Sequential Learning with Endogenous Consideration Sets

Daniel Fershtman†        Alessandro Pavan‡

June 2020

Abstract

We introduce a model of sequential learning where the decision maker (DM) can expand her consideration set by searching for new alternatives to explore. The DM faces uncertainty about prospective alternatives outside of the consideration set. We characterize the optimal policy and relate the decision dynamics to the properties of the search technology. We show that the decision to search depends on the composition of the consideration set only through the information that the latter contains about the probability of finding new alternatives. When the search technology is stationary, or improves over time, search is equivalent to replacement. With deteriorating technologies, instead, alternatives in the consideration set are revisited after search is launched and the DM proceeds as if each expansion were the last one. The analysis also yields a formula for the DM’s payoff under the optimal policy that can be used to price the access to new alternatives, as well as the option to expand the consideration set in the future. Finally, we show that the analysis can accommodate for certain irreversible choices that admit as a special case a generalization of Weitzman’s (1979) “Pandora’s boxes” problem in which the set of boxes is endogenous and each alternative may need to be explored multiple times before it can be selected.

*The paper supersedes a previous version circulated under the title “Searching for Arms”. For useful comments and suggestions we thank Dirk Bergemann, Eddie Dekel, Laura Doval, Kfir Eliaz, Stephan Lauermann, Benny Moldovanu, Eran Shmaya, Philip Reny, Rani Spiegler, Asher Wolinsky, and seminar participants at University of Bonn, Cornell University, Hebrew University of Jerusalem, HEC Paris, Paris Dauphine, Tel Aviv University, Toulouse School of Economics, Yale University, Warwick Economic Theory Workshop, 3rd Columbia Conference on Economic Theory (CCET), and 30th Stony Brook International Conference on Game Theory. We also thank Matteo Camboni for excellent research assistance. The usual disclaimer applies.

†Eitan Berglas School of Economics, Tel Aviv University. Email: danielfer@tauex.tau.ac.il
‡Department of Economics, Northwestern University. Email: alepavan@northwestern.edu
1 Introduction

An important implication of limited attention is that a decision maker (hereafter, DM) can consider only a subset of the alternatives available to her. This scarcity gives rise to the concept of “consideration sets,” which has recently been applied to a range of areas of study.\(^1\)

Classic models of sequential learning often involve a DM exploring a fixed set of options with unknown returns. Yet, a ubiquitous feature of dynamic decision problems is that the set of options in a DM’s consideration set is not predetermined. Rather, the DM may choose to expand her consideration set by seeking out additional options as part of her learning process, in response to information she has gathered. This paper studies the tradeoff between learning about alternatives within the DM’s consideration set and expansion of the latter through search for additional options.

Consider for example a consumer’s online search for a product. Time is costly, and the consumer faces a potentially huge amount of options, displayed sequentially (and stochastically, from the perspective of the consumer) across multiple pages. The information displayed in each ad is limited. By clicking on any of the ads, the consumer is directed to the vendor’s webpage from which she can gather further information and, when desirable, finalize a purchase. Rather than clicking on one of the ads on the existing page (or on one of the pages previously visited), the consumer can also move on to the next page of results, or switch to a different website. The decision to do so entails some time and/or cognitive cost, and depends on the information the consumer has gathered about the products thus far, as well as the relevance and number of suitable new options she expects to find by expanding her consideration set. Assuming the consumer behaves “optimally,” under what conditions does she seek additional new options to explore? If she does move beyond the first page of results, or switches to a different website, is she likely to go back to results she has already explored? How do such decisions depend on the gradual resolution of uncertainty and on the properties of the search technology (e.g., on how the search engine distribute ads on multiple consecutive pages). These questions are relevant for both advertisers on a platform and for the platform’s design of its search environment (e.g., how many ads to display on each page).\(^2\)

Similarly, parents choosing a prospective school district may trade-off further evaluations of schools they are already aware of (or that are in their vicinity) with the expansion of their consideration set by searching for new options in locations they may not know of ex-ante. R&D often involves pursuing a number of alternative technologies whose ability to produce the desirable

---

\(^1\)The concept of consideration sets also has a long tradition in the marketing literature.

\(^2\)Consumers typically limit their attention to a relatively small number of websites when buying products over the internet (De los Santos, Hortaçu, and Wildenbeest, 2012). According to Epstein and Robertson (2015), a “recent analysis of \(\sim\)300 million clicks on one search engine found that 91.5% of those clicks were on the first page of search results” although not necessarily on the order by which the ads were displayed on the same page (see, e.g., Jezierski and Segal (2015)). Millward Brown examined how likely customers are to view products beyond the first page of search results on Amazon, and found that approximately 70% did not click past the first page of results (see clavisinsight.com 2015).
breakthrough is unknown, but also spending time and/or effort to search for new alternatives to subsequently explore. The tradeoff between the two activities (exploration and expansion of the consideration set) evolves over time based on the development of existing projects and the ability to find new alternatives.

To study this tradeoff, we introduce a model of sequential learning among an endogenous set of alternatives. In each period, the DM can either focus attention on a single alternative within her current consideration set or choose to expand it. We refer to such expansion as “search” for new alternatives, as we assume it is costly and its outcome may be stochastic. Focusing attention on an alternative generates a signal that is informative about the alternative’s value (and independent of other alternatives) and may yield a payoff (positive or negative). The decision to expand the consideration set (i.e., to search) triggers the discovery of a stochastic set of new alternatives as a function of the state of the search technology. This state may evolve over time based on the past outcomes of search and on the information the DM possesses about the process governing the search results. For example, the state of the search technology may be stationary, yielding i.i.d. sets of new options from a known distribution. Alternatively, it may evolve over time in a non-stationary manner reflecting the DM’s beliefs about the set of alternatives outside her consideration set.

Our environment and characterization of the DM’s optimal policy extend Weitzman’s (1979) classic problem, and its solution based on independent “reservation prices,” to an environment in which the consideration set can be sequentially expanded and is endogenous to the DM’s problem. Alternatives are assigned independent reservation prices, which are used to determine the order in which attention is allocated among them. These reservation prices are equivalent to those which determine the optimal policy in the absence of the option to expand the consideration set. Expansion of the consideration set is also assigned a reservation price, which depends on the information the DM has about the “state” of the existing search technology, but also on the exploration and future expansion policy the DM plans to follow once she starts the search. At any given time, the DM’s decision under the optimal policy is determined by the action – focusing on one of the alternatives within the consideration set or expanding the latter – whose reservation price is the greatest.

Rather than committing to a consideration set up front and then proceeding to evaluate the alternatives in the optimal order, the DM chooses when to expand the consideration set based on the results of her past explorations and past search (i.e., expansion) outcomes. We show that, despite the fact that the size and composition of the consideration set may affect the DM’s beliefs about the results of future searches and in particular the likelihood of finding alternatives similar to those already in the consideration set, in expectation, the relative attention the DM allocates to any pair of alternatives within the consideration set does not vary with the results of past search outcomes.

Importantly, the reservation price the DM assigns to the option to expand the consideration set does not coincide with the value of the expansion. That is, the decision to expand is typically
sensitive to only partial information about the prospective alternatives the DM expects to add to the consideration set, even if such information is relevant for the expected continuation payoff under the optimal policy. The reservation price corresponding to the expansion of the consideration set is intricately linked to the reservation prices of the new alternatives expected to be introduced in the future. We show that if the search technology is stationary (or “improving” in an appropriate sense made precise below) then alternatives in the set at the time of the expansion never receive attention in the future, and hence are effectively discarded once the set is expanded. That is, search is equivalent to replacement. When, instead, the search technology “deteriorates” over time, alternatives are put on hold and returned to at a later stage, after the consideration set has been expanded. Furthermore, in this case, the DM proceeds as if each decision to expand were the last one (that is, the reservation price of expansion takes into account only information about alternatives that are expected to be added to the consideration set as the result of the current search, as if further expansions were not feasible).

More generally, we show that the decision to expand the consideration set depends on the composition of the set only through the information that the latter contains about the probability of finding new alternatives of different types. This result holds despite the fact that the alternatives in the set may share similarities with those that are expected to be found through the expansion and despite the fact that the opportunity cost of searching for new alternatives (which is linked to the value of continuing to explore the current set) depends on the entire composition of the consideration set. Likewise, at any point in time, the relative attention allocated to any pair of alternatives in the consideration set is invariant to the search technology, and in particular is independent of the probability that search will bring new alternatives similar to those already under consideration. Finally, improvements in the search technology yielding an increase in the probability that search brings alternatives of positive expected value (vis-a-vis the outside option) need not affect the decision to search, even at histories at which, prior to the improvement, the DM is indifferent between searching and continuing with the current consideration set.

The results are not specific to sequential learning. They apply to a broad class of dynamic experimentation problems with an endogenous set of alternatives. We start by considering the case in which the DM can revert her decision at all periods, as in the multi-armed bandit literature. We prove that the optimal policy takes the form of an index rule with a special index for search. The result is related to the optimality of index policies in the branching literature (e.g., Weiss, 1988, Weber, 1992, and Keller and Oldale, 2003). Our contribution is not in expanding the results in this operation research literature but in identifying conditions under which indexability applies to the sequential learning problem with endogenous consideration sets under examination. However, our proof of indexability is new to the literature and uses a recursive representation of the index of search which also yields a novel representation of the DM’s payoff under the optimal policy (with or without the possibility of expansion). The representation can be used to price the access to new alternatives as well as the option to expand the set in the future.

Next, we consider the case where, in addition to focusing attention on existing alternatives
and searching for new ones, the DM can also irreversibly commit to one of the alternatives in the consideration set, putting an end to the exploration. In general, the irreversibility of choice is known to preclude a tractable solution. We identify a condition under which the optimal policy remains indexable, which admits as a special case a generalization of Weitzman’s (1979) original problem to a setting in which (a) the consideration set (i.e., the set of boxes) is endogenous, (b) learning the value of an alternative in the consideration set may require multiple explorations, and (c) the DM may derive a positive payoff from exploring an alternative (possibly higher than the value she derives from irreversibly committing to it) for an arbitrary large (and possibly infinite) number of periods.\footnote{The condition is based on a certain “better-later-than-sooner” property guaranteeing that once an alternative reaches a state in which the DM can irreversibly commit to it, its “retirement value” (that is, the value of irreversibly committing to it) either drops below the value of the outside option, or improves over time. This “better-later-than-sooner” 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 approach is, however, quite different and accommodates for the possibility that the set of alternatives is endogenous.}

In Section 2, we kick off by considering the simplest extension of Weitzman’s (1979) model to a setting with endogenous consideration sets. As in Weitzman’s model, we assume that it takes a single exploration to learn an alternative’s value and that exploring an alternative without committing to it comes at a cost. We characterize the “prizes” that guide the DM’s sequential alternation between exploring the alternatives already in the consideration set and the expansion of the latter. We also derive an “eventual-purchase theorem” in the spirit of Choi, Dai and Kim (2018) (see also Armstrong and Vickers, 2015 and Armstrong, 2017) that relates the probability that each alternative is eventually selected to the primitives of the search problem (realized values and search technology) and discuss how the endogeneity of the consideration set affects the selection probabilities.

The possibility to expand the consideration set may change quite radically some of the comparative statics of the canonical model. To illustrate this possibility in concrete terms, we consider a stylized market in which three firms advertise on a platform. Each firm has two different products, and a representative consumer seeks to purchase at most one of the firms’ products. As in Weitzman (1979), the consumer must inspect a product to finalize the purchase. We consider two cases: one with a fixed consideration set, and one with an endogenous set. In the fixed-consideration-set case, there are four advertising slots, with equal visibility. Each firm is endowed with one slot, but one of the firms, chosen by the platform before the consumer’s search begins is given a second slot to advertise her other product. As in Weitzman (1979), the consumer sees all the four ads (i.e., the identity of the firm that was awarded the second slot) before starting the exploration. She then sequentially decides between inspecting a product (by clicking on a firm’s ad) and stopping and then either choosing an inspected product or her outside option (this version of the problem is thus identical to the one in Weitzman (1979)). In the second environment, instead, ads are displayed on two different pages. Each firm advertises on the first page, and one of the firms also advertises on the second page. The identity of such firm is unknown.
to the consumer who may have correct beliefs about the probability the slot is assigned to each firm but nonetheless does not know the realization of the relevant risk. This environment thus corresponds to a sequential learning problem with an endogenous consideration set. The consumer may expand her consideration set by visiting the second page but this entails a cost (the magnitude of which may be small and possibly due only to the postponing of the exploration of one of the alternatives already in the consideration set). The consumer’s optimal inspection strategy is an index policy with a special index for the decision to explore the second page. When the consideration set is exogenous (and contains all items), each firm benefits from an increase in the probability she receives the additional slot that permits it to display its second product. This is not the case when the consideration set is endogenous: a firm may be strictly worse off when the probability the slot on the second page is assigned to it increases.

The rest of the paper is organized as follows. The remainder of this section discusses the related literature. Section 2 considers the extension of Weitzman’s model to a setting with an endogenous set of boxes and contains the example discussed above. Section 3 presents the general model, contains all the main results, and discusses how the optimal policy in Weitzman (1979) must be amended to accommodate not only for endogenous consideration sets but also for gradual resolution of uncertainty. Section 4 discusses a several simple extensions, whereas Section 5 concludes. Most of the proofs are relegated to the Appendix.

1.1 Relation to the literature

The paper is part of a growing literature studying optimal sequential learning in settings in which the DM explores one alternative at the time. Most closely related are Ke, Shen, and Villas-Boas (2016), Austen-Smith and Martinelli (2018), Ke and Villas-Boas (2019), and Gossner, Steiner and Stewart (2019). Each of these papers considers a different cost and information structure and makes different assumptions on when learning stops. In contrast to our work, in all of these papers the set of alternatives the DM can focus her attention on is fixed ex-ante. In related work, Che and Mierendorff (2019) study a DM’s optimal sequential allocation of attention to two different signal sources biased towards alternative actions. In our model, instead, the DM learns about the value of different alternatives (drawn independently) while also expanding the set of alternatives to explore over time.

The paper is also related to the literature on consideration sets. A number of papers in the marketing literature study the formation of consideration sets (e.g., Hauser and Wernerfelt, 1990; Roberts and Lattin, 1991). Eliaz and Spiegler (2011) study implications of different consideration sets on firms’ behavior, assuming such sets are exogenous. Masatlioglu, Nakajima, and Ozbay (2012) and Manzini and Mariotti (2014), instead, identify consideration sets from choice behavior. Caplin, Dean, and Leahy (2018) provide necessary and sufficient conditions for rationally inattentive agents to focus on a subset of all available choices, thus endogenizing the consideration sets. Simon (1955) considers a sequential search model, in which alternatives are examined till a
“satisfying” alternative is found.\textsuperscript{4} Our analysis complements the one in this literature by providing a dynamic microfoundation for endogenous consideration sets. Rather than committing to a consideration set up front and proceeding to evaluate the alternatives in it, the DM expands the consideration set over time, in response to the results obtained from the exploration of the alternatives in the set.

As mentioned above, our results cover as special case an extension of Weitzman’s (1979) classic problem in which the set of boxes is endogenous and where exploration leads to a gradual resolution of uncertainty. 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 contrast to our work, in all of these papers, the DM faces a fixed set of alternatives. In independent work, Greminger (2020) considers a version of Weitzman’s (1979) problem in which the DM can bring to the consideration set new options over time. His problem is a special case of ours, both in terms of the search technology and of the DM’s payoff.

Our paper is also related to Garfagnini and Strulovici (2016), who study how successive (forward-looking) agents experiment with endogenous technologies. Trying a “radically” new technology reduces the cost of experimenting with similar technologies, which effectively expands the space of affordable technologies.\textsuperscript{5} While both their work and ours consider environments in which the set of alternatives/technologies is expanded over time, the two models, as well as the analysis and questions addressed, are fundamentally different. Schneider and Wolf (2019) study the time-risk tradeoff of an agent who wishes to solve a problem before a given deadline, and allocates her time between implementing a given method and developing (and then implementing) a new one. Related is also Fershtman and Pavan (2020), which considers the effects of “soft” affirmative action on minority recruitment in a setting in which the candidate pool is endogenous.\textsuperscript{6}

The paper is related to the multi-armed bandit literature pioneered by Gittins and Jones (1974), and in particular to the part of this literature that looks at “branching bandits,” where “arms” in addition to yielding payoffs branch into new arms. Versions of such problems are studied in Weiss (1988), Weber (1992), and Keller and Oldale (2003), among others. As anticipated above, our contribution is to show that, under appropriate conditions, the problem of sequential learning with endogenous consideration sets is a special case of the branching problem in which all arms but one do not branch, and where the “expansion” arm does not yield direct benefits but brings additional arms. We provide an independent proof of the optimality of an index policy for this problem which hinges on a recursive characterization of the index for the search arm and yields a novel representation of the DM’s expected payoff under an index policy.\textsuperscript{7}

\textsuperscript{4}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.

\textsuperscript{5}Technologies are also interdependent in their environment. In particular, a radically new technology is informative about the value of similar technologies.

\textsuperscript{6}The problem in that paper is a special version of the one considered in the present paper.

\textsuperscript{7}The reason why indexability is not obvious and does not follow from standard results in the bandit literature is that search is a “meta-arm” bringing alternatives with correlated returns that one needs to process optimally.
2 Pandora’s problem with an endogenous set of boxes

Consider the following extension of Weitzman’s (1979) “Pandora’s problem” in which a DM optimally constructs her consideration set over time, accounting for the costs of expanding it.

A DM must make a single choice among alternatives. An alternative is characterized by a pair \((F, \lambda)\), where \(F\) denotes the distribution over the alternative’s unknown value \(u\), and \(\lambda\) denotes the cost of inspecting the alternative to learn \(u\) (here, as in Weitzman’s setting, we consider the simple case where the value of an alternative is revealed immediately upon its first inspection; this assumption is relaxed in the general model in Section 3). Initially, the DM is aware of only a subset of alternatives – this is her initial consideration set – and has an outside option, normalized to zero.

The DM’s consideration set is endogenous to her decision problem, and adding alternatives into the consideration set is increasingly costly. More precisely, at each period \(t = 0, 1, \ldots\), the DM can either search for an additional alternative to add to her consideration set, inspect an alternative to learn its value, or stop and either recall an observed payoff \(u\) from one of the inspected alternatives or take her outside option (in which case the decision problem ends). Search expands the consideration set, introducing a new alternative for which \((F, \lambda)\) are drawn independently from a set \(A\) with known distribution \(F\) (we consider here the simple case where search yields a single new alternative; in the general model in Section 3, search may introduce a stochastic number of alternatives). The cost of adding an alternative to the consideration set is equal to \(c(m)\), where \(m\) is the number of past searches (i.e., expansions of the consideration set), and \(C\) is a positive, increasing function. Besides the direct costs of inspecting and adding new alternatives, the DM discounts the future according to \(\delta\).

How should the DM optimally (and dynamically) balance her tradeoff between bringing in alternatives to inspect and inspecting her current options? The following Proposition characterizes the DM’s optimal behavior.

**Proposition 1.** Denote by \(I(F, \lambda)\) the “reservation price” (equivalently, the index) of an alter-
native, as in Weitzman’s problem; that is,

$$\mathcal{I}(F, \lambda) = \frac{-\lambda + \delta \int_{\mathcal{I}(F, \lambda)}^{\infty} u dF(u)}{1 + \frac{\delta}{1-\delta} \Pr(u > \mathcal{I}(F, \lambda)|F, \lambda)}.$$

(1)

For any \( l \in \mathbb{R} \), denote \( \mathcal{A}(l) \equiv \{(F, \lambda) \in \mathcal{A} : \mathcal{I}(F, \lambda) > l\} \), and let \( \mathcal{I}^S(m) \) be defined by

$$\mathcal{I}^S(m) = -c(m) + \frac{\delta \int_{\mathcal{A}(\mathcal{I}^S(m))} \left( -\lambda + \delta \int_{\mathcal{I}^S(m)}^{\infty} u dF(u) \right) dF(F, \lambda)}{1 + \int_{\mathcal{A}(\mathcal{I}^S(m))} \left( \delta + \frac{\delta^2}{1-\delta} dF(u) \right) dF(F, \lambda)}.$$

(2)

The DM’s optimal policy is the following: (i) Add an alternative to the consideration set if \( \mathcal{I}^S(m) \) is positive and greater than the index of all uninspected options in the consideration set and all observed values of inspected options. (ii) Inspect an alternative in the consideration set if its index is positive, is the greatest in the consideration set, and exceeds \( \mathcal{I}^S(m) \) as well as the value of all inspected alternatives. (iii) Stop and choose an inspected alternative if its observed value is positive, is the highest observed value among inspected options, and exceeds \( \mathcal{I}^S(m) \) as well as the index of any option in the consideration set. (iv) Stop and take the outside option if \( \mathcal{I}^S(m) \), all indices of uninspected options, and all values of inspected options are negative.

The reservation price \( \mathcal{I}(F, \lambda) \) of an alternative has the following interpretation. Suppose there are only two alternatives. One is the alternative \( i \) characterized by \( (F, \lambda) \), and the other is a hypothetical alternative, \( j \), with a known value \( u_j \). The reservation price is the size of \( u_j \), multiplied by \((1 - \delta)\), for which the DM is indifferent between taking \( j \) and inspecting the alternative \( i \) while maintaining the option to recall \( j \) once the value \( u_i \) is observed.

The reservation price of search extends this interpretation. Suppose there are two options: the hypothetical alternative \( j \), and the option of expanding the consideration set. The reservation price of search is the value of \( j \) for which the DM is indifferent between taking \( j \) and expanding the consideration set, maintaining the option to take \( j \) either (a) once the characteristics of the new alternative \( (F, \lambda) \) are observed, or (b) after the value of this new alternative is observed, and under the assumption that \( j \) is recalled in (a) only if \( u_j \geq \mathcal{I}(F, \lambda) \).

Rather than committing to a consideration set up front and then proceeding to evaluate the alternatives within the set and choose among them, the size of the DM’s consideration set may reflect information the DM has learned over time. Note that \( \mathcal{I}^S \), which we interpret as the reservation price corresponding to expansion of the consideration set, is independent of any information about the composition of the DM’s consideration set beyond the number of times it

---

8Weitzman defines reservation prices as the solution to \( \lambda = \delta \int_{\mathcal{I}(F, \lambda)}^{\infty} (u - \mathcal{I}(F, \lambda)) dF(u) - (1 - \delta)\mathcal{I}(F, \lambda) \), which yields

$$\mathcal{I}(F, \lambda) = \frac{-\lambda + \delta \int_{\mathcal{I}(F, \lambda)}^{\infty} u dF(u)}{1 - \delta + \delta \Pr(u > \mathcal{I}(F, \lambda)|F, \lambda)}.$$

The reservation prices in (1) are the same, but multiplied by \((1 - \delta)\) to facilitate a comparison with the more general model introduced in the next Section.
has been expanded, and is easy to calculate (see, e.g., Subsection 2.2). Importantly, \( T^S \) need not be equivalent to the DM’s expected value of expanding her consideration set.

As in Weitzman’s problem, the optimal policy is driven by a comparison of independent reservation prices. The setting above and Proposition 1 therefore extend Weitzman’s problem, and its solution based on independent reservation prices, to an environment in which the consideration set can be sequentially expanded, and is endogenous to the DM’s problem. Importantly, as a consequence of Proposition 1, under the optimal policy, despite the fact that alternatives accumulate over time, the relative likelihood of selecting an option remains the same for any pair of options within the consideration set.

This setting is a special case of a more general model presented in Section 3. Before turning to the general model, we describe two applications of the environment above.

### 2.1 Consumer search and eventual purchase: endogenous consideration sets

Despite the importance of sponsored search in modern business activities, the few models of online consumer search that have been developed remain quite restrictive.\(^9\) For example, the literature has typically assumed that consumers click ads sequentially in the order they are displayed, and that click-through-rates depend on positions but not on the ads displayed at the various positions. Such assumptions, however, do not appear to square well with empirical observations.\(^10\) An important feature of online consumer search is that the consumer is presented with a potentially huge amount of search results, which are displayed in a sequence across multiple pages. Most of the options are initially unobservable, and require the consumer to incur the (time, or mental) cost of scrolling through alternatives. Clearly, consumers do not read all of the results. Instead, the set of search results a consumer reads (and considers clicking on) is partial, and fashioned by the way links are assigned by search engines to different pages.

As a first step toward a better understanding of consumer search in such markets, we apply Proposition 1 as follows. When a consumer enters a query on a search engine, a first list of alternatives is presented (the ones displayed on the first page). Reading the text displayed on a page is costly, and adds the alternatives displayed on the page to the consumer’s consideration set. The consumer may also click on one of the alternatives in the consideration set (that is, on one of the links displayed on one of the pages she has read already) in which case she is directed to a vendor’s website (also at a cost). Once the consumer visits a vendor’s website, she learns her valuation for the vendor’s product or service. At any point in time, the consumer can then stop and purchase a product among those offered by those vendors she visited. To recap, at each point in time, the consumer can either read the information displayed in response to the search query (i.e., search), click on one of the texts/links she has read to be redirected to the corresponding

---

\(^9\)Sponsored search advertising accounts for a large fraction of Internet advertising revenues (see, e.g., Edelman, Ostrovsky, and Schwarz, 2007).

\(^10\)For example, in an empirical analysis of consumer search in online advertising markets using data from Microsoft Live, Jezierski and Segal (2015) find that almost half of the users who click on a link do not click in sequential order of positions, and that click-through-rates do depend on the identity of competing ads.
vendor’s webpage (i.e., learn about a specific alternative’s value), or purchase a product, in which case the decision problem ends.\footnote{This formulation implicitly assumes the consumer does not click on a result without having read its text first, and that a purchase cannot be made without having visited a vendor’s webpage. Both assumptions seem quite natural in this context.}

Suppose the results are indexed by natural numbers in increasing order \( i = 1, 2, \ldots \). When (search) result \( i \) is read, and hence added to the consumer’s consideration set, the consumer receives information \((F_i, \lambda_i)\) about the result under consideration. Such information is drawn from a set \( \mathcal{A} \) according to a distribution \( \mathcal{F} \) that reflects the consumer’s beliefs over the way the search engine displays the results associated with the query the consumer submitted. The information \( F_i \) denotes the distribution over the consumer’s ultimate value \( u_i \) for the corresponding vendor’s product (which also includes the product’s price), whereas \( \lambda_i \) represent the cost the consumer assigns to visiting the vendor’s webpage and learning her value \( u_i \) for the vendor’s product (heterogeneity in both \( F_i \) and \( \lambda_i \) may reflect the consumer’s prior experiences, the vendors’ reputations, as well as the consumer’s expectations about the complexity of navigating within vendors’ websites).

Denote the index of product \( i \), once added to the consideration set, by \( I_i \equiv I(F_i, \lambda_i) \). The cost of reading a result is \( c(m) \), where \( c \) is increasing and \( m \) is the number of past searches. Denote by \( I_i^S \equiv I^S(i - 1) \), as defined in (2); upon reaching the \( i \)’th result, \( m = i - 1 \) results have previously been read.

In this formulation, we are assuming that the consumer reads the results in the order with which they are presented. Under such an assumption, the order by which the results are displayed matters but how the search engine bundles different results on different pages is inconsequential (see the next subsection for a case where the bundling plays a role). Also note that, while the consumer reads the results in the order they are presented, she need not click on them in the same order.

The consumer’s initial outside option, normalized to zero, is captured by product \( i = 0 \) (purchasing product 0 is therefore interpreted as quitting and taking the outside option).

This formulation corresponds to a special case of the model in the previous subsection. With the consumer’s optimal behavior described by Proposition 1, the model can then be used to endogenize the probability with which the consumer selects the various products. Choi, Dai and Kim (2018) (and, independently, Armstrong, 2017) derive a static condition characterizing eventual purchase decisions based on a comparison of “effective values,” in a model where consumers face a fixed set of alternatives, and therefore the optimal policy is as characterized by Weitzman. Building on Proposition 1, Proposition 2 below extends the characterization to the case of endogenous consideration sets.

**Proposition 2.** For all \( i \geq 1 \), let \( w_i = \min\{I_i, u_i, I_i^S\} \) denote the “effective value” of product \( i \). The consumer purchases product \( i \) if, for all \( j \neq i \), \( w_i > w_j \) (and only if \( w_i \geq w_j \), for all \( j \neq i \)).

Proposition 2 provides a micro-foundation for why (and the extent to which) higher positions
imply higher eventual purchase probabilities, a property typically exogenously assumed in existing models. The result generalizes Theorem 1 in Choi, Dai and Kim (2018) to a setting where the consumer’s consideration set is endogenous, and as in Choi, Dai and Kim (2018) suggests that consumers’ eventual purchase decisions can be represented as in canonical discrete-choice models, where consumers’ purchase decisions are based on their effective values.

The result follows from the following monotonicity property of the optimal policy in Proposition 1: If a consumer reads a result \( i \), all reservation values and observed values must be no greater than \( I_i^S \). Hence, if \( I_i \geq I_i^S \), she will proceed to inspect \( i \). Similarly, if \( i \) is inspected and \( u_i \geq I_i \), she will proceed to purchase \( i \) immediately. Note that an immediate implication is that, if the cost of reading \( c(\cdot) \) is strictly increasing, all thing equal, the further a result \( i \) is down the list the more likely it is that its effective value coincides with \( I_i^S \). Therefore, conditional on being read, results further down are more likely to be clicked on and purchased immediately.

A key difference with respect to an environment with an exogenous set of alternatives is that reading additional results is a substitute for inspection of ones that have already been read. This leads to very different dynamics. For example, suppose there is a fixed set of alternatives (and hence search is à la Weitzman), and consider the effects of a reduction in inspection costs or prices such that the reservation prices of all products increase uniformly. In the exogenous environment, the optimal policy (and hence eventual purchase decisions) remain unchanged. However, with an endogenous set of alternatives, this shifts the balance between expansion of the consideration set and inspection of alternatives, and therefore has implications for the size of the consideration set and the ultimate purchase decisions. The following subsection provides an illustration of this tradeoff and its implications.

### 2.2 Consumer search and competition among multi-product firms

To further highlight some of the new tradeoffs the possibility of expanding the consideration set introduces (as part of the DM’s optimal policy), consider the following stylized environment. There are three firms advertising on a platform. Each firm \( i \in \{1, 2, 3\} \) has two different products with similar characteristics (meaning that \( (F, \lambda) \) is the same for both products; the value the buyer attaches to the two products may be different though), and a representative consumer seeks to purchase at most one of these product. Formally, each firm \( i \)’s product is characterized by the pair \( (u_i, p_i) \), where \( u_i \) represents the product’s value to the consumer in case it is a good match for her tastes, and \( p_i \) is the probability with which the product is a good match for the consumer. If a product is not a good match, its value is 0.\(^{12}\) The consumer learns \( u_i \) and \( p_i \) by reading the product’s ad, but must inspect the product (e.g., by clicking on the ad to be directed to the vendor’s website) to learn whether it is a good match or not, and in order to finalize the purchase (as in the rest of the search literature, we assume that the consumer cannot purchase the product without visiting the vendor’s website). Whether or not a product is a good match for

\(^{12}\)Once again, \( u \) is the consumer’s value, net of the product’s price.
the consumer is independent across products. For example, firms may be hotel chains advertising their hotels in a given location, on a platform such as Kayak or Expedia.

We consider two environments. In the first, the consumer’s consideration set is fixed in advance, while in the second it is endogenously determined as part of the consumer’s optimal policy. As it will become clear from the discussion below, the two environments have very different implications for the desirability to increase the probability an ad is displayed.

**Fixed consideration set.** There are four slots for advertising products, each equally visible. Each firm receives a single slot, and the remaining slot, to be used for advertising an additional product, is assigned to one of the firms randomly. Specifically, suppose the probability with which firm $i$ gets to display its additional ad is $\gamma_i \in [0, 1]$, with $\gamma_1 + \gamma_2 + \gamma_3 = 1$. In this fixed-consideration-set scenario, we assume the identity of the firm that receives the additional slot is determined ex-ante. Once the consumer visits the platform’s webpage, she sees all four products (that is, all four products are in the consumer’s consideration set at the beginning of the exploration/inspection process). The consumer sequentially decides which product to inspect and when to stop, at which point she either chooses an inspected product or her outside option (normalized to zero). To keep things simple, suppose the consumer incurs no cost for inspecting a product other than the time cost of postponing the purchase of the final product. The consumer discounts time geometrically with a discount factor $\delta$. In other words, the consumer faces a standard problem à la Weitzman (1979), with four products: $(u_1, p_1), (u_2, p_2), (u_3, p_3)$, and $(u_j^s, p_j^s)$, where $j \in \{1, 2, 3\}$ is the identity of the firm selected by the platform to display the second ad and $(u_j^s, p_j^s)$ are the characteristics of the second ad. To make things simple, assume that $(u_j^s, p_j^s) = (u_j, p_j)$ for all $j \in \{1, 2, 3\}$, meaning that the two products that each firm displays are identical in the eyes of the consumer prior to visiting the firm’s webpage and learning whether each product is a good match or not.

It is easy to verify that, given $(u, p)$, the reservation price $I(u, p)$ is equal to

$$I(u, p) = \frac{(1 - \delta)pu}{1 - \delta + \delta p}.$$  

(3)

The optimal policy is to inspect products in descending order of their reservation prices, stopping when the remaining reservation prices are all smaller than the maximal realized value among the inspected products.

Firms are interested in maximizing the probability with which one of their products is selected. For simplicity, here we assume that each firm makes equal profits on each of its two products. Clearly, in this environment, any firm $i$ benefits from an increase in the probability $\gamma_i$ it is given the second slot.

**Endogenous consideration set.** Now suppose search results are displayed on two separate pages. All three firms advertise on the first page, but one of them, selected at random, is also offered the possibility to advertise on the second page.$^{13}$ Thus, in this case, there are three

---

$^{13}$This stylized setting easily extends to one with more than two pages and/or with a stochastic number of slots.
products in the consumer’s initial consideration set, one for each firm. The consumer has the option of expanding her consideration set by visiting the second page. If the consumer does so, the identity of the firm selected to display the additional ad is revealed to the consumer. As indicated above, the probability with which each firm \( i \) gets to display the additional ad is \( \gamma_i \in [0, 1] \). By visiting the second page, the consumer thus adds a new product to her consideration set with, with each product \((u_i, p_i)\) selected with probability \( \gamma_i \).

Again, there is no direct cost for either inspecting a product or for expanding the consideration set. The cost of each decision is simply the opportunity cost of waiting a period to make a different decision, with each period discounted according to \( \delta \).\(^{14}\)

The consumer’s optimal policy is characterized by Proposition 1, and is based on a comparison of reservation prices \((1)\) and \((2)\). Note that, in this environment, the reservation prices corresponding to each product are \((3)\), the same as in the case of a fixed consideration set.

Without loss of generality, suppose firms are ordered in decreasing order of reservation prices \( I_i \). It is easy to verify from \((2)\) that, in this case, the reservation price of expanding the consideration set (equivalently, its index) is equal to

\[
I_S^* = \delta^2 \max_{k \in\{1, 2, 3\}} \left\{ \frac{\sum_{i=1}^k \gamma_i p_i u_i}{1 + \sum_{i=1}^k \gamma_i \left(1 + \frac{p_i \delta}{1 - \gamma} \right)} \right\}. \tag{4}
\]

Importantly, note that \( I_S^* \) is different from the consumer’s expected payoff from expanding her consideration set. Whether or not \( I_S^* \) takes into account the benefits of inspecting a particular type of product depends on the relationship between the reservation price of that product and \( I_S^* \) itself. In particular,

\[
I_S^* = \begin{cases} 
\frac{\delta^2 \gamma_1 p_1 u_1}{1 + \gamma_1 \left(1 + \frac{p_1 \delta}{1 - \gamma} \right)} & \text{if } I_2 < \frac{\delta^2 \gamma_1 p_1 u_1}{1 + \gamma_1 \left(1 + \frac{p_1 \delta}{1 - \gamma} \right)} < I_1 \\
\frac{\sum_{i=1}^2 \gamma_i p_i u_i}{1 + \sum_{i=1}^2 \gamma_i \left(1 + \frac{p_i \delta}{1 - \gamma} \right)} & \text{if } I_3 < \frac{\sum_{i=1}^2 \gamma_i p_i u_i}{1 + \sum_{i=1}^2 \gamma_i \left(1 + \frac{p_i \delta}{1 - \gamma} \right)} < I_2 \\
\frac{\sum_{i=1}^3 \gamma_i p_i u_i}{1 + \sum_{i=1}^3 \gamma_i \left(1 + \frac{p_i \delta}{1 - \gamma} \right)} & \text{if } I_3 < \frac{\sum_{i=1}^3 \gamma_i p_i u_i}{1 + \sum_{i=1}^3 \gamma_i \left(1 + \frac{p_i \delta}{1 - \gamma} \right)} < I_3 
\end{cases} \tag{5}
\]

Suppose again that firms are interested in maximizing the probability with which a product of theirs is selected, given the consumer’s optimal policy. Perhaps surprisingly, in this case, a firm may suffer from an increase in the probability it is given a second slot, even if this implies that its competitors are less likely to display their ads and even if all products have a positive expected value in the eyes of the consumer.

\(^{14}\)The assumption that expanding the consideration set takes the same amount of time as inspecting a product is completely innocuous and made only for simplicity. In Section 4, we discuss how all of our analysis can be extended to accommodate for the possibility that the time it takes to evaluate an alternative and to expand the consideration set may differ.
Figure 1: The change $\phi(\zeta) - \phi(0)$ in the probability with which a product of firm 2 is selected, as a function of $\zeta$ (in blue). The horizontal gray lines represent the indices $I_1, I_2, I_3$, and the dark gray curve represents $I^S$ as a function of $\zeta$.

For concreteness, let $\delta = 0.9$ and suppose the three types of products are: $(u_1, p_1) = (10, \frac{1}{10})$, $(u_2, p_2) = (3, \frac{1}{3})$, and $(u_3, p_3) = (2, \frac{1}{2})$. Note that the lotteries corresponding to each of the products have the same mean value, but are mean preserving spreads of one another; hence $I_1 > I_2 > I_3$. Initially, $\gamma_1 = \gamma_2 = \frac{1}{4}$, and $\gamma_3 = \frac{1}{2}$. Using the above definitions of the indices, it is easily verified that $I_1 = 0.473$, $I_2 = 0.225$, $I_3 = 0.163$, and

$$I^S = \frac{\gamma_1 p_1 u_1 + \gamma_2 p_2 u_2}{1 + \gamma_1 \delta \left(1 + \frac{p_1 \delta}{1-\delta}\right) + \gamma_2 \delta \left(1 + \frac{p_2 \delta}{1-\delta}\right)} = 0.174.$$  

Also note that $I_3 < I^S < I_2$. As a result, $I^S$ does not take into account the benefits from inspecting firm 3’s additional product, in case firm 3 is the one that is selected to display on the second page.

Now suppose the search engine increases the probability firm 2 is selected on the second page, at the expense of firm 1. Precisely, suppose that $\gamma_2$ is increased by $\zeta \in [0, 0.25]$ while $\gamma_1$ is reduced by the same amount. Let $\phi(\zeta)$ denote the probability that one of firm 2’s products is ultimately chosen when firm 2 is given the second ad with probability $\gamma_2 + \zeta$. Figure 1 depicts the change $\phi(\zeta) - \phi(0)$ in the probability that one of firm 2’s products is selected as a function of $\zeta$, where $\phi(0) = (1 - p_1)(p_2 + (1 - p_2)\gamma_2 p_2) = 0.35$. The horizontal gray lines correspond to the indices $I_1$, $I_2$, and $I_3$, whereas the dark gray curve depicts $I^S$, as a function of $\zeta$. Note that $I^S$ is decreasing in $\zeta$. This is because $I_1 > I_2$. Hence, an increase in $\zeta$ implies a lower reservation value for expanding the consideration set. $I^S$ starts out above $I_3$, and intersects $I_3$ at $\zeta^*$ (the
vertical dashed line). For $\zeta < \zeta^*$, $I_3 < I_S < I_2$, whereas for $\zeta > \zeta^*$, $I_S < I_3$. The function $I^S(\zeta)$ has a kink at $\zeta = \zeta^*$ (see (5)). For $\zeta \in [0, \zeta^*)$, the consideration set is expanded before firm 3’s product is inspected, while for $\zeta \in (\zeta^*, 0.25]$ the opposite is true. The probability that one of firm 2’s products is chosen is equal to $\phi(\zeta) = (1 - p_1)(p_2 + (1 - p_2)(\gamma_2 + \zeta)p_2)$ for $\zeta \in [0, \zeta^*)$ and is equal to $\phi(\zeta) = (1 - p_1)(p_2 + (1 - p_2)(1 - p_3)(\gamma_2 + \zeta)p_2)$ for $\zeta \in (\zeta^*, 0.25]$, with a downward discontinuity at $\zeta = \zeta^*$ equal to $(1 - p_1)(1 - p_2)p_3\zeta^*(\gamma_2 + \zeta)p_2$. Furthermore, the downward drop in $\phi(\zeta)$ at $\zeta = \zeta^*$ makes $\phi(\zeta) - \phi(0)$ negative over $(\zeta^*, 0.25]$.

We summarize the above observation in the following proposition.

**Proposition 3.** If the consideration set is endogenous, reducing the probability that firm 2’s second product is advertised while increasing by the same amount the probability that firm 1’s second product is advertised may lead to an increase in the overall probability with which a product of firm 2 is purchased.

The result illustrates some of the new tradeoffs that emerge when the set of alternatives in a DM’s consideration set is endogenous. Contrary to the case of an exogenous consideration set, firm 2 suffers from displaying a second ad with a higher probability. This is because the increase in the probability that firm 2’s product is displayed on the next page reduces the overall attractiveness of expanding the consideration set inducing the consumer to inspect firm 3’s product prior to expanding the consideration set. This new effect is detrimental to firm 2 when $\zeta > \zeta^*$. 15

### 3 Gradual resolution of uncertainty

In the model in the previous section, the resolution of uncertainty concerning the profitability of each alternative takes a single exploration, as in Weitzman’s (1979) original work. In this section, we relax such an assumption. We first consider a different version of the problem in which the DM alternates between the exploration of the alternatives in the consideration set and the expansion of the latter, without having to commit irreversibly to any alternative. This version is identical to the one in the multi-arm bandit literature, modulo the endogeneity of the set of arms. We allow for general processes governing the gradual resolution of uncertainty, establish the optimality of a certain index policy, and relate the dynamics of search and exploration to the search technology and the primitives of the model. We then enrich this problem by adding to it an irreversible choice, like in Weitzman (1979). In each period the DM can explore one of the alternatives in the consideration set, expand the latter by searching for new alternatives, or commit irreversibly to one of the alternatives in the consideration set that have been explored for long enough. As explained above, Weitzman’s problem corresponds to the special case of this problem in which, in addition to the consideration set being exogenous and constant over time, it takes a single exploration to learn each alternative’s value. Below, we show how the analysis in the previous

---

15 This intuition clearly extends to a more general environment with more than two pages and any number of slots on each page.
section extends to a more general class of problems in which (a) the value of each alternative may never be fully revealed, with each exploration bringing new information, (b) commitment to any alternative in the consideration set may be possible only after an arbitrary number of explorations (in Weitzman (1979), this number is one), and (c) the payoff from selecting an alternative in the consideration set without committing to it may remain strictly positive in each period and evolve according to an arbitrary process.

For concreteness, we continue to conduct the analysis in the context of sequential learning. However, it should be clear that the analysis applies more generally to a broader class of problems in which the evolution of the different alternatives may originate in shocks other than the arrival of information, and where the endogeneity of the shocks may reflect for example a preference for variety, or habit formation.

3.1 General framework

In each period $t = 0, ..., \infty$, the DM can either explore an alternative in her consideration set, or expand her consideration set by searching for new alternatives. Denote by $C_t \equiv (0, ..., n_t)$ the period-$t$ consideration set, with $n_t \in \mathbb{N}$. The DM’s consideration set in period 0 is $C_0 \equiv (0, ..., n_0)$.

Given $C_t$, search brings a (stochastic) set of new alternatives $C_{t+1} \setminus C_t = (n_t + 1, ..., n_{t+1})$ that the DM can explore in subsequent periods and which are added to the current consideration set. Let $x_{it} \in \{0, 1\}$ denote the decision to focus on alternative $i$ in period $t$ (equivalently, to explore alternative $i$), with $x_{it} = 1$ if the DM focuses on $i$ and $x_{it} = 0$ otherwise. Let $x_t \equiv (x_{jt})_{j=0}^\infty$, and denote by $x \equiv (x_s)_{s=0}^\infty$ a complete sequence of attention/exploration decisions. Search involves a direct cost $c_t \geq 0$. Let $y_t \in \{0, 1\}$ denote the decision to search in period $t$, with $y_t = 1$ if the DM searches and $y_t = 0$ otherwise. Denote by $y \equiv (y_s)_{s=0}^\infty$ a complete sequence of search decisions. The period-$t$ overall decision is summarized by $d_t \equiv (x_t, y_t)$. A sequence of decisions $d \equiv (x, y)$ is feasible if for all $t \geq 0$, (i) $x_{jt} = 1$ only if $j \in C_t$, and (ii) $\sum_{j=0}^\infty x_{jt} + y_t = 1$. The DM may also have an outside option, which we normalize to zero.\(^\text{16}\)

Selecting alternative $i \in C_t$ at period $t$ generates a stochastic flow payoff $u_{it} \in \mathbb{R}$, the distribution of which is a function of the alternative’s state. The dependence of flow payoffs on the state of the various alternatives, as well as the description of the search technology are outlined below.

Each alternative has a type $\xi$, an element of an arbitrary type space $\Xi$, which determines the stochastic process governing the evolution of its state. The process corresponding to each alternative of type $\xi$ is Markov and time-homogeneous (i.e., invariant to calendar time). Slightly abusing notation, denote by $\omega^P = (\xi, \theta) \in \Omega^P = \Xi \times \Theta$ the alternative’s current state, where $\theta$ is an element of an arbitrary set $\Theta$. The superscript $P$ is meant to highlight that this is the state of the process at period $t$.

\(^{16}\)To allow for this, simply assume alternative $0$ is degenerate, with a deterministic payoff equal to zero at all periods. One can also let the search decision correspond to a particular alternative in the consideration set. While this poses no problem from a mathematical standpoint, the definition of “consideration set” favored in the literature, but also in the popular language, suggests it is best to keep search separated from the alternatives in the consideration set.
of a physical alternative in the consideration set, not the state of the search technology, or the overall state of the decision problem, which we will define below.

Depending on the application, $\theta$ may have different interpretations. In our sequential learning environment, it is natural to interpret $\theta$ as the history of signals the DM has received about the profitability of the alternative under consideration. However, it could also represent the DM's history of beliefs about that alternative, or a sufficient statistics of the latter (e.g., her current belief). In the case of habit-formation, but also in certain learning environments, it may represent the history of payoffs the DM received over time are governed by the same process describing the evolution of the alternative’s state (that is, they can be represented by deterministic function of the alternative’s state). The first time the DM focuses on an alternative of type $\xi$, the latter is in state $(\xi, \theta_0)$ where, without loss of generality, $\theta_0$ can be taken to be the same across all $\xi$. We do not model the mappings from states to payoffs as the analysis below does not require us to do it.

The process governing the cost incurred due to search, the number of new alternatives introduced as the result of search, and their types, is also Markov time-homogeneous. The search technology’s state is summarized by $\omega^S$, which consists of the history $((c_0, E_0), (c_1, E_1), \ldots, (c_m, E_m))$ of past search costs and of alternatives’ types added to the consideration set. Here $m \in \mathbb{N}$ denotes the number of times search has been carried out in the past, and $E_k = (n_k(\xi) : \xi \in \Xi)$ is a vector representing, for each alternative’s type $\xi$, the number of alternatives $n_k(\xi)$ of type $\xi$ found as the result of the $k$‘th search.\footnote{The first time search is carried out, its state is $(c_0, E_0)$, where $c_0$ can be taken arbitrarily (it plays no role in the analysis since the cost of the first search is $c_1$), and $E_0$ is a description of the types of alternatives in $c_0$.} Denote the set of possible states of search by $\Omega^S$. The distribution over the cost and the set of new alternatives added to the consideration set is denoted by $H_{\omega^S}$.\footnote{Note that this formulation allows the search technology to depend in flexible ways on the results of previous searches. The key assumptions are that the search process is time-homogeneous, and that the outcome of each new search is drawn from $H_{\omega^S}$ independently from the idiosyncratic and time-varying component $\theta$ of each alternative in the consideration set.}

Next, we define the state of the decision problem as follows. For each $\omega^P \in \Omega^P$, let $S^P(\omega^P) \in \mathbb{N}$ denote the number of alternatives within the consideration set in state $\omega^P$. The state of the problem is given by the pair $S = (\omega^S, S^P)$, where $S^P : \Omega^P \to \mathbb{N}$ is a function describing, for each $\omega^P \in \Omega^P$, the number of alternatives within the current consideration set in state $\omega^P$. Next let $\Omega = \Omega^P \cup \Omega^S$ and note that $\Omega^P \cap \Omega^S = \emptyset$. With an abuse of notation, we will sometime find it useful to denote the entire state of the decision problem as a function $S : \Omega \to \mathbb{N}$ that specifies, for each $\omega \in \Omega$, including $\omega \in \Omega^S$, the number of alternatives, including the option of searching, in state $\omega$. We will then denote by $S_t$ the state of the decision problem at the beginning of period $t$. Clearly with this representation, at each period $t$ there is a unique $\omega^* \in \Omega^S$ such that $S_t(\omega^S) = 1$ if $\omega^* = \hat{\omega}^*$ and $S_t(\omega^S) = 0$ if $\omega^* \neq \hat{\omega}^*$. The special case where the decision maker does not have
the option to search corresponds to the case where $S_t(\omega^S) = 0$ for all $\omega^S \in \Omega^S$, all $t$.

Defining the state of the decision problem this way allows us to keep track of all relevant information, and greatly facilitates the analysis.

### 3.2 Optimal policy

A policy $\chi$ for the above decision problem is a rule governing the decisions in each period – whether to focus attention on an alternative in the consideration set or searching for new ones – based on the available information. That is, given a sequence of feasible decisions $(d_t)_{t \geq 0}$, the state process $(S_t)_{t \geq 0}$ generates a natural filtration $\{\mathcal{F}_t\}_{t \geq 0}$. A policy $\chi$ is then a $\mathcal{F}_t$-measurable sequence of feasible decisions. Denote by $U_t \equiv \sum_{j=0}^{\infty} x_{jt} u_j - c_t y_t$ the realized period-$t$ net payoff. A policy $\chi$ is optimal if it maximizes the expected discounted sum of the net payoffs $E \left[ \sum_{t=0}^{\infty} \delta^t U_t | S_0 \right]$, where $\delta \in (0, 1)$ denotes the discount factor. To guarantee that the process of the expected net payoffs is well behaved, we assume that for any state $S$ and policy $\chi$, $\delta^T E \left[ \sum_{s=0}^{\infty} \delta^s U_s | S \right] \to 0$ as $T \to \infty$.\(^\text{19}\)

For each $\omega^P \in \Omega^P$, let

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

 denote the “index” of an alternative currently in state $\omega^P$, where $\tau$ denotes a measurable stopping time.\(^\text{20}\) The definition in (6) is equivalent to the definition in Gittins (1979). Given the state $S$, denote the maximal index among the alternatives within the DM’s consideration set by $I^*(S^P) = \max_{\omega^P \in \hat{\Omega}^P: S^P(\omega^P) > 0} \mathcal{I}(\omega^P)$. Note that $I^*(S)$ depends on $S$ only through the state of the alternatives in $S$.

We now define an index for search that depends on the state of the alternatives in the consideration set only through the information that the latter contain for the evolution of the search technology. Given the representation introduced above, such information is represented by the number of alternatives of each type $\xi \in \Xi$ brought to the set by past searches. Such information is already encoded in the state $\omega^S$. As a result, the index of search depends on the state of the entire decision problem only through $\omega^S$. 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 expected arrival of alternatives as the result of search implies that the maximization in the definition is not just over the stopping time, but also over the rule governing the allocation of future attention among the new alternatives and further

\(^{19}\)This property guarantees the solution to the Bellman equation corresponding to the above dynamic program coincides with the true value function; it is immediately satisfied if payoffs and costs are uniformly bounded.

\(^{20}\)Specifically, $\tau$ is a stopping time with respect to the process whose filtration is obtained by focusing attention on the alternative with initial state $\omega^S$ in all periods.

18
search. Denote by $\tau$ a measurable stopping time, and by $\pi$ a measurable rule prescribing, for any period $s$ between the current one and the stopping time $\tau$, either the selection of one of the new alternatives brought in by search or further search. Importantly, $\pi$ selects only among search and alternatives that were not already in the consideration set when search was launched.\(^{21}\)

Formally, given the state of the search technology $\omega^S \in \Omega^S$, its index is defined by

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

where $U^\pi_s$ denotes the stochastic net flow payoff obtained in period $s$, under the rule $\pi$.

Denote by $\chi^*$ the policy that selects at each period $t \geq 0$, given the overall state $S_t$ of the decision problem, search if and only if $I^S(\omega^S_t) \geq I^*(S^P_t)$, and otherwise an arbitrary alternative with index $I^*(S^P_t)$.\(^{22}\) Ties between the alternatives may be broken arbitrarily. In order to maintain consistency throughout the analysis, we assume that, in the case of a tie $I^S(\omega^S_t) = I^*(S^P_t)$, the tie is broken in favor of search.

To present the next result, we first introduce the following notation. Let $\kappa(v) \in \mathbb{N} \cup \{\infty\}$ denote the first time at which, when the DM follows an index policy, (i) the search technology reaches a state in which its index is no greater than $v$, and (ii) all alternatives – regardless of when they were introduced into the consideration set – have an index no greater than $v$. That is, $\kappa(v)$ is the minimal number of periods until all indices are weakly below $v$. In case this event never occurs, $\kappa(v) = \infty$. Note that between the current period and the first period at which all indices are weakly below $v$, if the DM searches, new alternatives are introduced, in which case the evolution of their indices must also be taken into account in the computation of $\kappa(v)$.

Let $V^*(S_0)$ denote the maximal expected (per-period) payoff the DM can attain, given the initial state $S_0$. That is, $V^*(S_0) = (1 - \delta) \mathbb{E} \left[ \sum_{t=0}^{\infty} \delta^t U^*_t | S_0 \right]$, with the expectation taken under the optimal rule.

**Theorem 1.** (i) The index policy $\chi^*$ is optimal in the sequential search problem with endogenous consideration sets.

(ii) The DM’s expected (per-period) payoff under the index policy $\chi^*$ is equal to

$$
\int_0^{\infty} \left( 1 - \mathbb{E} \left[ \delta^{\kappa(v)} | S_0 \right] \right) dv.
$$

\(^{21}\)That is, $\pi$ does not select among alternatives present before the launch of search. To make things clear, suppose search is launched in period $t$ and terminated in period $\tau > t$. Then at each period $t < s < \tau$, $\pi$ selects between search and alternatives available in period $s$ but which were not yet in the consideration set in period $t$.

\(^{22}\)Recall that $I^*(S^P_t)$ is the largest index of the “physical” alternatives in the consideration set.
(iii) The index of search, as defined in (7), is such that, for any $\omega^S \in \Omega^S$,

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

(9)

where $\tau^*$ is the first time $s \geq 1$ at which $I^S$ and all the indexes of the alternatives brought in by search fall below the value $I^S(\omega^S)$ of the search index when search was launched, and where the expectations are with respect to the process induced by the index rule $\chi^*$. 

**Proof of Theorem 1.** The proof consists of three steps. Step 1 below first establishes the result in part (iii) and then uses such recursive representation of the index of search to show that, when the DM follows an index policy, her expected (per-period) payoff satisfies the representation in (8), thus establishing part (ii). Steps 2 and 3, in the Appendix, then use this representation to show that the DM’s payoff under such a rule satisfies the Bellman equation for the dynamic program under consideration, thus proving the optimality of the index policy.

**Step 1.** We start by proving part (iii). Clearly, by definition, for all $\omega^S$,

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

To see that the opposite inequality must also hold, note that by definition of $\tau^*$ in (iii), between period $s = 0$ and period $\tau^*-1$, any option chosen under the index policy (whether exploring an alternative or searching again) must yield an expected per-period average payoff greater than $I^S(\omega^S)$. Therefore, following the index policy until $\tau^*-1$, the expected average payoff on the right-hand-side of (9) must be greater than $I^S(\omega^S)$.

Next, consider part (ii). We first introduce some notation. Define $S^1 \lor S^2 \equiv (S^1(\omega) + S^2(\omega) : \omega \in \Omega)$ and $S^1 \setminus S^2 \equiv (\max\{S^1(\omega) - S^2(\omega), 0\} : \omega \in \Omega)$. Any feasible state of the decision problem must specify one, and only one, state of the search technology (i.e., one state $\hat{\omega}^S$ for which $S_t(\hat{\omega}^S) = 1$ and such that $S_t(\omega^S) = 0$ for all $\omega^S \neq \hat{\omega}^S$). However, it will be convenient to consider fictitious (infeasible) states where search is not possible, as well as fictitious states with multiple search possibilities. If the state of the decision problem is such that either (i) the consideration set is empty, or (ii) there is a single alternative in the consideration set and the latter cannot be expanded, we will denote such state by $e_t(\omega)$, where $\omega \in \Omega$ is the state of the search technology in case (i) and of the single physical alternative in case (ii). Observe that the independence of the processes governing the evolution of the alternatives (conditional on their types $\xi$) along with the independence of these processes from the evolution of the search technology (conditional on the information about the types of the alternatives in the consideration set encoded into $\omega^S$) imply that, for any $v \in \mathbb{R}$ and states $S^1$ and $S^2$, $\kappa(v|S^1 \lor S^2) = \kappa(v|S^1) + \kappa(v|S^2)$. That is, the time it takes to bring all indexes below $v$ when the state of the decision problem is $S^1 \lor S^2$ is the sum of the times it takes to accomplish the same thing when the state is $S^1$ and $S^2$, respectively.
We construct the following stochastic process based on the values of the indices, and the expansion of the consideration set through search, under the index policy. Starting with the initial state $S_0 = (S_0^P, \omega_0^S)$, let $v^0 = \max\{I^*(S_0^P), I^S(\omega_0^S)\}$. Consider the first time $t^0$ in which, when the DM follows the policy $\chi^*$, all indices are strictly below $v^0$, with $t^0 = \infty$ if this event never occurs. Note that $t^0$ differs from $\kappa(v^0)$, as $t^0$ is the first time at which all indices are strictly below $v^0$, whereas $\kappa(v^0) = 0$ is the first time at which all indices are weakly below $v^0$.

Next let $v^1 = \max\{I^*(S_0^P), I^S(\omega_0^S)\}$, where $S_0^P = (S_0^P, \omega_0^S)$ is the state of the system at time $t^0$. Note that, by construction, and $t^0 = \kappa(v^1)$. Furthermore, if $t^0 < \infty$ then $v^0 > I^S(\omega_0^S)$ implies $\omega_t^S = \omega_0^S$. We can proceed in this manner to obtain a stochastic, strictly decreasing, sequence of values $(v^i)_{i \geq 0}$, with corresponding stochastic times $(\kappa(v^i))_{i \geq 0}$. Next, for any $i = 0, 1, 2, \ldots$, let $\eta(v^i) = \sum_{s=\kappa(v^i)}^{\kappa(v^{i+1})-1} U_s$ 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 $(\eta(v^i))_{i \geq 0}$ define the corresponding sequence of discounted accumulated net payoffs, with $\eta(v^0) = 0$ if $\kappa(v^0) = \infty$.

Denote by $\mathcal{V}(S_0)$ the expected (per-period) net payoff under the index policy $\chi^*$, given the initial state of the problem $S_0$. That is, $\mathcal{V}(S_0) = (1 - \delta)\mathbb{E}^x \left[ \sum_{t=0}^{\infty} \delta^t U_t | S_0 \right]$. Below we identify properties of the function $\mathcal{V}(S_0)$ that allow us to establish the optimality of $\chi^*$. It should be clear that the same properties apply to $\mathcal{V}(S_t)$, for any $t \geq 0$ (to see this, it suffices to note that the value of $S_0 = (\omega_0^S, S^P)$ plays no role in the arguments below). We start with the following result:

**Lemma 1.** Given the initial state $S_0$, the expected (per period) payoff under the policy $\chi^*$ is

$$\mathcal{V}(S_0) = \int_0^{\infty} \left( 1 - \mathbb{E} \left[ \delta^{\kappa(v)} | S_0 \right] \right) dv. \quad (10)$$

**Proof of Lemma 1.** First, observe that, by definition of the processes $(\kappa(v^i))_{i \geq 0}$ and $(\eta(v^i))_{i \geq 0}$,

$$\mathcal{V}(S_0) = (1 - \delta)\mathbb{E} \left[ \sum_{i=0}^{\infty} \delta^{\kappa(v^i)} \eta(v^i) | S_0 \right].$$

Next, using the definition of the indices (6) and (7), along with the properties of the optimal stopping times in the definition of the indices described above, we have that

$$v^i = \frac{(1 - \delta)\mathbb{E} \left[ \eta(v^i) | S_{\kappa(v^i)} \right]}{\mathbb{E} \left[ 1 - \delta^{\kappa(v^{i+1}) - \kappa(v^i)} | S_{\kappa(v^i)} \right]}, \quad (11)$$

To see why (11) holds, recall that, at period $\kappa(v^i)$, given the state of the system $S_{\kappa(v^i)}$, the optimal stopping time in the definition of the index $v^i$ is the first time at which the index of the corresponding alternative (if $v^i$ corresponds to a “physical” alternative), or the index of search and all alternatives introduced through future searches (in case $v^i$ corresponds to the index of the search technology), drop below $v^i$.23

23Note that if, at period $\kappa(v^i)$, there are multiple options (“physical” alternatives and search) with index $v^i$, the
Rearranging, multiplying both sides of (11) by \( \delta(\kappa(v)) \), and using the fact that \( \delta(\kappa(v)) \) is known at \( \kappa(v) \), we have that

\[
\sum_{i=0}^{\infty} v^i \left( \delta(\kappa(v)) - \delta(\kappa(v+1)) \right) = \int_0^\infty v d\delta(\kappa(v)),
\]

for a particular path with \( \kappa(v^3) = \infty \).

Taking expectations of both sides of the previous equality given the initial state \( S_0 \), and using the law of iterated expectations, we have that

\[
\mathbb{E} \left[ \sum_{i=0}^{\infty} v^i \left( \delta(\kappa(v)) - \delta(\kappa(v+1)) \right) \mid S_0 \right] = \left( 1 - \delta \right) \mathbb{E} \left[ \delta(\kappa(v)) \eta(v) \mid S_0 \right].
\]

If follows that

\[
\mathcal{V}(S_0) = \mathbb{E} \left[ \sum_{i=0}^{\infty} v^i \left( \delta(\kappa(v)) - \delta(\kappa(v+1)) \right) \mid S_0 \right].
\]

Next, note that \( \delta(\kappa(v)) = 0 \) whenever \( \kappa(v) = \infty \), and that for any \( i = 0, 1, ..., \kappa(v) = \kappa(v^{i+1}) \) for all \( v^{i+1} < v < v^i \). It follows that (12) is equivalent to

\[
\mathcal{V}(S_0) = \mathbb{E} \left[ \int_0^\infty v d\delta(\kappa(v)|S_0) \right] = \mathbb{E} \left[ \int_0^\infty \left( 1 - \delta(\kappa(v)) \right) dv \mid S_0 \right] = \int_0^\infty \left( 1 - \mathbb{E} \left[ \delta(\kappa(v)|S_0) \right] \right) dv.
\]

average sum \( \mathbb{E} \left[ \eta(v^i) \mid S_{\kappa(v^i)} \right] \) of the discounted net payoffs across all alternatives with index \( v^i \) until the indices of all such options drop below \( v^i \) (in case search is included, also those of the new arriving alternatives), per unit of average discounted time, \( \mathbb{E} \left[ 1 - \delta(\kappa(v^{i+1})) - \kappa(v^i) \mid S_{\kappa(v^i)} \right] / (1 - \delta) \), is the same as the average sum of the discounted net payoffs of each individual option with index \( v^i \) normalized by the average discounted time until the index of that alternative falls below \( v^i \). This follows from the independence of the processes. Hence, Condition (11) holds irrespectively of whether, at \( \kappa(v^i) \), there is a single or multiple options with index \( v^i \).
The construction of the integral function in Lemma 1 is illustrated in Figure 2.

**Steps 2 and 3.** In the Appendix, we use the result in Lemma 1 to characterize the difference in the DM’s expected payoff between (i) following the index policy $\chi^*$ from the outset and (ii) search in the first period and then reverting to $\chi^*$ from the next period onward. To do so, we introduce an additional fictitious “retirement option,” which is available at all periods and which yields a constant payoff $M < \infty$. This characterization is instrumental for Step 3, which proves the optimality of $\chi^*$ by showing that $V$ solves the Bellman equation. Together with some arguments from dynamic programming, this result implies that $\chi^*$ is indeed an optimal policy thus establishing Part (i) in the Theorem.

### 3.3 Search and exploration dynamics

Equipped with the result in Theorem 1, we now highlight a few important properties of the dynamics of learning and expansion of the consideration set as a function of the search technology.

**Corollary 1.** At any point in time, the decision to search for new alternatives depends on the composition of the consideration set only through the information the latter contains about the likelihood that new searches will bring alternatives of different types.

The result is an immediate implication of the optimal policy being an index policy. Note that the result is true despite the fact that the opportunity cost of search, which coincides with the value of continuing exploring alternatives in the consideration set, typically depends on the composition of the set over and above the information the latter contains for search outcomes.

**Corollary 2.** At any point in time, the relative likelihood of selecting any pair of alternatives in the consideration set is invariant to the state of the search technology.

Corollary 2 is also an immediate implication of Theorem 1. More generally, the optimal policy being an index policy implies that the relative frequency with which the DM explores any two alternatives in the consideration set is invariant to what the decision maker expects to find by expanding the consideration set. This is true despite the fact that the expansion of the consideration set may bring alternatives that are more similar to some alternatives in the consideration set than others.

**Corollary 3.** An improvement in the search technology yielding an increase in the probability that search brings an alternative of positive expected value (vis-a-vis the outside option) need not affect the decision to search even at histories at which, prior to the improvement, the DM is indifferent between searching and continuing with the current consideration set.

The result follows from the fact that improvements in the search technology need not imply an increase in the search index. This is because the optimal stopping time in the definition of the search index coincides with the first time at which the index of search and the indexes of all
alternatives brought in by search fall below the value of the search index at the time search was launched. As a result, any marginal improvement in the search technology affecting only those alternatives whose index at the time of arrival is below the value of the search index does not affect the decision to search. Note that the result hinges on the fact that the DM needs to explore the various alternatives that search brings to determine their values. When, instead, search is stochastic, but the value of the alternatives search brings is revealed to the DM upon arrival (as in the standard undirected search literature), any marginal improvement in the search technology necessarily breaks the indifference in favor search.

**Definition 1.** A search technology is **stationary** if \( H_{\omega S} = H \) for all \( \omega \in \Omega^S \), **deteriorating** if \((-c_k, E_k)\) is decreasing in \(k\) in the sense of first-order stochastic dominance, and **improving** if \((-c_k, E_k)\) is increasing in \(k\) in the sense of first-order stochastic dominance.\(^{24}\)

**Corollary 4.** If the search technology is stationary, for any two states \(S, S'\) at which the DM optimally searches, \(V^*(S) = V^*(S')\).

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

**Corollary 5.** 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 consideration set at period \(t\).

Since the state of an alternative changes only when the DM focuses on it, if, in period \(t\), \(I^S \geq I^*(S^P)\), under a stationary or improving search technology, the same inequality remains true in all subsequent periods. In this case, search corresponds to disposal of all alternatives within the current consideration set. Each time the DM launches himself into search, he starts fresh.

**Corollary 6.** If the search technology is stationary or deteriorating, at any history, the decision to search is the same as in a fictitious environment where the DM expects he will have only one further opportunity to search.

The result follows again from the fact that the optimal stopping time in the definition of the search index is the first time at which the index of any alternative brought in by search, and the index of search itself, drop below the value of the search index at the time search is initiated. If the search technology is stationary or deteriorating, the index of search falls (weakly) below its initial value immediately after search is launched. Hence, \(I^S(\omega^S)\) is independent of any information pertaining to future states of the search technology, conditional on \(\omega^S\).

\(^{24}\)That is, the search technology is deteriorating if for any \(k\) and any upper set \(Z \in \mathbb{R} \times \mathbb{R}^{|\Xi|}\), \(\Pr((-c_{k+1}, E_{k+1}) \in Z) \leq \Pr((-c_k, E_k) \in Z)\). This definition is quite strong. In more specific environments in which there is an order on the set of types \(\Xi\), weaker definitions are consistent with the results in the corollaries below.
Corollary 7. Suppose the DM does not have the option to search (i.e., \(S_0\) is such that \(S_0(\omega_S) = 0\) for all \(\omega_S \in \Omega^S\)). Let \(\hat{S}_0\) denote the state that coincides with \(S_0\) except for the fact that \(\hat{S}_0(\hat{\omega}_S) = 1\) for some \(\hat{\omega}_S \in \Omega^S\). The DM’s willingness-to-pay to have access to a search technology in state \(\hat{\omega}_S\) is equal to

\[
P^*(S_0; \hat{\omega}_S) = \int_0^\infty \left( E[\delta^v(\omega) | S_0] - E[\delta^v(\omega) | \hat{S}_0] \right) dv.
\]

The result in Corollary 7 can be used to price access to a 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 to an agent with exogenous outside option equal to \(v \in \mathbb{R}_+\) to drop off and take the outside option, both when search is available and when it is not. Then by integrating over the relevant values of the outside option one can compute \(P^*(S_0; \hat{\omega}_S)\) and hence the maximal price that the DM is willing to pay to access the search technology.

3.4 Irreversible choice

In many decision problems, in addition to learning about existing options and searching for new ones, the DM can decide to irreversibly commit to one of the alternatives. Consider Weitzman’s (1979) search problem, where uncertainty about the reward from each box is resolved immediately after opening the box, and where a box can be chosen only if it was previously opened. Under these assumptions, whether or not the choice of a box is reversible is irrelevant: since there is no additional information to be learned about the opened box, there is no reason for the DM to change her selection. Doval (2018) studies a generalization of Weitzman’s problem where a box may be selected even if it has not been previously opened. She shows that the optimal solution takes the form of an index policy only under certain conditions, and studies the case in which the index policy need not be optimal.

In this section, we extend our analysis to a general model of learning, searching for new alternatives, and irreversible choice in which, at each period, the DM can (i) focus attention on one of the alternatives within her consideration set, (ii) expand her consideration set through search, or (iii) irreversibly select an alternative from her consideration set based on possibly partial information about its value. Formally, we modify the general model of Section 3 as follows. In addition to the actions \(x_t\) and \(y_t\) defined above, we introduce an additional action, \(z_{jt} \in \{0, 1\}\), representing the irreversible choice of an alternative \(j\) in the consideration set in period \(t\), with \(z_{jt} = 1\) if the DM irreversibly chooses alternative \(j\), and \(z_{jt} = 0\) otherwise. The period-\(t\) complete decision is then summarized by \(d_t \equiv (x_t, y_t, z_t)\), with \(z_t = (z_{jt})_{j=0}^\infty\). A sequence of decisions \(d\) is feasible if, for all \(t \geq 0\), (i) \(x_{jt} = 1\) or \(z_{jt} = 1\), only if \(j \in C_t\), (ii) \(\sum_{j=1}^\infty x_{jt} + y_t + z_{jt} = 1\), and (iii) \(z_{jt} = 1\) if \(z_{js} = 1\) for some \(s < t\). Together, the last two assumptions imply that, once an alternative is chosen (that is, once the DM has committed to it), there are no further decisions to be made.
To allow for the possibility that the DM must explore an alternative multiple times before she can irreversibly commit to it, we assume the DM must focus on each type-\(\xi\) alternative at least \(M_\xi \geq 0\) times before she can choose it (the case \(M_\xi = \infty\) corresponds to the model with no irreversible choice of Section 3). Focusing on an alternative (i.e., exploring it without committing to it – formally captured by \(x_{jt} = 1\)) changes its state and may yield a flow payoff/cost, as in the baseline model. Irreversibly selecting alternative \(j\) of type \(\xi\) in period \(t\) (formally captured by \(z_{jt} = 1\)) yields a flow payoff to the DM from that moment onward, the value of which may be only imperfectly known to the DM at the time the irreversible decision is made. Denote by \(R(\omega^P)\) the expected flow value from choosing an alternative when its current state is \(\omega^P = (\xi, \theta)\).

Note that since the choice is irreversible, \(R(\omega^P)\) admits two equivalent interpretations: (i) If the alternative is chosen, an immediate expected payoff equal to \(R(\omega^P)/(1 - \delta)\) is obtained and there are no further payoffs; (ii) payoffs continue to accrue at all subsequent periods after the irreversible choice is made, with each expected payoff equal to \(R(\omega^P)\).

Now suppose that each alternative’s states can be partially ordered, based on the number of times the DM has focused on the alternative. Formally, suppose the set \(\Theta\) takes the product form \(\Theta = \Theta' \times \mathbb{N}\), with \(m \in \mathbb{N}\) denoting the number of times the DM focused on the alternative and \(\theta'\) all additional information. For any \(\omega^P = (\xi, (\theta', m))\), we say that \(\hat{\omega}^P = (\xi, (\hat{\theta}', \hat{m}))\) if and only if \(\hat{m} \geq m\). Denote this relation by \(\hat{\omega}^P \succeq \omega^P\).

**Condition 1.** A type-\(\xi\) alternative has the **better-later-than-sooner property** if, for any \(\omega^P = (\xi, (\theta', m))\), with \(m \geq M_\xi\), and any \(\hat{\omega}^P \succeq \omega^P\), either \(R(\hat{\omega}^P) \geq R(\omega^P)\), or \(R(\hat{\omega}^P), R(\omega^P) \leq 0\).

The following environments are examples of settings satisfying Condition 1.

**Example 1** (Weitzman’s extended problem.). Consider the following extension of Weitzman’s original problem: (i) The set of boxes is endogenous; (ii) each alternative of type \(\xi\) requires \(M_\xi\) explorations before the alternative’s value is revealed, and, as in Weitzman’s original problem, the DM can irreversibly commit (i.e., select) an alternative only if its value has been fully revealed, i.e., only after \(M_\xi\) explorations, where \(M_\xi\) can be stochastic (in this case, \(\theta'\) specifies also whether or not the alternative can be retired and any information the DM may have about the value of \(M_\xi\)); (iii) the flow payoff from exploring an alternative without committing to it is equal to the cost of exploring the alternative (with the later evolving stochastically with the number of past explorations) and is equal to zero for any exploration \(t > M_\xi\); (iv) the payoff \(R(\omega^P)\) from irreversibly committing to a an alternative whose value has been revealed (i.e., after the \(M_\xi\)’s exploration) remains constant after the \(M_\xi\)’s exploration and is equal to the box’s prize. Clearly, this problem satisfies the better-later-than-sooner property.

**Example 2** (Purchase/Lease problem.). In each period, an apartment owner either chooses one of the real-estate agents of her knowledge to lease her apartment, or searches for new agents. In addition, the owner can use one of the agents to sell the apartment. The decision to sell the apartment is irreversible. Once the apartment is sold, the owner’s problem is over. The
(expected) flow value $u_{jt}$ the owner assigns to leasing the apartment through agent $j$ of type $\xi$ in state $\omega^P = (\xi, (\theta', m))$ is a function of the information $\theta'$ the owner has accumulated over time about agent $j$’s ability to deal with all sorts of problems related to tenants. The (expected) value $R(\omega^P)$ the owner assigns to selling the apartment through the same agent may also depend on the agent’s expertise with tenant-related problems but is primarily a function of the familiarity the agent has with the apartment which is determined by the number of times $m$ the agent has been hired by the owner in the past. If the agent has no or little past experience selling apartments (this information is contained in $\xi$), $R(\omega^P) \leq 0$. Else, for any $\theta'$ and $\hat{\theta}'$, $R(\xi, (\hat{\theta}', \hat{m})) \geq R(\xi, (\theta', m))$ if and only if $\hat{m} \geq m$. Clearly, this problem too satisfies the better-later-than-sooner property of Condition 1. Contrary to Weitzman’s extended problem discussed in the previous example, the DM may derive a higher (expected) value from using an alternative without irreversibly committing to it (i.e., from leasing instead of selling through an agent) for an arbitrary long, possibly infinite, number of periods.

For each alternative in state $\omega^P \in \Omega^P$, let

$$I^P(\omega^P) \equiv \sup_{\pi, \tau} \mathbb{E} \left[ \sum_{s=0}^{\tau-1} \delta^s U^\pi_s | \omega^S \right],$$

be its new index, where $\tau$ is a stopping time, and $\pi$ is a rule specifying whether the DM focuses attention on the alternative or irreversibly commits to it (i.e., chooses it). Similarly, let $I^S(\omega^S)$ denote the index of search in the problem with irreversible choice, with the index as defined in (15) but with the rule $\pi$ now specifying whether to keep searching, exploring one of the alternatives introduced through search, or irreversibly committing to one of the alternatives that the new search brought to the consideration set.

Amend the definition of the index policy $\chi^*$ as follows. At each period $t \geq 0$, given the state $S_t$: (a) search if $I^S$ is greater than the index $I^P$ and the expected value $R$ of each alternative in the consideration set; (b) focus attention on an alternative if its index $I^P$ is greater than its expected value $R$, as well as the index of search, and both the index and expected value $R$ of any other alternative in the consideration set; (c) choose (i.e., irreversibly commit to) an alternative if its retirement value $R$ is greater than its index $I^P$, as well as the index of search and both the index and expected value of any other alternative in the consideration set.

**Theorem 2.** Suppose Condition 1 is satisfied for all $\xi \in \Xi$. The index policy $\chi^*$ is optimal in the extended model with irreversible choice.

### 3.5 Pandora’s problem with endogenous boxes and gradual resolution of uncertainty

Equipped with the above results, we now revisit the extension of Weitzman’s (1979) search problem introduced in Example 2 above. Recall that the problem is a generalization of the one in
Section 2, in which (a) fully learning the value of each alternative (which, as in Weitzman (1979) original work is necessary to irreversibly commit to an alternative) requires an arbitrary number of explorations \( M^= \), and (b) each exploration brings additional information about the alternative’s value drawn from an arbitrary distribution. For simplicity, we assume here that \( M^= \) is known and that there are no direct costs associated with the exploration of the various alternatives or with search.\(^{25}\)

More precisely, denote by \( \omega^P = (\xi, \theta', n) \) the state of each alternative, represented by the alternative’s characteristics, as encapsulated in \( \xi \) (e.g., the prior distribution from which the alternative’s prize is drawn), the history of signal realizations \( \theta' \), and the number of times \( n \) the alternative has been explored. When a new alternative is discovered, \( \xi \) is drawn from \( A \) from a known distribution \( F \). Denote the reservation price of each alternative in state \( \omega^P \) by

\[
I^P(\omega^P) = \operatorname{sup}_{\tau > 0} \mathbb{E} \left[ \frac{\delta^\phi \left( 1 - \delta^{\tau - \phi} \right) 1_{\{\phi < \tau\}} \tilde{u}_{[\omega^P]} }{1 - \mathbb{E} [\delta^{\tau} | \omega^P]} \right],
\]

where \( \tau \) is a (stochastic) stopping-time, \( \phi \) is the (stochastic) time at which all information is revealed about the alternative, and \( \tilde{U} \) denotes the value of the alternative. Similarly, let

\[
I^S(m) = \operatorname{sup}_{\tau > 0} \frac{\delta \int_A \left( \mathbb{E} \left[ \delta^\phi \left( 1 - \delta^{\tau - \phi} \right) 1_{\{\phi < \tau\}} \right] \right) dF(\xi)}{1 - \int_A \left( \mathbb{E} [\delta^{\tau} | \xi] \right) dF(\xi)}
\]

denote the reservation price of search, where \( \tau, \phi \) and \( u \) are as defined above.

In this “multiple-exploration” extension of the environment in Section 2, the DM’s optimal policy is the following.

**Proposition 4.** The DM’s optimal policy is the following: (i) Add an alternative to the consideration set if \( I^S(m) \) is positive and greater than the index of all uninspected options in the consideration set and all observed values of inspected options. (ii) Inspect an alternative in the consideration set if its index is positive, is the greatest in the consideration set, and exceeds \( I^S(m) \) as well as the value of all inspected alternatives. (iii) Stop and choose an inspected alternative if its observed value is positive, is the highest observed value among inspected options, and exceeds \( I^S(m) \) as well as the index of any option in the consideration set. (iv) Stop and take the outside option if \( I^S(m) \), all indices of uninspected options, and all values of inspected options are negative.

We conclude by noting that the above generalization of the Weitzman (1979) model also accommodates for a generalization of the “eventual purchase theorem” of Section 2 (Proposition 2) to a setting in which the resolution of uncertainty is gradual. Consider the environment described in Section 2, to one where it takes \( M^= \) explorations (each with a cost \( \lambda \)) to learn a

\(^{25}\)Introducing direct costs here complicates the notation and the expressions for the reservation prices below, but is otherwise innocuous.
product’s value. Denote by $I_i^k$ the index of product $i$ after it has been added to the consideration set and has been explored $k$ times.

**Proposition 5.** For all $i \geq 1$, denote by $w_i = \min\{I_i^S, I_i^1, I_i^2, ..., I_i^{M-1}, u_i\}$ the “effective value” of product $i$. The consumer purchases product $i$ if, for all $j \neq i$, $w_i > w_j$ (and only if $w_i \geq w_j$ for all $j \neq i$).

### 4 Extensions

In this section, we discuss how the results above accommodate easily accommodate for a few simple extensions that may be relevant for applications.

**Relative length of expansion.** In order to allow for frictions in searching for new options, we assume that whenever the DM searches, she cannot focus her attention on (that is, explore) any of the alternatives in the consideration set. In reality, the length of time search occupies relative to learning about alternatives may differ. For example, the online search for alternative providers of a given service may take seconds, but searching for a potentially suitable candidate for a given position make take longer than an interview. The assumption that the length of time search occupies is identical to a period of focusing attention on an alternative is innocuous. The results immediately extend to a setting where the number of periods search or focusing on an alternative occupy before a different action can be taken can differs as a function of its state; in particular, the length of time may differ not only between search and alternatives, but also across alternatives, and even over time. Since the length of time a focusing on an alternative takes can be made arbitrary large, by rescaling the payoffs and adjusting the discount factor appropriately, the relative length of time for which search creates an interruption can be made arbitrarily small. The result therefore also apply to problems in which search and learning occur “almost” in parallel.

**Multiple expansion possibilities.** As illustrated in an example in the online appendix, if there are multiple options for search for which the outcome is correlated, an index policy cannot be guaranteed to be optimal. However, the analysis readily extends to an environment in which there are multiple search possibilities with independent outcomes, by allowing for the possibility of multiple “search arms”. For example, a researcher may choose in which field to search for a new project. A department with a single new faculty position may choose in which field to search for candidates. The analysis can also be extended to allow the results of search to include not just physical alternatives, but also new search possibilities.

---

26 These lengths of time may also be stochastic, and need not be finite. More generally, all of the results may be extended to a semi-Markov environment, where time is not slotted.

27 Such correlation arises naturally in an environment in which the DM can choose how much to invest in search, with different levels of investment corresponding to different “intensities” of search.
No discounting. The proofs of the results above rely on the assumption that $\delta < 1$. As discussed above, an important special case of our analysis is an extension of Weitzman’s problem with search for new boxes. Many applications of Weitzman’s framework assume no discounting (i.e., $\delta = 1$). Our results extend to this case since, as noted in Olszewski and Weber (2015), Weitzman’s problem with $\delta = 1$ is a special case of a multi-armed bandit problem with undiscounted “target processes”, in which once an arm reaches a certain (target) state, payoffs no longer accrue. A well known result for such problems is that the finiteness they impose allows to take the limit as $\delta \to 1$ (see, e.g., Dumitriu, Tetali, and Winkler, 2003).

5 Concluding remarks

We introduce a model of sequential learning in which the set of alternatives the DM may explore is endogenously determined by her deliberate decision to search for additional options. The DM’s consideration set is therefore constructed sequentially in response to information she has learned over time. We characterize the DM’s optimal policy, and study how the tradeoff between focusing attention on alternatives within the DM’s consideration set and expanding it depends on the search technology. This tradeoff hinges a reservation price (“index”) corresponding to expansion of the consideration set. Using a recursive characterization of this reservation price and of the DM’s expected payoff under the optimal policy, we derive key properties of this reservation price and its implications for how the consideration set is optimally formed over time.

Our framework may be of interest for the study of dynamic decision processes in a variety of economic settings given the DM’s inability to consider all feasible alternatives from the beginning, whether due to limited attention or exogenous constraints on the environment.

References


**A Proofs**

**Proof of Proposition 1.** The environment is a special case of the one studied in Section 3.4. Condition 1 is satisfied in this environment since $M_\xi = 1$ for all $\xi \in \Xi$ and all uncertainty is resolved upon the first inspection. As a result, Theorem 2 applies. It remains to verify that the policy described in the statement of the Proposition coincides with $\chi^*$. Indeed, the reservation prices (1) are a special case of (13) in this setting, and since the search technology is deteriorating, $T^S$ is a special case of (1).

**Proof of Proposition 2.** Since product 0 corresponds to the outside option, one of the products is necessarily purchased. Let $i \neq j$ such that $w_i > w_j$. It remains to show that product $j$ will not be purchased.

Case 1: $j > i$. Suppose $w_j = T^S_j$, then $\min\{I_i, u_i\} \geq w_i > T^S_j$. The consumer reads $j$’s result only after clicking on $i$’s result. Once she clicks on $i$’s result, however, she will not read $j$’s result, since $u_i > T^S_j$. Next suppose $w_j = I_j$. Then $\min\{I_i, u_i\} \geq w_i > I_j$. The consumer therefore click on $j$’s result only after clicking on $i$’s. But again, once she has done so, she will not click on $j$’s result, since $u_i > I_j$. Now suppose $w_j = u_j$. Then even if she clicks on $j$’s result, since
If \( u_i \geq w_i > u_j \), she will either recall a previous product or continue to search and find a better product.

Case 2: \( j < i \). Since \( \mathcal{I}_i^S \geq w_i > w_j = \min\{\mathcal{I}_i, u_j, \mathcal{I}_j^S\} \), \( j \) cannot be purchased before \( i \)'s result is read. Since \( \mathcal{I}_i^S \leq \mathcal{I}_j^S \) and \( w_i > w_j \), it must be that \( \min\{\mathcal{I}_i, u_i\} \geq \min\{\mathcal{I}_i, u_i, \mathcal{I}_j^S\} = w_i > w_j = \min\{\mathcal{I}_j, u_j\} \). Arguments analogous to those in Case 1 establish that \( j \) cannot be purchased.

Therefore, in both cases, product \( j \) will not be purchased.

**Proof of Theorem 1: Steps 2 and 3.** Step 2. We use the result in Lemma 1 to characterize how much the DM obtains from following \( \chi^* \) from the outset rather than being forced to focus on a specific alternative in the first period and then reverting to \( \chi^* \) from the next period onward. The characterization permits to establish in Step 3 the optimality of \( \chi^* \) through dynamic programming.

Given the initial state \( S_0 \), for any \( \omega^P \in \{\mathcal{\omega}^P \in \Omega^P : S_0^P(\mathcal{\omega}^P) > 0\} \), denote by \( \mathbb{E} \left[ \mathcal{\nu} | \omega^P \right] \) the immediate expected payoff from focusing on an alternative in state \( \omega^P \) (the expectation is taken under the distribution \( H_{\omega^P} \)) and by \( \mathcal{\omega}^P \) the new state of that alternative. Let

\[
V^P(\omega^P | S_0) \equiv (1 - \delta) \mathbb{E} \left[ \mathcal{\nu} | \omega^P \right] + \delta \mathbb{E} \left[ \mathcal{V} \left( S_0 \setminus e(\omega^P) \right) | \omega^P \right]
\]

(16)
denote the payoff from starting with an alternative in state \( \omega^P \) and then following \( \chi^* \) from the next period onward. Similarly, let

\[
V^S(\omega^S | S_0) \equiv -(1 - \delta) \mathbb{E} \left[ \mathcal{c} | \omega^S \right] + \delta \mathbb{E} \left[ \mathcal{V} \left( S_0 \setminus e(\omega^S) \right) | \omega^S \right]
\]

(17)
denote the payoff from first searching, when the state of search is \( \omega^S \), and then following \( \chi^* \) from the next period onward, where \( \mathbb{E} \left[ \mathcal{c} | \omega^S \right] \) is the immediate expected cost incurred from searching, \( \mathcal{\omega}^S \) is the new state of search after search is carried out, and \( W^P(\mathcal{\omega}^S) \) is the state of the new alternatives obtained as the result of search, with \( \mathcal{c} \) and \( W^P(\mathcal{\omega}^S) \) jointly drawn from the distribution \( H_{\mathcal{\omega}^S} \).

We introduce a fictitious “retirement option” which is available at all periods and yields a constant reward \( M < \infty \) when chosen. Denote the state corresponding to this fictitious option by \( \omega^A_M \), and enlarge \( \Omega^P \) to include \( \omega^A_M \). Similarly, let \( e(\omega^A_M) \) denote the state of the problem when only the retirement option with fixed reward \( M \) is available. Since the payoff from the retirement option is constant at \( M \), if \( v \geq M \), then \( \kappa(v|S_0 \vee e(\omega^A_M)) = \kappa(v|S_0) \). If, instead, \( v < M \), then clearly \( \kappa(v|S_0 \vee e(\omega^A_M)) = \infty \). Hence, Lemma 1, adapted to the fictitious environment that

---

28 Note that \( W^P(\mathcal{\omega}^S) \) is a deterministic function of the new state \( \mathcal{\omega}^S = ((c_0, E_0), ..., (c_m, E_m)) \) of search. To see this, recall that \( E_m = (a_m(\xi) : \xi \in \Xi) \) is the list of the number of alternatives of each type introduced in the last search. The state \( W^P(\mathcal{\omega}^S) : \Omega \to \mathbb{N} \) is then obtained from \( E_m \) by letting \( W^P(\mathcal{\omega}^S)(\omega) = 0 \) for all \( \omega \in \Omega^S \cup \{(\xi, r^s) : s > 0 \text{ or } E_m(\xi) = 0\} \) and \( W^P(\mathcal{\omega}^S)(\omega) = E_m(\xi) \) for all states \( (\xi, r^s) \in \Omega^S \) such that \( E_m(\xi) > 0 \).
includes the retirement option, implies
\[
V(S_0 \lor e(\omega_M^A)) = \int_0^\infty \left( 1 - \mathbb{E} \delta^v(S_0 \lor e(\omega_M^A)) \right) dv
\]
\[
= M + \int_M^\infty \left( 1 - \mathbb{E} \delta^v(S_0) \right) dv.
\]

The definition of \( \chi^* \), along with Conditions (16) and (17), also implies the following.

**Lemma 2.** For any state of the search technology \( \omega^S \), state of an alternative \( \omega^P \), and payoff \( M \) of the retirement option:

\[
V(e(\omega^S) \lor e(\omega_M^A)) = \begin{cases} 
V^S(\omega^S | e(\omega^S) \lor e(\omega_M^A)) & \text{if } M \leq T^S(\omega^S) \\
M > V^S(\omega^S | e(\omega^S) \lor e(\omega_M^A)) & \text{if } M > T^S(\omega^S)
\end{cases}
\]

and similarly

\[
V(e(\omega^P) \lor e(\omega_M^A)) = \begin{cases} 
V^P(\omega | e(\omega^P) \lor e(\omega_M^A)) & \text{if } M \leq T^P(\omega^P) \\
M > V^P(\omega | e(\omega^P) \lor e(\omega_M^A)) & \text{if } M > T^P(\omega^P).
\end{cases}
\]

**Proof of Lemma 2.** First note that the index corresponding to the retirement option is equal to \( M \). Hence, if \( M \leq T^S(\omega^S) \), \( \chi^* \) will start by prescribing search. If, instead, \( M > T^S(\omega^S) \), \( \chi^* \) will opt for the retirement option forever, yielding an expected payoff of \( M \). To see why in this case \( M > V^S(\omega^S | e(\omega^S) \lor e(\omega_M^A)) \), observe that the payoff \( V^S(\omega^S | e(\omega^S) \lor e(\omega_M^A)) \) from starting with search and then following \( \chi^* \) in each subsequent period is

\[
V^S(\omega | e(\omega^S) \lor e(\omega_M^A)) = \mathbb{E} \left[ \sum_{s=0}^{\hat{\tau}-1} \delta^s U_s^\pi + \frac{\delta^\pi}{1-\delta} M | \omega^S \right],
\]

for some stopping and selection rules \( \hat{\tau}, \hat{\pi} \), since once the DM, under \( \chi^* \), opts for the retirement option, he will continue to do so in all subsequent periods. By definition of \( T^S(\omega^S) \),

\[
M > T^S(\omega^S) = \sup_{\pi, \tau} \mathbb{E} \left[ \sum_{s=0}^{\tau-1} \delta^s U_s^\pi | \omega^S \right] \geq \mathbb{E} \left[ \sum_{s=0}^{\hat{\tau}-1} \delta^s U_s^\pi | \omega^S \right],
\]

and rearranging,

\[
M \mathbb{E} \left[ \sum_{s=0}^{\hat{\tau}-1} \delta^s | \omega^S \right] > \mathbb{E} \left[ \sum_{s=0}^{\hat{\tau}-1} \delta^s U_s^\pi | \omega^S \right].
\]

34
Therefore,
\[
V^S(\omega|e(\omega^S) \lor e(\omega_A^S)) = \mathbb{E} \left[ \sum_{s=0}^{t-1} \delta^s U^*_s + \frac{\delta^t M}{1 - \delta} |\omega^S| \right] < \mathbb{E} \left[ M \sum_{s=0}^{t-1} \delta^s + \frac{\delta^t M}{1 - \delta} |\omega^S| \right] = M.
\]

Similar arguments establish Condition (20).

Next, for any initial state of \( S_0 \), and any state \( \omega^P \in \{ \hat{\omega}^P \in \Omega^P : S_0(\hat{\omega}^P) > 0 \} \) of the alternatives in the consideration set corresponding to \( S_0 \), let \( D_P(\omega^P|S_0) \equiv V(S_0) - V_P(\omega^P|S_0) \) denote the payoff differential between (i) starting with \( \chi^* \) right away and (ii) focusing first on an alternative in state \( \omega^P \) and then following \( \chi^* \) thereafter. Similarly, let \( D_S(\omega^S|S_0) \equiv V(S_0) - V^S(\omega^S|S_0) \) denote the payoff differential between (i) starting with \( \chi^* \) and (ii) starting with search in state \( \omega^S \) and then following \( \chi^* \). We now establish the following lemma.\(^{29}\)

**Lemma 3.** Let \( S_0 \) be the initial state, with \( \omega^S \in \Omega^S \) the state of the search technology as specified in \( S_0 \). Then

\[
D^S(\omega^S|S_0) = \int_0^{I^*(S^P_0)} D^S(\omega^S|e(\omega^S)) d\mathbb{E} \left[ \delta^v|S_0 \setminus e(\omega^S) \right].
\]

Similarly, for any alternative in the consideration set of type \( \omega^P \in \{ \hat{\omega}^P \in \Omega^P : S_0^P(\hat{\omega}^P) > 0 \} \),

\[
D_P(\omega^P|S_0) = \int_0^{\max(I^*(S^P_0 \setminus E(\omega^P)), I^S(\omega^S))} D_P(\omega^P|e(\omega^P)) d\mathbb{E} \left[ \delta^v|S_0 \setminus e(\omega^P) \right].
\]

**Proof of Lemma 3.** Given the state \( S_0 \lor e(\omega_A^S) \) and fixing \( \omega^S \in \Omega^S \), we have

\[
D^S(\omega^S|S_0 \lor e(\omega_A^S)) = V(S_0) + \int_0^M \mathbb{E} \left[ \delta^v|S_0 \right] dv + \mathbb{E} \left[ \tilde{c}|\omega^S \right]
\]

\[
- \delta \mathbb{E} \left[ V(S_0 \lor e(\omega^S) \lor e(\hat{\omega}^S) \lor W^P(\hat{\omega}^S)) + \int_0^M \mathbb{E} \left[ \delta^v|S_0 \lor e(\omega^S) \lor e(\hat{\omega}^S) \lor W^P(\hat{\omega}^S) \right] dv |\omega^S \right],
\]

where the equality follows from combining (17) with (18). Similarly,

\[
D^S(\omega^S|e(\omega^S)) = V(\omega^S) + \int_0^M \mathbb{E} \left[ \delta^v|\omega^S \right] dv + \mathbb{E} \left[ \tilde{c}|\omega^S \right]
\]

\[
- \delta \mathbb{E} \left[ V(e(\hat{\omega}^S) \lor W^P(\hat{\omega}^S)) + \int_0^M \mathbb{E} \left[ \delta^v|S_0 \lor e(\omega^S) \lor e(\hat{\omega}^S) \lor W^P(\hat{\omega}^S) \right] dv |\omega^S \right].
\]

Differentiating (23) and (24) with respect to \( M \), using the independence across alternatives

\(^{29}\)In the statement of the lemma, \( S_0 \setminus e(\omega^S) \) is the state of a fictitious problem in which search is not possible and \( S_0^P \setminus e(\omega^P) \) is the state of the consideration set obtained from \( S_0^P \) by subtracting an alternative in state \( \omega^P \).
and search, and the property that \( \kappa(v|S^1 \lor S^2) = \kappa(v|S^1) + \kappa(v|S^2) \), it is easily verified that

\[
\frac{\partial D^S(\omega^S|S_0 \lor e(\omega^A_M))}{\partial M} = \mathbb{E}\left[\delta^\kappa(M)|S_0 \lor e(\omega^S)\right] \frac{\partial D^S(\omega^S|e(\omega^S) \lor e(\omega^A_M))}{\partial M}.
\]

(25)

That is, the improvement in \( D^S(\omega^S|S_0 \lor e(\omega^A_M)) \) as a result of a slight increase in the retirement option \( M \) is the same as in a setting with only search and the retirement option, \( D^S(\omega^S|e(\omega^S) \lor e(\omega^A_M)) \), discounted by the expected time it takes (under \( \chi^* \)) until there are no indices with value strictly higher than \( M \), in an environment with only a consideration set in state \( S_0^P \) (where \( S_0^P \) is the same state of the consideration set as in \( S_0 \)). Similar arguments imply that, for any \( \omega^P \in \{\hat{\omega}^P \in \Omega^P : S_0(\hat{\omega}^P) > 0\} \),

\[
\frac{\partial D^P(\omega^P|S_0 \lor e(\omega^A_M))}{\partial M} = \mathbb{E}\left[\delta^\kappa(M)|S_0 \lor e(\omega^S)\right] \frac{\partial D^P(\omega^P|e(\omega^P) \lor e(\omega^A_M))}{\partial M}.
\]

(26)

Let \( M^* \equiv \max \{ I^*(S_0^P), I^S(\omega^S) \} \). Integrating (25) over the region \((0, M^*)\) of payoffs of the retirement option and rearranging, we have

\[
D^S(\omega^S|S_0 \lor e(\omega^A_M)) = D^S(\omega^S|S_0 \lor e(\omega^A_{M^*})) - \int_0^{M^*} \mathbb{E}\left[\delta^\kappa(v)|S_0 \lor e(\omega^S)\right] \frac{\partial D^S(\omega^S|e(\omega^S) \lor e(\omega^A_M))}{\partial v} dv
\]

\[
= D^S(\omega^S|S_0 \lor e(\omega^A_{M^*})) - \left[ \mathbb{E}\left[\delta^\kappa(v)|S_0 \lor e(\omega^S)\right] D^S(\omega^S|e(\omega^S) \lor e(\omega^A_M)) \right]_0^{M^*}
\]

\[
+ \int_0^{M^*} D^S(\omega^S|e(\omega^S) \lor e(\omega^A_M)) d\mathbb{E}\left[\delta^\kappa(v)|S_0 \lor e(\omega^S)\right]
\]

\[
= D^S(\omega^S|S_0 \lor e(\omega^A_{M^*})) - D^S(\omega^S|e(\omega^S) \lor e(\omega^A_M))
\]

\[
+ \int_0^{M^*} D^S(\omega^S|e(\omega^S) \lor e(\omega^A_M)) d\mathbb{E}\left[\delta^\kappa(v)|S_0 \lor e(\omega^S)\right]
\]

where the second equality follows from integration by parts and the third from the fact that \( \mathbb{E}\left[\kappa(M^*)|S_0 \lor e(\omega^S)\right] = 0 \) together with \( \mathbb{E}\left[\kappa(M^*)|S_0 \lor e(\omega^S)\right] = \infty \), as the DM can always receive her outside option. The outside option also implies that \( D^S(\omega^S|S_0 \lor e(\omega^A_M)) = D^S(\omega^S|S_0) \). It can also easily be verified that \( D^S(\omega^S|S_0 \lor e(\omega^A_{M^*})) = D^S(\omega^S|e(\omega^S) \lor e(\omega^A_{M^*})) \). Therefore, we have

\[
D^S(\omega^S|S_0) = \int_0^{M^*} D^S(\omega^S|e(\omega^S) \lor e(\omega^A_M)) d\mathbb{E}\left[\delta^\kappa(v)|S_0 \lor e(\omega^S)\right].
\]

(27)

Similar arguments yield

\[
D^P(\omega^P|S_0) = \int_0^{M^*} D^P(\omega^P|e(\omega^P) \lor e(\omega^A_M)) d\mathbb{E}\left[\delta^\kappa(v)|S_0 \lor e(\omega^P)\right].
\]

(28)

\[\text{This follows immediately from the observation that } V(S_0 \lor e(\omega^A_{M^*})) = V(e(\omega^S) \lor e(\omega^A_{M^*})) = M^* \text{, and similarly } \mathbb{E}\left[ V(S_0 \lor e(\omega^S) \lor e(\omega^S_{M^*}) \lor V(\omega^P) \lor e(\omega^A_{M^*})) | \omega^S \right] = \mathbb{E}\left[ V(e(\omega^S) \lor W^P(\omega^P) \lor e(\omega^A_{M^*})) | \omega^S \right]. \text{ Intuitively, any alternative with index strictly below } M^* \text{ will never be receive attention given the presence of a retirement option with payoff } M^*. \]
To complete the proof of Lemma 3, consider first the case where, given $S_0$, $\chi^*$ specifies focusing on an alternative (i.e., $M^* \neq I^S(\omega^S)$). Then Condition (21) in the lemma follows from (27) by noting that $M^* = I^\ast(S_0^P)$. Observe that, for any state $\omega^P \in \Omega^P$, if $\omega^P > \max\{I^\ast(S_0^P \setminus e(\omega^P)), I^S(\omega^S)\}$ then $M^* = I^P(\omega^P)$, in which case the integrand $D^P(\omega^P|e(\omega^P) \vee e(\omega^A))$ in (28) is equal to zero over the region $[0, I(\omega^P)]$, and hence $\max\{I^\ast(S_0^P \setminus e(\omega^P)), I^S(\omega^S)\}$. That is, in this case, Condition (22) clearly holds. Next, pick any state $\omega^P \in \Omega^P$ such that $I^P(\omega^P) < M^*$. Condition (22) follows directly from (28) by noting that, in this case, $M^* = \max\{I^\ast(S_0^P \setminus e(\omega^P)), I^S(\omega^S)\}$. Next, consider the case where, given $S_0$, $\chi^*$ specifies search (i.e., $M^* = I^S(\omega^S)$). Then, for any $\omega^P \in \Omega^P$, $\max\{I^\ast(S_0^P \setminus e(\omega^P)), I^S(\omega^S)\} = M^*$, in which case Condition (22) in the lemma follows directly from (28). That Condition (21) also holds follows from the fact that, in this case, $D^S(\omega^S|S_0) = 0$ and the integrand $D^S(\omega^S|e(\omega^S) \vee e(\omega^A))$ in (27) is equal to zero over the entire region $[0, I^S(\omega^S)]$. \hfill \Box

Step 3. Using Lemma 3, we can now directly verify that the average payoff under $\chi^*$ satisfies the dynamic programming equation, and hence is optimal. Let $V^\ast(S_0) \equiv (1-\delta)\sup_{\chi} E^X \left[ \sum_{s=0}^{\infty} \delta^s U_s | S_0 \right]$ denote the value function for the dynamic programming problem.

**Lemma 4.** For any state of the problem $S_0$, with $\omega^S$ denoting the state of the search technology as specified under $S_0$,

1. $V(S_0) \geq V^S(\omega^S|S_0)$, with the inequality holding as an equality if and only if $I^S(\omega^S) \geq I^\ast(S_0^P)$;

2. For any $\omega^P \in \{\hat{\omega}^P \in \Omega^P : S_0(\hat{\omega}^P) > 0\}$, $V(S_0) \geq V^P(\omega^P|S_0)$ with the inequality holding as an equality if and only if $I^P(\omega^P) = I^\ast(S_0^P) \geq I^S(\omega^S)$.

Hence, for any $S_0$, $V(S_0) = V^\ast(S_0)$, and $\chi^*$ is optimal.

**Proof of Lemma 4.** First, use (19) to note that the integrand in (21) is non-negative for all $0 \leq v \leq I^\ast(S_0^P)$, and that the entire integral in (21) is equal to zero if and only if $I^\ast(S_0^P) \leq I^S(\omega^S)$. This establishes Condition 1 in the lemma. Similarly, use (20) to observe that for any $\omega^P \in \{\hat{\omega}^P \in \Omega^P : S_0(\hat{\omega}^P) > 0\}$, the integrand in (22) is non-negative for any $0 \leq v \leq \max\{I^\ast(S_0^P \setminus e(\omega^P)), I^S(\omega^S)\}$, and that the entire integral in (22) is equal to zero if and only if $I^P(\omega^P) \geq \max\{I^\ast(S_0^P \setminus e(\omega^P)), I^S(\omega^S)\}$, which is the case if and only if $I^P(\omega^P) = I^\ast(S_0^P) \geq I^S(\omega^S)$. This establishes Condition 2 of the lemma.

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

$$V(S_0) = \max \left\{ V^S(\omega^S|S_0), \max_{\omega^P \in \{\hat{\omega}^P \in \Omega^P : S_0(\hat{\omega}^P) > 0\}} V^P(\omega^P|S_0) \right\}.$$ 

Hence $V$ solves the Bellman equation. That $\delta^T E^X [\sum_{s=T}^{\infty} \delta^s U_s | S] \rightarrow 0$ as $T \rightarrow \infty$ guarantees $V(S_0) = V^\ast(S_0)$, and hence the optimality of $\chi^*$.
This completes the proof of the theorem. ■

**Proof of Theorem 2.** To ease the notation, assume the initial consideration set is empty. It will be evident from the arguments below that the optimality of \( \chi^* \) does not hinge on this assumption.

Consider first an environment where \( M_{\xi} = 0 \) for all \( \xi \). It will become evident from the arguments below that the result easily extends to environments where \( M_{\xi} > 0 \), as well as to environments where \( M_{\xi} \) is stochastic, and gradually learned over time. Consider the following **auxiliary environment**, where all choices are reversible. Suppose that whenever an alternative of type \( \xi \) is brought into the consideration set, an additional auxiliary alternative is also introduced into the consideration set. Whenever the DM focuses on an auxiliary alternative, it yields a fixed flow payoff of \( R(\xi, (\theta'_0, 0)) \) where \( \theta'_0 \) is an arbitrary element of a set \( \Theta' \). Next, suppose that, whenever a non-auxiliary alternative in state \( \omega^P \) receives attention for the \( m \)-th time, a new auxiliary alternative is immediately introduced into the consideration set as well. Whenever pulled, this auxiliary option yields a fixed payoff of \( R(\omega^P) \), with the state \( \omega^P = (\xi, (\theta', m)) \) as defined in the main text. We say that an auxiliary alternative **corresponds to an (non-auxiliary) alternative** \( j \) if it has been introduced as a result of alternative \( j \) either being brought into the consideration set (through search) or receiving attention at some prior period.

In this auxiliary environment, define the index of search as in (7), with the rule \( \pi \) specifying whether to keep searching or explore one of the alternatives introduced through search, including the auxiliary alternatives brought in by the explorations of the physical alternatives introduced through search. For each alternative in state \( \omega^P \), define its new index as in (13), with the rule \( \pi \) in the definition of the index specifying for each period prior to stopping whether to focus on the alternative itself or to of the auxiliary alternatives introduced as the result of the alternative’s future explorations (importantly, \( \pi \) excludes any auxiliary alternative introduced in periods prior to the one in which the index is computed). Finally, let the index of any auxiliary alternative coincide with the alternative retirement payoff \( R \).

It should be clear that the same steps as in the proof of Theorem 1 imply that, in this auxiliary environment, the index policy based on the above new indices is optimal.\(^{31}\) We now interpret focusing attention on an auxiliary alternative corresponding to alternative \( j \) as irreversibly choosing alternative \( j \). Note that in this auxiliary environment, once the DM focuses attention on an auxiliary alternative, she continues to do so in all subsequent periods, since the index \( R(\omega^P) \) of the auxiliary alternatives does not change.

For the auxiliary environment to be formally equivalent to the primitive one (in the sense of generating the same dynamics and yielding the DM the same payoff) the following property must be true. For any alternative \( j \) and any period \( t > 0 \), if an auxiliary alternative corresponding to \( j \) is introduced in period \( t \), then under the optimal policy in the auxiliary environment, any other auxiliary alternative corresponding to alternative \( j \) introduced prior to period \( t \) is never selected.

\(^{31}\)The proof must be adjusted to allow for the auxiliary alternatives introduced as the result of the DM exploring the physical alternatives. Since all the steps are virtually the same, however, the proof is omitted.
That is, for each physical alternative, the “newest” auxiliary alternative corresponding to it must have the highest expected value $R$ among all other auxiliary alternatives corresponding to the same physical alternative. Condition 1 guarantees that this is the case. The same condition also guarantees that the policy $\pi$ in the definition of the index of the physical alternatives in the auxiliary problem coincides with the one in (13) where the selection $\pi$ was restricted to be over the exploration of the alternative under consideration and the retirement of the latter in its most recent state.

Finally, note that the proof immediately extends to settings in which $M^{\xi} > 0$ by assuming that, in the auxiliary environment described above, an auxiliary alternative is introduced into the consideration set only when its corresponding physical alternative has been explored more than $M^{\xi}$ times, with $M^{\xi}$ possibly stochastic and learned over time (in this latter case, the part of the state $\theta'$ other than the number of past exploration may also contain information about $M^{\xi}$).

Proof of Proposition 4. The result follows immediately from Theorem 2, noting that the perceived value of the alternative, $u$, given the current information about it, is precisely the expected flow value of choosing an alternative $R(\omega^P)$. The reservation prices (14) are a special case of (13) in this setting, and, since the search technology is deteriorating, (15) is a special case of (9).

Proof of Proposition 5. The proof follows arguments analogous to those in the proof of Proposition 2, and is therefore omitted.

39
In this appendix, we briefly illustrate, by means of an example, why bandit problems in which alternatives take the form of “meta arms”, i.e., sub-decision problems with their own sub-decisions, typically do not admit an index solution. This is so 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 an index solution. This is the case even if a subset of dependent alternatives evolves independently of all other alternatives, and even if one knows how to optimally choose among the dependent alternatives in each given subset in isolation.

Consider the following extension of the environment described in the paper. There are \(k \in \mathbb{N}\) sets of arms, \(K_1, \ldots, K_k\). Arms from different sets evolve independently of one another, but the state of each arm within a set may depend on the state of other arms from the same set. More generally, suppose that each arm corresponds to a “meta arm”, the activation of which involves decisions other than when to stop using it. Each meta arm has its own decision process which is independent of the other meta arms.

Clearly, the model in the main paper is a special case of this richer setting. Suppose that, for each set of arms \(K_i\), one can compute the optimal sequence of pulls, independently of the other sets of arms. Equivalently, suppose that for each “meta arm” one can compute the optimal sequence of decisions that define the usage of that arm, independently of the solution to the other meta arms’ problems. It is tempting to conjecture that one may then assign an independent index to each set of arms \(K_i\) (alternatively, to each “meta arm”) and that the optimal policy for the overall problem reduces to an index policy, whereby the meta arm with the highest index is selected in each period.

Perhaps surprisingly, the optimal policy for this enriched problem does not admit an index representation. When arms are not defined as in the canonical multi-armed bandit problem, but rather feature a more complicated internal decision problem (preserving the independence across arms), the optimal policy typically is not an index policy. The following example illustrates.

**Example 3.** There are two arms. Arm 1 yields a reward of 1000 when it is first pulled. In all subsequent pulls, it yields a reward \(\lambda\), where \(\lambda\) is initially unknown and may be either 1 or 10, with equal probability. After the first pull of arm 1, \(\lambda\) is perfectly revealed and is fully persistent. Arm 2 is a “meta arm” corresponding to the following decision problem. When the decision maker pulls arm 2 for the first time, she must also choose how to pull it. There are two ways to pull this arm, 2(A) and 2(B). If the decision maker selects 2(A), the arm yields a reward of 100 for a single period, followed by no rewards thereafter. If, instead, the decision maker selects 2(B), the arm yields a reward equal to 11 in each of its subsequent pulls. The choice of which version of arm 2 to use must be made the first time that arm 2 is pulled and can not be reversed.
Assume $\delta = 0.9$. The optimal policy for this problem is the following. In period 1, arm 1 is pulled. If $\lambda = 10$, then arm 2 in version 2(A) is then pulled for a single period, followed by arm 1 again in all subsequent periods. If, instead, $\lambda = 1$, arm 2 is then pulled in version 2(B) in all subsequent periods. Note that, under the optimal policy, the decision of how to use arm 2 depends on the realization of arm 1’s first pull. It is then evident that the optimal policy is not an index policy, no matter how one defines the indices. This is because an index policy requires that both the index of each arm and its utilization (when an arm can be used in different versions, as in the case of “meta arm” 2 in this example) be invariant in the results of the activation of all other arms.