes), 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 these difficulties in the Supplement.

subsequent state $\hat{\omega}^S$, as defined in Part (ii), $J(\hat{\omega}^S)$ first-order stochastically dominates $J(\omega^S)$.[21]

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

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

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

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

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

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

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

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

7. ***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*

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

[21] That is, the search technology is deteriorating if, regardless of the outcome of past searches, for any k and any upper set $A \subset \mathbb{R} \times \mathbb{N}^{|\Xi|}$ (that is, any set $A \subset \mathbb{R} \times \mathbb{N}^{|\Xi|}$ such that for each $a_1, a_2 \in \mathbb{R} \times \mathbb{N}^{|\Xi|}$ with $a_2 \geq a_1$, $a_2 \in A$ if $a_1 \in A$), one has that $\Pr((-c_{k+1}, E_{k+1}) \in A) \leq \Pr((-c_k, E_k) \in A)$. This definition is quite strong. In more specific environments, such as those in Section 4 where there is an order on the set of categories Ξ, weaker definitions are consistent with the results below.

Part (1) of the proposition is an implication of Theorem 1. The result is not trivial, however, 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.

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

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

Part (4) of the proposition says that the continuation value when search is launched is invariant to the state of the CS. This follows from the fact that, without loss of optimality, the DM never comes back to any alternative in the CS after search is launched. The same property holds in case of improving search technologies.

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

Part (6) follows again from the recursive characterization of the stopping time in the index of search, as per Part (2) of Theorem 1. Recall that this time coincides with the first time at which the index of any physical alternative brought to the CS by the current or future searches, and the index of search itself, drop weakly below the value of the search index at the time the current search is launched. If the search technology is stationary, or deteriorating, the index of search falls (weakly) below its current value immediately after search is launched. Hence, $\mathcal{I}^S(\omega^S)$ is invariant to the outcome of any search following the current one, conditional on ω^S.

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

4 Applications

We now turn to applications. We distinguish between the case in which the type of each arm is discovered after its first exploration (as in Weitzman (1979)) or after (possibly infinitely) many explorations (as in Gittins and Jones, 1974). We refer to the first situation as "immediate resolution of uncertainty" and the second one as "gradual resolution of uncertainty." As the results below illustrate, the distinction plays an important role for the dynamics of exploration and CS expansion and for the predictions that the analysis delivers in applications.

4.1 Recruitment

Problem. A committee searches for a candidate to fill an available position. The hiring decision is subject to a third party's approval (e.g., a university Dean, upper management, or a board of directors). There are different categories of candidates, representing, for instance, heterogeneity in race, gender, or field of expertise. Each ξ-candidate, with $\xi \in \Xi$, can either be qualified ($\mu = 1$) or unqualified ($\mu = 0$) for the position, and such qualification is unknown. Conditional on their category, candidates' qualification is independent. Let $p^\xi(\emptyset)$ denote the prior probability that a ξ-candidate is qualified and $p^\xi(\theta)$ the posterior probability of the same event given the history $\theta = (\vartheta_s)_{s=1}^m$ of past evaluations. The cost of evaluating a candidate with history θ of past evaluations is $\lambda^\xi(\theta) \geq 0$.

A ξ-candidate can he hired only if he is believed to be qualified with sufficiently high probability. Formally, for any $\xi \in \Xi$, there exists a threshold $\Psi^\xi \in (0, 1]$ such that a ξ-candidate with history θ can be hired if and only if $p^\xi(\theta) \geq \Psi^\xi$. The threshold Ψ^ξ can be thought of as reflecting standards and/or priorities set by the third party, which may, but need not, coincide with the preferences of the search committee. To avoid trivialities, assume that $p^\xi(\emptyset) < \Psi^\xi$, so that each candidate must be evaluated at least once to be hired. The evaluation of each ξ-candidate ends when the candidate is deemed acceptable, i.e., when $p^\xi(\theta) \geq \Psi^\xi$.[22]

The value to the committee of hiring a ξ-candidate who is perceived as qualified with prob-

[22]This assumption is vacuously satisfied when the evaluation technology exhibits immediate resolution of uncertainty, as in Weitzman's (1979) Pandora's boxes problem. For more general technologies, the assumption guarantees that the optimal policy remains indexable despite the presence of a stopping problem generated by the fact that the hiring of a candidate ends the entire recruitment process. See also the discussion in Section S.5 in the Supplement. The assumption that each candidate's exploration ends when the candidate is deemed acceptable by the third party can be dispensed with when each evaluation either reveals that the candidate is unqualified ($\mu = 0$) or leads to an upward revision of the posterior that he is qualified, as in the no-news-is-good-news case often considered in the strategic experimentation literature.

ability $p^\xi(\theta)$ is $v^\xi f^\xi(p^\xi(\theta))$, where $v^\xi > 0$, and $f^\xi : [0,1] \to \mathbb{R}_+$. The case where f^ξ is constant corresponds to the case where the committee does not care about a candidate's qualification, for example because the latter may reflect dimensions that are relevant for the third party but not the committee. The committee's value of not hiring any candidate is zero.

Optimal policy. The model in Section 2 assumes no irreversible stopping. In the problem described above, instead, the recruitment process ends when the committee offers the position to one of the acceptable candidates. Nevertheless, we can map this problem with stopping into the general model of Section 2 as follows. Consider a fictitious environment in which (a) hiring is temporary, i.e., the committee can take the position back from a candidate hired in the past and re-assign it to another candidate, (b) each candidate can be hired irrespectively of whether or not he was found acceptable, and (c) the evaluation of each candidate that has not been found acceptable yet is obtained by hiring the candidate for one period. In this fictitious environment, the flow payoff of temporarily hiring an acceptable candidate with posterior $p^\xi(\theta) \geq \Psi^\xi$ is equal to $v^\xi f^\xi(p^\xi(\theta))(1-\delta)$, whereas the flow payoff of temporarily hiring a candidate that has not been deemed acceptable yet (i.e., with posterior $p^\xi(\theta)<\Psi^\xi$) is equal to the cost $-\lambda^\xi(\theta)$ of evaluating the candidate, when the hiring of such a candidate generates a new signal realization, and is equal to some loss $-L < 0$ otherwise. Clearly, this fictitious setting is a special case of the general model in Section 2. It is also easy to see that the recruitment problem in this fictitious environment is a relaxation of the one in the primitive environment of interest.

Next, observe that, in this fictitious environment, once a candidate's evaluation leads to a posterior $p^\xi(\theta) \geq \Psi^\xi$, the index of selecting such a candidate remains constant at all subsequent periods and equal to $v^\xi f^\xi(p^\xi(\theta))(1-\delta)$. Likewise, once a candidate's evaluations are over (meaning that no further information about the candidate's qualification can be obtained) and $p^\xi(\theta)<\Psi^\xi$, the index of selecting such a candidate at any subsequent period is equal to $-L$. Clearly, in this fictitious environment, under the optimal rule, once an acceptable candidate is selected, he is selected at all subsequent periods (this is because none of the indexes, including the index of search, changes). Because the recruitment problem in the fictitious environment is a relaxation of the one in the primitive environment, the optimal rule in the primitive environment coincides with that in the fictitious environment.

Hard affirmative action (HAA). The recruitment problem described above can be used, among other things, to study how changes in the acceptance thresholds Ψ^ξ affect the outcome of the recruitment process. Note that the thresholds Ψ^ξ have implications not only for the sequencing of the evaluations of the existing candidates, but also for the committee's incentives to expand the candidate pool. To make things concrete, suppose there are two categories of candidates, $\Xi = \{\alpha, \beta\}$, and suppose that the third party lowers the acceptance threshold for the α-candidates, with the intent of favoring the recruitment of a candidate from that category. When the α category corresponds to a "minority," such a policy can be interpreted as a form of "hard" affirmative action, or HAA (see, e.g., Schuck, 2002).[23]

The following result shows that such policies are guaranteed to deliver the desired outcomes when either (i) the CS is exogenous or (ii) the CS is endogenous but the candidates' evaluation exhibits immediate resolution of uncertainty, as in Weitzman's (1979) Pandora boxes problem. Interestingly, these policies are not guaranteed to deliver the desired results and can even hurt minorities under more general evaluation technologies involving gradual resolution of uncertainty.

Proposition 2 (Hard affirmative action). *Suppose the acceptance threshold for any α-candidate is lowered from Ψ^α to $\hat{\Psi}^\alpha < \Psi^\alpha$, whereas the one for any of the β-candidates is kept at Ψ^β.*

1. *If the candidates' evaluation technology exhibits immediate resolution of uncertainty (as in Weitzman, 1979), the ex-ante probability that the position is assigned to an α-candidate is (weakly) higher under the new threshold than under the original one, regardless of the search technology.*

2. *If, instead, the candidates' evaluation technology features gradual resolution of uncertainty, the ex-ante probability that the position is given to an α-candidate can be strictly lower under the new threshold than under the original one (that is, HAA can backfire).*

3. *Suppose the reduction in the acceptance threshold for the α-candidates results in an increase in the indexes $\mathcal{I}^P(\alpha, \theta)$ of the α-candidates. Such a reduction can reduce the ex-ante probability the position is assigned to an α-candidate only if the CS is endogenous.*

When the evaluation technology exhibits immediate resolution of uncertainty, a reduction in the acceptance threshold for the α-candidates strictly increases the index of evaluating any of the α-candidates. When the CS is exogenous, such an increase implies that, ex-ante, the position is more likely to be assigned to an α-candidate. Interestingly, and perhaps surprisingly, this is true also with an endogenous CS, no matter the technology governing the discovery of new candidates. The reason why the result is not immediate is that, with an endogenous CS, a reduction in the acceptance threshold for the α-candidates also increases the search index, as the new α-candidates brought to the CS are more likely to be hired after being evaluated. The latter property changes how the committee alternates between evaluating existing candidates and searching for new ones, with effects on the ex-ante probability that the position is given to an α-candidate.

The reason why, despite these effects, when the resolution of uncertainty is immediate, HAA is guaranteed to help the α-candidates is the following. Suppose that, when the acceptance threshold is $\hat{\Psi}^\alpha < \Psi^\alpha$, at a given history, the committee prefers searching for new candidates

[23] Employment quotas are another form of HAA. In contrast, policies that change the composition of the CS, for example by requiring that search be conducted in areas populated primarily by minorities are examples of soft affirmative action, or SAA (see, e.g., Fershtman and Pavan, 2021).

than evaluating one of the α-candidates in the CS. This means that $\hat{\mathcal{I}}^S(\omega^S) \geq \hat{\mathcal{I}}^P(\alpha, \emptyset)$, where $\hat{\mathcal{I}}^S(\omega^S)$ and $\hat{\mathcal{I}}^P(\alpha, \emptyset)$ denote, respectively, the search-index and the index for an α-candidate that has not been evaluated yet, under the new acceptance threshold $\hat{\Psi}^\alpha$. The recursive properties of the search index identified in the previous section imply that, under the optimal stopping time τ^* that characterizes the search index (the one in part 2 of Theorem 1), stopping occurs whenever the search index and the indexes of all the candidates identified through the expansion of the CS starting from state ω^S are weakly below $\hat{\mathcal{I}}^S(\omega^S)$. Furthermore, the policy π^* that maximizes the objective function in the definition of the search index prescribes selecting at each period the alternative (search, evaluation of a candidate not evaluated yet, or assignment of the position to one of the acceptable candidates) for which the index is the highest. Because $\hat{\mathcal{I}}^S(\omega^S) \geq \hat{\mathcal{I}}^P(\alpha, \emptyset)$, the above properties imply that, under the policies (τ^*, π^*) of part 2 of Theorem 1 characterizing the search index $\hat{\mathcal{I}}^S(\omega^S)$, none of the α-candidates is evaluated and hence assigned the position. In turn, this means that the search index $\hat{\mathcal{I}}^S(\omega^S)$ under the new acceptance threshold $\hat{\Psi}^\alpha$ is the same as the search index $\mathcal{I}^S(\omega^S)$ under the old threshold Ψ^α, i.e.,$\hat{\mathcal{I}}^S(\omega^S) = \mathcal{I}^S(\omega^S)$. Because, under the new acceptance threshold $\hat{\Psi}^\alpha$, the index $\hat{\mathcal{I}}^P(\alpha, \emptyset)$ of any α-candidate that has not been evaluated yet is strictly higher than the index $\mathcal{I}^P(\alpha, \emptyset)$ of any such candidate under the original threshold Ψ^α (i.e.,$\hat{\mathcal{I}}^P(\alpha, \emptyset) > \mathcal{I}^P(\alpha, \emptyset)$), this also means that, whenever the committee prefers expanding the CS to evaluating any of the α-candidates in the CS under the new acceptance threshold, the same is true under the old threshold. Likewise, whenever, under the new threshold, the committee prefers expanding the CS to assigning the position to any of the α-candidates in the CS found acceptable (that is, whenever $\hat{\mathcal{I}}^S(\omega^S) > v^\alpha f^\alpha(p^\alpha(\theta))$, for all α-candidates in the CS with $p^\alpha(\theta) \geq \hat{\Psi}^\alpha$), the committee would have expanded the CS also under the original threshold Ψ^α. To see this, observe that some of the α-candidates in the CS that are acceptable under the threshold $\hat{\Psi}^\alpha$ may not have been acceptable under the original threshold Ψ^α; assigning the position to any such candidate would have not been feasible under the original threshold. Thus consider α-candidates that are acceptable under both the original and the new threshold. That the committee prefers expanding the CS than assigning the position to any such α-candidate implies that $\hat{\mathcal{I}}^S(\omega^S) > v^\alpha f^\alpha(p^\alpha(\theta))$, with $p^\alpha(\theta) \geq \Psi^\alpha$. Again, the recursive characterization of the search index in part 2 of Theorem 1 implies that, under the optimal policies (τ^*, π^*) yielding the search index $\hat{\mathcal{I}}^S(\omega^S)$, the position is never assigned to any α-candidate with posterior below $p^\alpha(\theta)$. This means that $\hat{\mathcal{I}}^S(\omega^S) = \mathcal{I}^S(\omega^S)$. Hence, if the committee prefers expanding the CS to assigning the position to an α-candidate with posterior $p^\alpha(\theta) > \Psi^\alpha > \hat{\Psi}^\alpha$ under the new threshold, the same must be true under the old threshold. These properties, along with the fact that, under the new threshold, whenever the committee prefers exploring or assigning the position to a β-candidate to searching, exploring an α-candidate in the CS, or assigning the position to one of the acceptable α-candidates in the CS, the same is true under the original threshold, imply that the ex-ante probability the position is assigned to an α-candidate is higher under the new threshold.

Things are different when the candidates' evaluation features gradual resolution of uncertainty. In this case, HAA (in the form of a reduction in the acceptance threshold for the α-candidates) can backfire by reducing the probability the position is assigned to an α-candidate. This can happen because of two reasons. First, it may reduce the indexes $\mathcal{I}^P(\alpha, \theta)$ of the α-candidates. This happens when the early evaluations yield positive results that qualify the candidate, but also prevent a more thorough evaluation of the candidate (recall that a candidate's evaluation terminates as soon as the candidate is deemed acceptable). When this happens, the α-candidates may suffer from HAA even when the CS is exogenous. The second reason is more relevant and specific to settings in which the CS is endogenous. Even when the reduction in the acceptance threshold of the α-candidates yields higher indexes $\mathcal{I}^P(\alpha, \theta)$, the α-candidates may suffer because the reduction may induce the committee to stop evaluating those candidates whose earlier evaluations delivered negative outcomes and instead search for new candidates.[24] The change in the candidate pool in turn may favor candidates from other categories, at the expense of those candidates whose threshold has been lowered. That is, the reduction in Ψ^α may lead to a reduction in the ex-ante probability the position is given to an α-candidate. In the Supplement, we identify sufficient conditions for this to happen. Importantly, when the indexes $\mathcal{I}^P(\alpha, \theta)$ increase in response to a reduction in the acceptance threshold of the α-candidates, HAA can backfire *only* when the CS is endogenous, for it is only because of the crowding out effects of search on the evaluations of the α-candidates that the position is more likely to be given to one of the β-candidates.

4.2 Online consumer search

Problem and optimal policy. Consider the problem of a consumer searching online for a product to purchase. In this environment, the consumer brings a product to her CS by reading the product's ad. Because the consumer does not know which products are advertised at the various positions following the submission of a query on a search engine, it is plausible that the consumer reads the ads in the order they are displayed by the platform. After reading a new ad, the consumer brings the corresponding product to her CS. At that moment the consumer decides whether to read the next ad or click on one of the products whose ad the consumer has read already. After clicking on a product's link, the consumer is directed to the vendor's website where she learns her value for the vendor's product (net of the product's price). The consumer then decides whether or not to finalize the purchase. The purchase of a product brings to an end the consumer's search process. While the consumer naturally reads the ads in the order in which

[24] Note that, when the acceptance threshold Ψ^α is reduced, the indexes $\mathcal{I}^P(\alpha, \theta)$ of the α-candidates always increase if the committee does not value the candidates' qualification (i.e., if there exists $k > 0$ such that $f^\alpha(p) = k$ for all p).

they are displayed, she clicks on the links of the products whose ads have been read in the order of her choice. As we show in the Supplement, this problem too can be recasted in a way that permits one to use the results in the previous section to describe the consumer optimal search strategy.

In this problem, each category $\xi \in \Xi$ corresponds to a different ad's type, with each type indexing a different (absolutely continuous) distribution F^{ξ} from which the consumer's net value v for the corresponding ad's product is drawn, and a different inspection cost λ^{ξ} to learn such a value.[25] Each position $m \in \mathbb{N}$ is occupied by the ad of one and only one firm, with the same firm possibly advertising at multiple positions. Reading the m-th ad reveals to the consumer the ad's type for the firm advertising on the m-th position. We denote the ad's type of the firm occupying the m-th position by $\xi(m) \in \Xi$. The consumer believes that $\xi(m)$ is drawn from Ξ according to a distribution $\rho(m) \in \Delta(\Xi)$ that may depend on m but is invariant in the ads' types of those firms occupying the upstream positions $l < m$. For example, the consumer may expect lower positions to be occupied, on average, by lower-quality ads (that is, by products that are more costly to learn about, i.e., have a higher λ^{ξ}, and/or that deliver, on average, lower values, i.e., have a "smaller" F^{ξ}, in the sense of FOSD), but does not change her beliefs based on the ads' types $\xi(l)$, $l < m$, encountered at upstream positions.

Let $c(m)$ denote the cost of reading the m-th ad. As we show in the Supplement, the results in Theorem 1 and Proposition 1 above imply that the index for the decision to click on the m-th ad, after discovering the ad's type $\xi(m)$, is equal to

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

whereas the index for the decision to read the m-th ad is equal to

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

where, for any $l \in \mathbb{R}$, $\Xi(l) \equiv \left\{\xi \in \Xi : \mathcal{I}^P(\xi, \emptyset) > l\right\}$ denotes the set of ads' types whose clicking index exceeds l.

One can use the model to endogenize the probability with which the consumer reads the ads, clicks on them, and finalizes her purchases. Furthermore, one can derive a structural relationship between the various positions and their click-through-rates (CTRs), accounting for the uncertainty that the consumer faces about the ads displayed at the various positions – a feature that

[25] The assumption that each F^{ξ} is absolutely continuous is made in order 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.

the model with exogenous CSs does not capture.[26] We refer the reader to the Supplement, where we also show how the properties of Proposition 1 above permit one to express the consumer's eventual purchase decisions in terms of comparisons of the products' "discovery values". In a similar setting, but with an exogenous CS, Choi, Dai and Kim (2018) – and, independently, Armstrong (2017) – derive a static condition characterizing eventual purchases, based on a comparison of "effective values." In the Supplement, we show that such a characterization extends to search problems with an endogenous CS by combining effective values with discovery ones, with the latter accounting for the fact that products need to be brought to the CS before they can be explored.[27]

Non-sequential-non-cascading clicking. In their analysis of consumer demand for search advertising, Jeziorski and Segal (2015) document three properties of users' behavior that, while ubiquitous, are inconsistent with existing models of search advertising:

1. *Non-sequential clicking*: Nearly half of users do not click on ads sequentially in the order of the positions in which ads are presented.

2. *Non-cascading clicking*: Over half of users who click more than once click on a higher position after having clicked on a lower position.

3. *Externalities from lower positions*: The rate at which an ad at a given position is clicked depends on which ads are displayed below it.

The above properties are consistent with what is predicted by the model of consumer search described above. Perhaps surprisingly, this is so even when the consumer expects lower positions to be occupied by less attractive ads and the cost for bringing ads to the CS to increase with the number of ads read. For simplicity, let $\Xi = \mathbb{N}$, with higher ξ denoting "higher" distributions F^{ξ} and lower inspection costs λ^{ξ}; that is, for any $\xi', \xi'' \in \Xi$ with $\xi'' > \xi'$, $F^{\xi''} \succeq_{FOSD} F^{\xi'}$ and $\lambda^{\xi''} \leq \lambda^{\xi'}$ (with one of the two relationships strict). Let $\underline{\xi} \equiv \inf \Xi$ and $\overline{\xi} \equiv \sup \Xi$. Suppose the cost of reading $c(m)$ is non decreasing in m and the consumer expects lower positions to be occupied by less attractive ads, in the sense that, for all m, the distribution $\rho(m) \in \Delta(\Xi)$ over Ξ first-order-stochastically dominates, weakly, the corresponding distribution $\rho(m+1) \in \Delta(\Xi)$. Then $\mathcal{I}^S(m+1) \leq \mathcal{I}^S(m)$ for all m. Furthermore, for all m, $\mathcal{I}^S(m) \leq \mathcal{I}^P(\overline{\xi}, \emptyset)$, that is, the index of search is smaller than the index of any ad whose type is the most attractive one. We then have the following result:

Proposition 3 (**Clicking behavior**). *Suppose that cost of reading $c(m)$ is non decreasing in m and that the consumer expects lower positions to be occupied by less attractive ads, in the sense that,*

[26] A position's CTR is the probability that the ad displayed at such position is clicked.

[27] Greminger (2022) also finds that a certain version of the eventual purchase theorem extends to problems with an endogenous CS. However, none of the results here have a counterpart in that paper which instead focuses on the comparison between directed and undirected search.

for all m, $\rho(m) \succeq_{FOSD} \rho(m+1)$. The consumer's search for the optimal product is consistent with non-cascading and non-sequential clicking, and generates externalities from lower positions.

Importantly, these dynamics do not obtain under the search model of Athey and Ellison (2011). They can emerge in models of consumer search with an exogenous CS à la Weitzman, 1979 if one assumes a specific relationship between the positions and the ads' types. In such a model, the consumer's beliefs over the ads displayed at the various positions are degenerate, as the consumer knows (exogenously) which ad occupies each position. Hence, there is no reason why firms would want to bid more for higher positions, in contrast with what is documented in the literature on bidding for sponsored search.

The model of consumer search with an endogenous CS can also generate dynamics under which the probability of non-cascading and non-sequential clicking is non-monotone in the positions. To see this, suppose that $\Xi = \{\underline{\xi}, \overline{\xi}\}$ and that the conditions in Proposition 3 hold. Because $\mathcal{I}^S(m)$ is decreasing in m and $\mathcal{I}^S(m) \leq \mathcal{I}^P(\overline{\xi}, \emptyset)$ for all m, when $\mathcal{I}^S(0) > \mathcal{I}^P(\underline{\xi}, \emptyset)$, there exists a position m^* such that: (a) for any $m < m^*$, the consumer clicks on the m-th ad immediately after reading it if and only if it is of type $\overline{\xi}$, whereas (b) for any $m > m^*$, the consumer clicks on the m-th ad immediately after reading it, regardless of the ad's type. When, for any $m < m^*$, the probability that the m-th ad is of type $\overline{\xi}$ is strictly decreasing in m, we then have that the probability of non-sequential and non-cascading clicking is single-peaked (increasing in m for $m < m^*$, and equal to zero for $m > m^*$). In turn, it can be shown that the probability that the m-th ad is occupied by a firm of type $\overline{\xi}$ is indeed decreasing in m when firms' profits for selling to the consumer are drawn from a distribution that is related to the ads' types by MLRP and the assignment of the ads is governed by an auction that induces monotone bidding.

Implications for equilibrium bidding in sponsored-search auctions. We now show how the results above can be put to work to identify important properties of bidding in sponsored-search auctions. For simplicity, suppose there are only two firms, with each firm advertising a single product (the results below extend to auctions with more than two positions and more than two bidders). The two firms compete for placing their ads on a platform offering two different positions. The platform uses the ascending-clock version of the generalized second price (GSP) auction of Edelman et al. (2007) to allocate the two positions. The firm dropping out first is allocated the second position and pays nothing, whereas the other firm is allocated the first position and pays the price at which the other firm drops out, per click. As in the analysis above, each firm's product can be of multiple types, with each type $\xi \in \Xi$ parametrizing the attractiveness of the firm's product/ad (formally, the distribution F^{ξ} from which the consumer's value is drawn) and the consumer's exploration cost λ^{ξ}. Let z denote the profit each firm derives from selling its product and assume that each z is drawn from $[\underline{z}, \bar{z}]$ according to a distribution F_z, with the draws independent across firms. Assume that the payoff the consumer expects from learning her value for each firm's product (net of the exploration cost λ) exceeds her outside option, and that this is true for all possible types ξ.

The search model introduced above delivers a structural characterization of the CTRs, the purchasing probabilities, and the firms' values per click (VPC), for each possible profile of firms' types (ξ_1, ξ_2) and each possible assignment of the positions to the firms. It also delivers a characterization of the same variables from the perspective of an observer who does not know the firms' types. Formally, and consistently with the notation above, an assignment is a vector $(\xi(1), \xi(2))$ where the first entry denotes the type of firm occupying the top position, and the second entry the type of firm occupying the second position. For each position $m = 1, 2$ and each assignment $(\xi(1), \xi(2))$, let $P(m; \xi(1), \xi(2))$ denote the probability that the consumer purchases the product advertised in position m under the assignment $(\xi(1), \xi(2))$. Clearly, the consumer does not know the assignment $(\xi(1), \xi(2))$ at the beginning of the search process and learns it by reading the various ads.

Let $CTR(m; \xi(1), \xi(2))$ denote the probability that the consumer clicks on the ad displayed in the m-th position under the assignment $(\xi(1), \xi(2))$. Finally, for any position m, assignment $(\xi(1), \xi(2))$, and unit profit z, let

$$VPC(m; \xi(1), \xi(2), z) \equiv z \frac{P(m; \xi(1), \xi(2))}{CTR(m; \xi(1), \xi(2))} \tag{7}$$

denote the value-per-click (VPC) that a firm with unit profit z assigns to occupying the m-th position under the assignment $(\xi(1), \xi(2))$. Note that, contrary to Edelman et al. (2007) and Athey and Ellison (2011), the values per click here are not only heterogeneous across firms but also position-specific, reflecting the property that the probability the consumer finalizes a trade after clicking on a firm's ad depends on the position at which the ad is displayed, a property also documented by the empirical literature. In the Supplement, we provide a parametric example where all the above variables can be computed in closed form.

Suppose that the two firms commonly know the attractiveness of their ads/products, possibly as a result of past experiences with consumers who searched similar products on the same platform. We then have the following result:

Proposition 4 (Firms' bidding behavior). *Consider the sponsored-search model described above and assume that firms do not follow weakly dominated strategies. For any profile of ad types* (ξ_1, ξ_2)*, there exists a threshold* $b(z; \xi_1, \xi_2)$ *such that each firm with unit profit* z *drops out at price* $b(z; \xi_1, \xi_2)$ *irrespective of whether its ad is more or less attractive than the rival's.*

To gather some intuition, consider a state (ξ_1, ξ_2) in which firm 1's ad is more attractive than firm 2's (in the sense that $\lambda^{\xi_1} < \lambda^{\xi_2}$, meaning that the cost to the consumer to learn her value for firm 1's product is lower than for firm 2's product, and $F^{\xi_1} \succ_{FOSD} F^{\xi_2}$, meaning that product 1 delivers, on average, more utility than product 2).

Consider the interesting case in which the consumer finds it optimal to click on the ad encountered at the first position before reading the ad in the second position, irrespective of whether the ad in the first position is firm 1's or firm 2's. That is, the index $\mathcal{I}_1 \equiv \mathcal{I}^P(\xi(1), \emptyset)$ for clicking on the ad displayed in the first position is higher than the index $\mathcal{I}^S(2)$ for reading the ad in the second position, irrespective of the type $\xi(1)$ of ad encountered in the top position. When this property does not hold, firms are indifferent between advertising in the first and second position given that the consumer always clicks first the ad of the most attractive firm irrespectively of the position at which the ad is displayed. Firm 1 then finds it optimal to drop out at a price b defined by

$$[VPC(1;\xi_1,\xi_2,z) - b]\, CTR(1;\xi_1,\xi_2) = VPC(2;\xi_2,\xi_1,z) CTR(2;\xi_2,\xi_1). \tag{8}$$

The above indifference condition reflects the fact that both the CTRs and the VPCs are not only position-specific but also ad-specific. Equivalently, using the relation between VPCs, CTRs and selling probabilities P in Condition (7) above, we have that the price at which firm 1 drops out is equal to $z[P(1;\xi_1,\xi_2) - P(2;\xi_2,\xi_1)]$. Because the first position is clicked with probability one, the firm optimally drops out when the price reaches a value equal to the extra profit the firm expects from placing its ad on the first position instead of the second one. Note that the term in square brackets is the difference between the probability the firm assigns to selling its product when listed in the top position (that is, under the assignment (ξ_1, ξ_2)) and when listed in the second position (that is, under the assignment (ξ_2, ξ_1)).

Likewise, firm 2, given its markup z, drops out when the price reaches the value b implicitly defined by

$$[VPC(1;\xi_2,\xi_1,z) - b]\, CTR(1;\xi_2,\xi_1) = VPC(2;\xi_1,\xi_2,z) \cdot CTR(2;\xi_1,\xi_2). \tag{9}$$

Note that Condition (9) differs from Condition (8) because the two firms expect different CTRs and have different VPCs for the two positions. Equivalently, using again the relationship in (7), we have that the price at which firm 2 drops out is equal to $z[P(1;\xi_2,\xi_1) - P(2;\xi_1,\xi_2)]$. Clearly, the probability $P(1;\xi_2,\xi_1)$ that firm 2 assigns to selling its good when advertising in the top position is different from the probability $P(1;\xi_1,\xi_2)$ that firm 1 assigns to selling its product when occupying the same position, reflecting the difference in the distributions from which the consumer's values are drawn, the exploration costs, and the probability the consumer clicks on the second ad when encountering an ad of type ξ_1 or of type ξ_2 in the first position. However, the differential in the probability of selling when occupying the first and second position is the same for the two firms; that is,

$$P(1;\xi_2,\xi_1) - P(2;\xi_1,\xi_2) = P(1;\xi_1,\xi_2) - P(2;\xi_2,\xi_1). \tag{10}$$

As a consequence, in equilibrium, the price at which the two firms drop out is the same, despite the fact that one firm is more attractive than the other. Because, in each state, the two firms follow identical bidding strategies, a consumer who understands the rules of the auction should hold beliefs about the type of firms displaying in the second position that are invariant in the type of firm encountered in the first position.

The property in Proposition 4 has important implications for the efficiency of the equilibrium allocations under the ascending-clock implementation of the GSP auction considered in the literature. To see this, consider any welfare objective that assigns strictly positive weight to consumer surplus. We then have the following result:

Corollary 1 (**Inefficiency of equilibrium ad allocations**). *Assume that firms do not follow weakly dominated strategies. In each state in which the attractiveness of the firms' products is not homogeneous across firms, the positions are assigned inefficiently with strictly positive probability.*

Efficiency requires that, in each state in which the attractiveness of the two products is different (that is, $\xi_1 \neq \xi_2$), whenever the difference $|z_1 - z_2|$ between the two firms' profits is small, the firm whose product is the most attractive be assigned the top position. This is because the top position is clicked more often. Hence, when the first position is occupied by the most attractive firm, the chances the consumer purchases the product she values the most (formally, for which her ex-post net value v is the highest) are higher than when the top position is occupied by the least attractive firm. Hence, no matter the weight the planner assigns to consumer surplus in the welfare objective function, as long as the latter is strictly positive, the auction allocates the positions inefficiently with positive probability. The result is a direct consequence of the fact that, in each state, the two firms follow symmetric bidding strategies, which implies that the assignment of the two positions is based entirely on the firms' unit profits z and not their attractiveness.

The result in the corollary applies also to settings with more than two firms and more than two positions. To see this, it suffices to note that the situation described above continues to represent a valid description of the problem each firm faces when there are only two firms left in the auction.[28] The two key properties responsible for the result are (1) that firms possess some information about their attractiveness at the bidding stage, and (2) that the differential in the selling probabilities is equalized across the remaining firms, which is always the case when the remaining firms expect the buyer to purchase one of their products with certainty.

Detrimental effect of additional ad space. The model above can also be used 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 is endogenous. To

[28]The selling probabilities $P(m; \xi(1), \xi(2))$, click-through rates $CTR(m; \xi(1), \xi(2)))$, and values-per-click $VPC(1; (\xi_1, \xi_2, z)$, for positions $m = 1, 2$ should then be interpreted as given the assignment $(\xi(m))_{m>2}$ of the lower positions observed by the two remaining firms.

see this, consider the following situation. There are three types of ads, i.e., $\xi \in \Xi = \{A, B, C\}$. The consumer's initial CS contains three products, each from a different firm and each of a different type. By searching, the consumer is presented with a fourth product whose ad's type is drawn from Ξ according to $\rho \in \Delta(\Xi)$. As above, the consumer believes that an ad of type ξ, when clicked upon, yields the consumer a net value v drawn from a distribution F^{ξ}, independently across products.[29] The cost to the consumer of learning the value of a product whose ad's type is ξ is λ^{ξ}. The consumer has unit demand and each firm makes the same profit from selling one of its products.

Suppose that each firm's ads are all of the same type and that the ads of different firms are of different type (in other words, in this example, ξ also indexes the identity of each firm). As we show in the Supplement, an increase in the probability that the new search brings an additional product of type ξ (equivalently, an additional product from 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 the new product. When strong enough, such an effect may reduce the probability that one of firm ξ's products is eventually purchased, and hence firm ξ's profits. See the Supplement for the details.

4.3 Administration of medical treatments

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. 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, but may also reflect the compensation the physician receives, as well as the physician's reputation, which may be linked to whether or not the administered treatments are successful. Importantly, each outcome also provides the DM with information about the treatment's efficacy.

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 up 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 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} from which μ is drawn, as defined in the general model of Section 2, is Bernoulli with parameter $p^{\xi}(\emptyset)$).

[29] 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.

Administering 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 $\mu = 0$), a bad outcome is observed with probability $q_0^\xi = 1$.[30] Hence, in this example, the distribution $G_\xi(\vartheta^m; \mu)$ from which each ϑ is drawn, as defined in Section 2, 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 outcomes, denote by $p^\xi(\vartheta^{m-1})$ the posterior probability that the specific ξ-treatment is effective. The distribution $L_\xi(\vartheta^{m-1}; \mu)$ 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 $p^\xi(\vartheta^{m-1})$.

At any point in time, the physician can expand the CS by searching for new treatments. 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 search, the physician does not know which type of treatment she will encounter when expanding her CS). Hence, in this example, for each m, the distribution $J((c_k, E_k)_{k=0}^{m-1})$ from which (c_m, E_m) is drawn, as defined in Section 2, 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 $E_k = (0, 1)$.

Using Proposition 1, we can arrive at the following characterization of the indexes for the optimal policy (the derivations are in the proof of Proposition 5 in the Appendix). For any $\omega^P = (\xi, \theta)$, the index of a treatment in state ω^P is equal to

$$\mathcal{I}^P(\omega^P) = \frac{\left(1 - \delta + \delta q_1^\xi\right) p^\xi(\theta) q_1^\xi v^\xi}{1 - \delta + \delta p^\xi(\theta) q_1^\xi}. \tag{11}$$

The index of search is invariant to ω^S and equal to[31]

$$\mathcal{I}^S = \frac{(1-\delta)\left\{\sum_{\xi \in \{\alpha,\beta\}} \rho^\xi \mathbb{E}\left[\sum_{s=0}^{\tau^{\xi*}-1} \delta^s u_s \middle| (\xi, \emptyset)\right] - c\right\}}{1 - \sum_{\xi \in \{\alpha,\beta\}} \rho^\xi \mathbb{E}\left[\delta^{\tau^{\xi*}} \middle| (\xi, \emptyset)\right]}, \tag{12}$$

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.

The model can be used to study, among other things, the effects of improvements in a cat-

[30] The results below extend to $q_0^\xi \in (0, 1]$, provided that $1 - q_0^\xi < q_1^\xi$, which guarantees that the administration of the treatments carries useful information.

[31] The expectations in the formula in (12) are under a rule selecting in each period the ξ-treatment brought to the CS by search.

egory of treatments on the administration of such treatments, as well as on the discovery and administration of treatments from other categories. These improvements may occur because of technological advancement, changes in pricing, or changes in the incentive schemes offered to the physician for the administration of the treatments. With an endogenous CS, these improvements also affect the desirability of expanding the CS by searching for new treatments when the administration of those already in the CS yields negative outcomes. The expansion of the CS, in turn, may favor the discovery and administration of treatments from categories other than the one that experienced the boost in attractiveness. These spillover effects, which are absent in the canonical model with an exogenous CS, have been discussed at large in the medical literature, which has documented how the discovery and subsequent administration of certain drugs occurred due to the improvement of drugs meant for alternative uses (e.g., several central nervous system therapies were discovered as the result of the search for drugs for the cardiovascular system). When each treatment's efficacy is learned gradually over time, that is, after (possibly infinitely) many administrations, as is typically assumed in the medical literature, such spillover effects may crowd out the administration of the treatments that have become more attractive, as the next proposition shows.

Proposition 5 (**Administration of medical treatments**). *Assume the α-treatments become more attractive, either as the result of an increase in the ex-ante probability $p^{\alpha}(\emptyset)$ they are effective, or because the value v^{α} the physician assigns to administering such treatments, when delivering good outcomes, increases. (a) When the CS is exogenous, or the efficacy of each treatment is discovered after its first administration (i.e., $q_1^{\xi} = 1$, as in Weitzman's (1979) model), such improvements always lead to an increase in the administration of the α-treatments.*[32] *(b) When, instead, the CS is endogenous and learning the efficacy of the treatments may require (possibly infinitely) many administrations ($q_1^{\xi} \in (0,1)$), these improvements may lead to a reduction in the administration of the α-treatments.*[33]

As explained above, when the CS is endogenous, an improvement in the α-treatments may also increases the value of expanding the CS; this is because such expansions are expected to bring α-treatments that are seen as more attractive. When the resolution of uncertainty is immediate (i.e., when the efficacy of each treatment is discovered after its first administration or, more generally, when subsequent administrations do not deliver additional information), the increase in the value of expanding the CS (as captured by the increase in the search index) never crowds out the administration of the α-treatments. This is because, at any history at which the DM prefers searching for new treatments to administering one of the α-treatments in the CS, she

[32]Formally, the ex-ante expected discounted number of times the α-treatments are administered increases with $p^{\alpha}(\emptyset)$ and v^{α}.

[33]Formally, the ex-ante expected discounted number of times the α-treatments are administered may strictly decrease with $p^{\alpha}(\emptyset)$ and v^{α}.

would have found it optimal to do the same also before the improvement. The arguments are similar to those for recruitment.

This conclusion, however, does not extend to settings with gradual resolution of uncertainty. In fact, in this case, the improvements in the α-treatments may induce the DM to search for new treatments at histories at which the DM would have administered one of the α-treatments in the CS that yielded negative outcomes in the past. The expansion of the CS, in turn, may favor the discovery of β-treatments and lead to an overall reduction in the administration of the α-treatments (measured by the ex-ante number of times such treatments are administered). We identify the conditions for such a phenomenon in the Appendix. Interestingly, the 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 effect 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 indexes of the two categories). In this case, the increase in the search index is small compared to the increase in the indexes of the α-treatments that delivered bad outcomes. The instances at which search crowds out the administration of the α-treatments are then rare. The direct effect of the improvement in the α-treatments then prevails over the indirect effect of the improvement in the attractiveness of search, and there is no crowding out. The phenomenon in the proposition thus occurs only for intermediate values of ρ^β. Similar non-monotonicities apply to the other parameters of the model.

Analogous phenomena naturally arise in many R&D environments, where improvements in certain categories of products spur investment in basic research which in turn may lead to the discovery of products from categories other than those that experienced the improvement, and with these other categories eventually crowding out the development and utilization of those products that saw the improvement in the first place. These dynamics are fairly natural and have been observed across a variety of industries. Nonetheless, as the proposition shows, they are not consistent with models featuring an exogenous CS or immediate resolution of uncertainty.

5 Conclusions and extensions

Summary. We introduce a model of experimentation in which the decision maker alternates between exploring alternatives in the consideration set and searching for new alternatives to explore in the future. Each search brings stochastically a new set of alternatives of different types that is added to the current consideration set. 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 comparison of independent indexes, where the index for search is computed in recursive form, accounting for future optimal decisions.

The analysis may be of interest to 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.

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

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. Prominent examples include the recruitment problem of Subsection 4.1 and the online consumer search problem of Subsection 4.2. As mentioned above, in general, such problems do not admit an index solution. In the Supplement, we derive a sufficient condition under which the optimality of an index policy extends to these 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 hiring committee must collect enough positive signals about a candidate's qualifications before hiring him; and a consumer must visit a vendor's webpage at least once to finalize a transaction with that vendor). 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 (as when a candidate is proved to be unqualified), or improves, weakly, with the number of explorations. 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.[34] Furthermore, because the time that each exploration takes can be arbitrary, by re-scaling 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 therefore also apply to problems in which search and learning occur "almost" in parallel.

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

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).

General "states" and multiple search options. The general model in Section 2 is quite flexible. In particular, the results in Theorem 1 and Proposition 1 apply also to problems in which the evolution of the "state" of each alternative is driven by shocks other than the arrival of information, as in the case of firms improving their products. In some of these problems, the decision maker may also face multiple search opportunities, that is, may choose "where and how" to expand the CS. The results in Theorem 1 and Proposition 1 extend to these settings provided that each "search arm" brings a set of alternatives independently from any other search arm.

References

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

Au, P. H., and M. Whitmeyer (2023). "Attraction versus persuasion: Information provision in search markets." *Journal of Political Economy,* 131(1), 202-245.

Athey, S. and G. Ellison (2011). "Position auctions with consumer search." *The Quarterly Journal of Economics* 126.3, 1213-1270.

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

Bardhi, A., Y. Guo, and B. Strulovici (2024). "Early-Career Discrimination: Spiraling or Self-Correcting?" mimeo New York University and Northwestern University.

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.

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.

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.

Edelman, B., Ostrovsky, M., and M. Schwarz (2007). "Internet advertising and the generalized second-price auction," *American Economic Review*, 97 (1), 242–259.

Edelman, B., and M. Schwarz (2010). "Optimal auction design and equilibrium selection in sponsored search auctions." *American Economic Review* 100(2), 597–602.

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

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

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

Gittins, J. 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.

Gomes, R. and K. Sweeney (2014). "Bayes–Nash equilibria of the generalized second-price auction," *Games and Economic Behavior* 86, 421–437.

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

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

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

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.

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 (2022). "Dynamically aggregating diverse information," *Econometrica*, 90(1), 47-90.

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

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

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

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

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.

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

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

Schuck, P. H. (2002). "Affirmative action: Past, present, and future." *Yale Law & Policy Review* 20 (1).

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.

6 Appendix

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

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

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

strictly greater than $\mathcal{I}^S(\omega^S)$, contradicting the optimality of $\hat{\tau}$ in the definition of $\mathcal{I}^S(\omega^S)$.[35] This implies that $\hat{\tau}$ is weakly greater than τ^*, where the latter is the first time at which the index of search and the index of each alternative brought to the CS by the search under consideration are weakly below $\mathcal{I}^S(\omega^S)$. Moreover, since at τ^* the index of search and of each alternative brought to the CS by the search under consideration are weakly below $\mathcal{I}^S(\omega^S)$, if $\hat{\tau} > \tau^*$, the average payoff per unit of average discounted time between τ^* and $\hat{\tau}$ must be equal to $\mathcal{I}^S(\omega^S)$. Hence, under the optimal selection rule in the definition of $\mathcal{I}^S(\omega^S)$, the average payoff per unit of average discounted time from 0 to τ^* must also be equal to $\mathcal{I}^S(\omega^S)$. This implies that the optimal stopping time in the definition of $\mathcal{I}^S(\omega^S)$ can be taken to be τ^*. Because the index policy χ^* selects in each period between 0 and τ^* the alternative for which the average payoff per unit of average discounted time is the largest (including search), we have that the optimal selection rule π in the definition of $\mathcal{I}^S(\omega^S)$ must coincide with the index policy χ^*. That $\mathcal{I}^S(\omega^S)$ satisfies the recursive representation in Part (2) then follows from the arguments above.

Part (3). We construct the following stochastic process based on the values of the indexes, and the expansion of the CS through search, *under the index policy* χ^*. Starting with the initial 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 $\omega_{t(v^0)}^S = \omega_0^S$.

[35]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).

We can proceed in this manner to obtain a strictly decreasing sequence of values $(v^i)_{i\geq 0}$, with corresponding stochastic times $(\kappa(v^i))_{i\geq 0}$. Note that the values v^i are all non-negative, as the DM's outside option is normalized to zero.

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

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

$$v^i = \frac{(1-\delta)\mathbb{E}^{\chi^*}\left[\eta^i | \mathcal{S}_{\kappa(v^i)}\right]}{\mathbb{E}^{\chi^*}\left[1 - \delta^{\kappa(v^{i+1})-\kappa(v^i)} | \mathcal{S}_{\kappa(v^i)}\right]}. \tag{13}$$

To see why (13) holds, recall that, at period $\kappa(v^i)$, given the state of the decision problem $\mathcal{S}_{\kappa(v^i)}$, the value of the highest index is v^i. Now suppose that the alternative corresponding to v^i is a physical alternative and that all other physical alternatives' indexes, as well as the index of search, are strictly below v^i. Recall that the optimal stopping time τ in the definition of the index of the physical alternative corresponding to v^i in (1) is the first period (strictly above $\kappa(v^i)$) at which the alternative's index falls below v^i. While it is convenient to take this fall to be weak, it is well known that one can equivalently take the fall to be strict. That is, stopping at the first period at which the index reaches a value equal to or smaller than the value at the time the index was computed is optimal, but so is stopping at the first period at which the index reaches a value strictly below the one at the time the index was computed. Now recall that $t(v^i)$ is the first time at which all indexes are strictly below v^i. Because the CS in period $\kappa(v^i)$ contains only one alternative with index equal to v^i (the physical one under consideration), $t(v^i)$ also coincides with the first period at which the index of the specific alternative under consideration drops strictly below v^i. Recall that v^{i+1} is the largest index at period $t(v^i)$ and that $t(v^i) = \kappa(v^{i+1})$. The definition of the index in (1), along with the optimality of stopping at the first time the index drops strictly below its initial value, and the definition of η^i, then imply that

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

which corresponds to the formula in (13).

Next, suppose that the alternative with the highest index at period $\kappa(v^i)$ is search, and that all physical alternatives in the CS in period $\kappa(v^i)$ have an index strictly smaller than v^i. As shown in the proof of Part (1) of Theorem 1 above, the optimal stopping time in the definition of the index of search in (2) is the first period (strictly above $\kappa(v^i)$) at which the index of search and of all the alternatives introduced through search, fall weakly below v^i. Equivalently, as discussed above, the optimal stopping time can also be taken to be the first period at which the index of search and of all the alternatives introduced through search fall strictly below v^i. Because all physical alternatives in the CS at period $\kappa(v^i)$ have an index strictly below v^i, such a period coincides with $t(v^i)$, that is, with the first period at which the index of search and of all alternatives in the CS are strictly below v^i. Using the above property of the optimal stopping time in the definition of the search index in (2), along with the fact that $t(v^i) = \kappa(v^{i+1})$ and the definition of η^i, we then have that the search index evaluated at period $\kappa(v^i)$ also satisfies the condition in (13).

Finally, suppose that, at period $\kappa(v^i)$, there are multiple options ("physical" alternatives and/or search) with index v^i. Then observe that the average sum $\mathbb{E}^{\chi^*}\left[\sum_{s=\kappa(v^i)}^{\kappa(v^{i+1})-1} \delta^{s-\kappa(v^i)} U_s | \mathcal{S}_{\kappa(v^i)}\right]$ of the discounted net payoffs from utilizing all options whose period-$\kappa(v^i)$ index is equal to v^i till the first period $t(v^i) = \kappa(v^{i+1})$ at which the indexes of all options are strictly below v^i, normalized by the average per unit discounted time $\mathbb{E}^{\chi^*}\left[\sum_{s=\kappa(v^i)}^{\kappa(v^{i+1})-1} \delta^{s-\kappa(v^i)} | \mathcal{S}_{\kappa(v^i)}\right]$ is the same as the average sum $\mathbb{E}^{\chi^*}\left[\sum_{s=\kappa(v^i)}^{T-1} \delta^{s-\kappa(v^i)} U_s | \mathcal{S}_{\kappa(v^i)}\right]$ of the discounted net payoffs from utilizing each individual option with index (at period $\kappa(v^i)$) equal to v^i till the first time T at which that option's index (and, in case the option is search, also the indexes of all alternatives brought to the CS by the search initiated at $\kappa(v^i)$) fall strictly below v^i, normalized by the average discounted time $\mathbb{E}^{\chi^*}\left[\sum_{s=\kappa(v^i)}^{T-1} \delta^{s-\kappa(v^i)} | \mathcal{S}_{\kappa(v^i)}\right]$. This follows from the independence of the processes. Hence, Condition (13) also holds when, at $\kappa(v^i)$, there are multiple options with index v^i.

Multiplying both sides of (13) by $\delta^{\kappa(v^i)}$, rearranging terms, and using the fact that $\delta^{\kappa(v^i)}$ is known at $\kappa(v^i)$, we have that

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

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

$$\mathcal{V}(\mathcal{S}_0) = \mathbb{E}^{\chi^*}\left[\sum_{i=0}^{\infty} v^i\left(\delta^{\kappa(v^i)} - \delta^{\kappa(v^{i+1})}\right) | \mathcal{S}_0\right]. \tag{14}$$

Next, note that $\delta^{\kappa(v^i)} = 0$ whenever $\kappa(v^i) = \infty$, and that, for any $i = 0, 1, ..., \kappa(v) = \kappa(v^{i+1})$

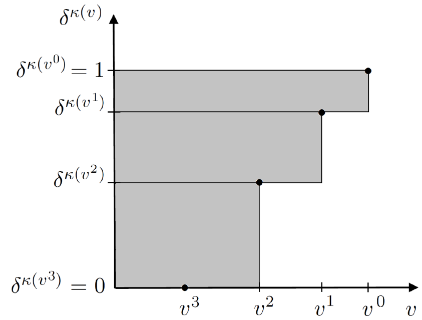

Figure 1: An illustration of the function $\delta^{\kappa(v)}$ and the region $\sum_{i=0}^{\infty} v^i \left(\delta^{\kappa(v^i)} - \delta^{\kappa(v^{i+1})}\right) = \int_0^\infty v \mathrm{d}\delta^{\kappa(v)}$, for a particular path with $\kappa(v^3) = \infty$.

for all $v^{i+1} < v < v^i$. It follows that (14) is equivalent to

$$\mathcal{V}(\mathcal{S}_0) = \mathbb{E}^{\chi^*}\left[\int_0^\infty v \mathrm{d}\delta^{\kappa(v)} | \mathcal{S}_0\right] = \mathbb{E}^{\chi^*}\left[\int_0^\infty \left(1 - \delta^{\kappa(v)}\right) \mathrm{d}v | \mathcal{S}_0\right] = \int_0^\infty \left(1 - \mathbb{E}^{\chi^*}\left[\delta^{\kappa(v)} | \mathcal{S}_0\right]\right) \mathrm{d}v. \quad (15)$$

The construction of the integral function (15) is illustrated in Figure 1.

Proof of Proposition 2–Part (1). Let $\hat{\mathcal{I}}^P(\alpha, \theta)$ and $\hat{\mathcal{I}}^S(\omega^S)$ denote the physical and search indices under the reduced threshold $\hat{\Psi}^\alpha$ and $\mathcal{I}^P(\alpha, \theta)$ and $\mathcal{I}^S(\omega^S)$ their counterparts under the original threshold Ψ^α. We first prove the following property.

Lemma 1. *Suppose the evaluation of the candidates satisfies the property of immediate resolution of uncertainty. For any $\omega^S \in \Omega^S$, if $\hat{\mathcal{I}}^S(\omega^S) > \hat{\mathcal{I}}^P(\alpha, \emptyset)$, then $\mathcal{I}^S(\omega^S) > \mathcal{I}^P(\alpha, \emptyset)$.*

Proof of Lemma 1. We introduce the notation:

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

so that $\mathcal{I}^S(\omega^S) = \sup_{\pi,\tau} \mathcal{I}^S(\omega^S; \pi, \tau)$. That is, $\mathcal{I}^S(\omega^S; \pi, \tau)$ denotes the value of the objective being maximized in the definition of the search index.

That $\Psi^\alpha > \hat{\Psi}^\alpha$ implies that $\mathcal{I}^P(\alpha, \emptyset) \le \hat{\mathcal{I}}^P(\alpha, \emptyset)$. Suppose that $\hat{\mathcal{I}}^S(\omega^S) > \hat{\mathcal{I}}^P(\alpha, \emptyset)$. Let $(\hat{\pi}^*, \hat{\tau}^*)$ be the optimal rules in the maximization in the definition of $\hat{\mathcal{I}}^S(\omega^S)$, i.e., $\hat{\mathcal{I}}^S(\omega^S) = \hat{\mathcal{I}}^S(\omega^S; \hat{\pi}^*, \hat{\tau}^*)$. Then, by Part (2) of Theorem 1, none of the $\alpha-$candidates identified through search (starting from ω^S) is inspected under $\hat{\pi}^*$. This is because $\hat{\pi}^*$ prescribes to stop the first time at which all of the indices of newly found candidates and of search itself are weakly below $\hat{\mathcal{I}}^S(\omega^S)$, and $\hat{\mathcal{I}}^P(\alpha, \emptyset)$ is strictly below $\hat{\mathcal{I}}^S(\omega^S)$. Note that this holds for any search technology.

Now suppose it were the case that $\mathcal{I}^P(\alpha, \emptyset) \ge \mathcal{I}^S(\omega^S)$, so that $\mathcal{I}^S(\omega^S) \le \mathcal{I}^P(\alpha, \emptyset) \le \hat{\mathcal{I}}^P(\alpha, \emptyset) < \hat{\mathcal{I}}^S(\omega^S)$. Because $(\hat{\pi}^*, \hat{\tau}^*)$ prescribes not to evaluate any of the α-candidates after bringing them to the CS, the payoff (per unit of discounted time) under $(\hat{\pi}^*, \hat{\tau}^*)$ is the same no matter whether the acceptance threshold for the α-candidates is Ψ^α or $\hat{\Psi}^\alpha$, implying that $\mathcal{I}^S(\omega^S;\hat{\pi}^*,\hat{\tau}^*) = \hat{\mathcal{I}}^S(\omega^S;\hat{\pi}^*,\hat{\tau}^*)$. Because $\mathcal{I}^S(\omega^S) = \sup_{\pi,\tau}\mathcal{I}^S(\omega^S;\pi,\tau)$, it must be that

$$\mathcal{I}^S(\omega^S) \geq \mathcal{I}^S(\omega^S;\hat{\pi}^*,\hat{\tau}^*) = \hat{\mathcal{I}}^S(\omega^S),$$

which contradicts the assumption that $\mathcal{I}^S(\omega^S) < \hat{\mathcal{I}}^S(\omega^S)$. Hence the result. □

Next note that $\hat{\mathcal{I}}^P(\alpha,\emptyset) \geq \mathcal{I}^P(\alpha,\emptyset)$ implies that, for any $\omega^S \in \Omega^S$, $\hat{\mathcal{I}}^S(\omega^S) \geq \mathcal{I}^S(\omega^S)$. In turn, this means that, after any history of past evaluations and searches, if the committee evaluates a β-candidate under the new acceptance thresholds, it also evaluates a β-candidate under the original thresholds. Furthermore, Lemma 1 establishes that, whenever, under the new thresholds, the committee expands the CS before evaluating an α-candidate, it would have done the same under the original thresholds. The result in the proposition then follows from these properties along with the following properties: (a) the set of signal realizations for which α-candidates, after being evaluated, can be hired is larger under the new thresholds; (b) the index for evaluating the α-candidates is larger under the new thresholds, that is, $\hat{\mathcal{I}}^P(\alpha,\emptyset) \geq \mathcal{I}^P(\alpha,\emptyset)$; (c) the search technology is invariant to changes in the acceptance threshold; (d) the probability that, under the index policy, an α-candidate is hired increases whenever, at a given history, the committee evaluates an α-candidate instead of evaluating a β-candidate or searching, and when the committee searches instead of evaluating a β-candidate (it decreases, whenever the committee evaluates a β-candidate instead of evaluating an α-candidate or searching).

Note that property (d) above is a direct implication of the optimal policy being an index policy, along with the evaluation of candidates involving immediate resolution of uncertainty: a candidate who, after being evaluated, is below the acceptance threshold is never reconsidered, implying that how far the posterior of such a candidate is relative to the acceptance threshold has no effect on the sequence of future decisions. ■

The proofs for Parts 2 and 3 of the Proposition are in the Supplement.

Proof of Proposition 3. The result follows from the properties preceding the proposition, namely that $\mathcal{I}^S(m)$ declines with m and that, for any m, $\mathcal{I}^S(m) \leq \mathcal{I}^P(\bar{\xi},\emptyset)$. These properties imply that, when m is small, after bringing the m-th ad to the CS and finding that it is of type $\xi < \bar{\xi}$, the consumer may prefer not to click on the ad (if $\mathcal{I}^S(m+1) > \mathcal{I}^P(\xi,\emptyset)$) and instead read the next ad. Later in the process, though, say after bringing the m'-th product to the CS, with $m' > m$ such that $\mathcal{I}^S(m'+1) < \mathcal{I}^P(\xi,\emptyset)$, the consumer may start clicking on some of the ads encountered at positions m'', with $m < m'' < m'$. These ads are clicked in the order of their indexes $\mathcal{I}^P(\xi,\emptyset)$. Because, with positive probability, $\xi(n) < \xi(n+j)$ for $m < n < n+j < m'$, the clicking over the ads displayed between positions m and m' can be non-sequential and non-cascading.

Proof of Proposition 4. Consider any state of the world in which the two firms' attractiveness differs, that is, $\xi_1 \neq \xi_2$, and, without loss of generality, assume that firm 1's product is the most attractive one. As explained in the main text, when the index for clicking $\mathcal{I}^P(\xi_2, \emptyset)$ of the least attractive firm is smaller than the index for reading the second ad, each firm drops out instantaneously because it expects the consumer to click on the ad of the most attractive firm first, irrespective of the position at which such an ad is displayed. That the two firms drop out at the same price then follows from the analysis in the main text after the proposition by observing that, given any assignment $(\xi(1), \xi(2))$, $P(2; \xi(1), \xi(2)) = 1 - P(1; \xi(1), \xi(2))$, which in turn is due to the assumption that the payoff the consumer expects from discovering her value for each firm's product exceeds her outside option. That the firms' bidding strategies are symmetric in states in which their attractiveness coincides is obvious (and hence the proof is omitted). ■

Proof of Proposition 5. The proof is in two steps. Step 1 establishes that the indexes take the form in (11) and (12) in the main text, whereas Step 2 establishes the properties in the proposition.

Step 1. Recall that the optimal stopping time in the index definition of an arm 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 (11) then uses the fact that a good outcome perfectly reveals that the treatment is effective, in which case $\tau^* = \infty$ in the definition of the arm's index, 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 (2) of Theorem 1, together with Part 6 of Proposition 1, we have that the index for search is invariant to ω^S and equal to (12). To see this, recall that the search technology is stationary (and, hence, weakly deteriorating) in this problem. Part 6 of Proposition 1 then implies that the index is the same as in a fictitious environment with a single opportunity to search. Part (2) of Theorem 1 in turn implies that the optimal stopping time in the definition of the search index is the first time at which the posterior belief about the efficacy of the new treatment brought to the CS is such that the treatment's index drops below the index of search when the latter was launched.

Step 2. Consider Part (a) first. Note that the indexes of the α-treatments (both the index

$$\mathcal{I}^P(\alpha, \emptyset) = \frac{(1 - \delta + \delta q_1^\alpha)\, p^\alpha(\emptyset) q_1^\alpha v^\alpha}{1 - \delta + \delta p^\alpha(\emptyset) q_1^\alpha}$$

for those treatments that have not been administered yet, and $p^\alpha(\theta) v^\alpha (1 - \delta)$ for those that have been administered already) are increasing in both v^α and $p^\alpha(\emptyset)$. Also note that, under immediate resolution of uncertainty (i.e., when $q_1^\alpha = 1$, or more generally, whenever any administration following the first one does not provide any further information), if the DM administers a treatment that has been administered already in the past, she then does so in all subsequent periods. The problem under consideration is thus similar to the one considered in the recruitment application, where the assignment of a job to a candidate that has been evaluated already brings to an end the exploration process. The result thus follows from arguments similar to those in the proof of Part (1) in Proposition 2.

Next, consider Part (b). We first identify conditions under which the exploration and CS expansion 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 strictly higher before the improvement. The conditions below are sufficient but not necessary for the result. However, they make it 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 the α-treatments improve. Such an improvement may take the form of (i) 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 (ii) 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. Note that 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 otherwise depends on $\theta \neq \emptyset$ only through the number s of bad outcomes recorded in θ. Hence, 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 $\mathcal{I}^P(\xi, \theta) = q_1^\xi v^\xi$. 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

$$
\begin{aligned}
&\mathcal{I}^P(\alpha, \emptyset) > \mathcal{I}^P(\beta, \emptyset) > \mathcal{I}^P(\alpha, 1) \\
&> \frac{-c(1-\delta) + \delta \left\{ \rho^\alpha \Lambda^\alpha(\emptyset) \left[1 - \delta^2 (1 - q_1^\alpha)^2\right] v^\alpha + \rho^\beta \Lambda^\beta(\emptyset) \left[1 - \delta(1 - q_1^\beta)\right] v^\beta \right\}}{1 - \delta^2 \left\{ \rho^\alpha \delta \left[1 - 2\Lambda^\alpha(\emptyset) + q_1^\alpha \Lambda^\alpha(\emptyset)\right] + \rho^\beta \left[1 - \Lambda^\beta(\emptyset)\right] \right\}} \\
&> \max\{\mathcal{I}^P(\beta, 1), \mathcal{I}^P(\alpha, 2)\}.
\end{aligned} \tag{16}
$$

We then argue that the formula for the index of search simplifies to

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

The result follows from Part (2) of Theorem 1, along with Part (6) in Proposition 1. In particular, the order in (16) implies that the optimal stopping time in the formula for the search index 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.[36]

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

$$\hat{\mathcal{I}}^P(\alpha,\emptyset) > \mathcal{I}^P(\beta,\emptyset) > \frac{-c(1-\delta)+\delta\left\{\rho^{\alpha}\hat{\Lambda}^{\alpha}(\emptyset)\left[1-\delta(1-q_1^{\alpha})\right]\hat{v}^{\alpha}+\rho^{\beta}\Lambda^{\beta}(\emptyset)\left[1-\delta(1-q_1^{\beta})\right]v^{\beta}\right\}}{1-\delta^2\left[1-\rho^{\alpha}\hat{\Lambda}^{\alpha}(\emptyset)-\rho^{\beta}\Lambda^{\beta}(\emptyset)\right]}$$
$$> \max\{\hat{\mathcal{I}}^P(\alpha,1),\mathcal{I}^P(\beta,1)\}, \tag{17}$$

where the hat on the indexes $\hat{\mathcal{I}}$ indicates that they are computed after the improvement. 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 (2) of Theorem 1 along with Part (6) in Proposition 1. Given (17), 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 (16) 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 second administration 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 Part (5) in Proposition 1, 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

[36] Recall that search itself occupies one period.

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

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

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 (17), 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.[37] 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 improve is equal to

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

It is easily verified that Conditions (16) and (17) are consistent with $A > \hat{A}$ over an open set of parameter values such that $\varepsilon_p \geq 0$ and $\varepsilon_v \geq 0$, with at least one inequality strict. ■

[37] 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.