Searching for "Arms": Experimentation with Endogenous Consideration Sets

Supplement

Daniel Fershtman **Alessandro Pavan**

July 2025

Abstract

This document contains additional material. All sections, conditions, and results specific to this document have the suffix "S" to avoid confusion with the corresponding parts in the main text. Section S.1 shows how the characterization of the optimal policy (Theorem 1 in the main text) extends to a broader class of decision problems where, in addition to learning about existing options and searching for new ones, the DM can irreversibly commit to one of the alternatives, bringing to an end the exploration process. As a special case, Section S.2 shows how to use Theorem 1 and Proposition 1 in the main body to arrive at a solution to the extension of Weitzman's Pandora's boxes problem in which the DM brings (stochastically) new boxes to the CS over time, as a function of previous explorations.

S.1 Decision problems with irreversible choice

Consider the following amendment to the general model of Section 2 in the main text. At any period t, in addition to exploring an alternative in the CS or expanding the latter, the DM can *irreversibly commit* to any alternative in the CS, provided that the alternative has been explored at least M_ξ times (with ξ denoting the alternative's category).[1] Once the DM irreversibly commits to an alternative, there are no further decisions to be made. Irreversibly committing to an alternative 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. In particular, denote by $R(\omega^P)$ the *expected flow value* from irreversibly committing to an alternative whose current state is $\omega^P = (\xi, \theta)$. Note that $R(\omega^P)$ admits two equivalent interpretations: (i) the DM obtains an immediate expected payoff equal to $R(\omega^P)/(1-\delta)$ after which there are no further payoffs; (ii) payoffs continue to accrue

[1] If $M_\xi = 0$, the DM can irreversibly commit to any ξ-alternative without first exploring it.

at all subsequent periods after the irreversible choice is made, with each expected flow payoff equal to $R(\omega^P)$.

For any $\omega^P = (\xi, \theta)$ and $\hat{\omega}^P = (\hat{\xi}, \hat{\theta})$, say that $\hat{\omega}^P$ "follows" ω^P if and only if $\hat{\xi} = \xi$, $\theta = (\vartheta_1, ..., \vartheta_m)$, for some m, and $\hat{\theta} = (\vartheta_1, ..., \vartheta_m, ..., \vartheta_{\hat{m}})$ for some $\hat{m} \geq m$. Denote this relation by $\hat{\omega}^P \succeq \omega^P$.

Condition 1. A category-ξ alternative has the *better-later-than-sooner property* if, for any $\omega^P = (\xi, \theta)$ such that $\theta = (\vartheta_1, ..., \vartheta_m)$, 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 S.1 (Weitzman's generalized problem). Consider the following extension of Weitzman's original problem: (i) The set of boxes is endogenous; (ii) each category-ξ box requires M_ξ explorations before the box's value is revealed; (iii) the DM can irreversibly commit (i.e., select) a box only if its value has been revealed, i.e., only after M_ξ explorations, where M_ξ can be stochastic; (iv) the flow payoff from exploring a box without committing to it is equal to the cost of exploring the box (with the latter evolving stochastically based on the number of past explorations) and is equal to zero for any exploration $t > M_\xi$; (v) the payoff $R(\omega^P)$ from irreversibly committing to a box whose value has been revealed (i.e., after the M_ξ-th exploration) remains constant after the M_ξ-th exploration and is equal to the box's prize.

Example S.2 (Purchase/Lease problem). In each period, an apartment owner either chooses one of the real-estate agents she knows 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 category ξ in state $\omega^P = (\xi, \theta)$ is a function of the information $\theta = (\vartheta_1, ..., \vartheta_m)$ 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, $R(\omega^P) \leq 0$. Else, for any $\theta = (\vartheta_1, ..., \vartheta_m)$ and $\hat{\theta} = (\vartheta_1, ..., \vartheta_m, ..., \vartheta_{\hat{m}})$ such that $\hat{m} \geq m$, $R(\xi, \hat{\theta}) \geq R(\xi, \theta)$. Contrary to Weitzman's generalized problem above, the DM may derive a higher (expected) value from using an alternative without irreversibly committing to it (i.e, from leasing instead of selling) for an arbitrary long, possibly infinite, number of periods.

To accommodate for irreversible choice, we need to modify the definition of the index of each alternative in state $\omega^P \in \Omega^P$ as follows:

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

where τ is a stopping time, and where π is a rule specifying whether the DM explores the alternative, or irreversibly commits to it. Similarly, modify the index of search $\mathcal{I}^S(\omega^S)$ by letting the rule π now specify not only whether the DM keeps searching or explores one of the alternatives brought to the CS through search, but also whether or not she irreversibly commits to one of the alternatives that the new search brought to the CS.

Next, amend the definition of the index policy χ^* as follows. At each period $t \geq 0$, given the state $\mathcal{S}_t$ of the decision problem, the policy specifies to (a) search if $\mathcal{I}^S$ is greater than the index $\mathcal{I}^P$ of any alternative in the CS and the expected "retirement" value R of each alternative in the CS; (b) experiment with an alternative in state ω^P if its index $\mathcal{I}^P$ is greater than its expected retirement value R, as well as the index of search, and both the index and the expected retirement value of any other alternative in the CS; (c) choose (i.e., irreversibly commit to) an alternative in state ω^P if its retirement value R is greater than its index $\mathcal{I}^P$, as well as the index of search and both the index and the expected retirement value of any other alternative in the CS.

We then have the following result:

Theorem S.1 (Indexability with irreversible choice)**.** *Suppose Condition 1 is satisfied for all $\xi \in \Xi$. The conclusions in Theorem 1 in the main text apply to the problem with irreversible choice under consideration. However, the stopping time τ^* in the characterization of the index of search is now the first time (strictly above the one at which the index is computed) at which $\mathcal{I}^S$, all the indexes of the alternatives brought to the CS by search, and all retirement values of such alternatives fall below the value $\mathcal{I}^S(\omega^S)$ of the search index when the latter is computed.*

The result is established by considering a fictitious problem without irreversible choice in which, each time the DM experiments with an alternative and changes its state to ω^P, an "auxiliary alternative" with constant flow payoff equal to $R(\omega^P)$ is added to the CS and remains available in all subsequent periods, irrespectively of possible changes in the state of the alternative that generated it. The better-later-than-sooner property of Condition 1 guarantees that, if the DM ever selects one of these auxiliary alternatives, she necessarily picks the one corresponding to the latest exploration of the alternative that generated it. This last property in turn implies that both (a) the non-perishability of the auxiliary alternatives and (b) the reversibility of choice in the fictitious problem play no role, which in turn implies that the optimal policy in the fictitious problem coincides with the one in the primitive problem.

Proof of Theorem S.1. To ease the notation, assume the initial CS is empty. It will be evident from the arguments below that the optimality of χ^* does not hinge on this assumption. Consider first an environment where $M_\xi = 0$ for all ξ. It will also become evident from the arguments below that the result easily extends to environments where $M_\xi > 0$, as well as to environments where M_ξ is stochastic and learned over time.

Consider the following *fictitious environment*, where all choices are *reversible*. Whenever an alternative of category ξ is brought to the CS, an additional *auxiliary* alternative is also introduced into the CS, yielding a fixed flow payoff of $R(\xi, \emptyset)$.[2] Furthermore, whenever a non-auxiliary alterna-

[2]Recall that $R(\xi, \emptyset)$ is the retirement value of a physical alternative of category ξ that has never been explored.

tive in state ω^P is explored, a new auxiliary alternative yielding a fixed payoff of $R(\tilde{\omega}^P)$ is also added to the CS, where $\tilde{\omega}^P$ denotes the new state of the explored alternative drawn from H_{ω^P}, as in the baseline model.[3] We say that an auxiliary alternative *corresponds to a (non-auxiliary) alternative in state* ω^P if it has been introduced to the CS as the result of either search (in which case $\theta = \emptyset$) or the exploration of an alternative in state ω^P. In this auxiliary environment, define the index of search as in the main text, with the rule π specifying whether to keep searching or exploring one of the alternatives introduced through search, including the auxiliary alternatives brought to the CS by search or by the explorations of the alternatives brought to the CS through search. For each alternative in state ω^P, define its new index as in (S.1), with the rule π in the definition of the index specifying for each period prior to stopping whether to explore the alternative itself or one of the auxiliary alternatives introduced as the result of the alternative's current and future explorations (i.e., following the period at which the index is computed; importantly, π 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's retirement value, as specified by the function R.

It is easy to see that the same steps as in the proof of Theorem 1 in the main text imply that, in this auxiliary environment, the index policy based on the above new indices is optimal.[4] It is also easy to see that the DM's problem in the auxiliary environment is a relaxation of the problem in the primitive environment in which (a) all decisions are reversible, and (b) alternatives can be retired also in states that are not feasible any more due to the subsequent explorations of the same alternative. Hereafter, we argue that the DM's payoff in the primitive environment under the proposed index policy is the same as under the corresponding index policy in the fictitious environment. To see this, first observe that, in the fictitious environment, once the DM explores an auxiliary alternative, she continues to do so in all subsequent periods, since the indexes $R(\omega^P)$ of the auxiliary alternatives do not change. This implies that the reversibility of choice in the fictitious environment plays no role. Next, observe that Condition 1 implies that, in the fictitious environment, if the DM selects an auxiliary alternative, she always picks the one corresponding to the "newest" state of the corresponding non-auxiliary alternative that created it; this is because the latest has the highest expected value R among all the auxiliary alternatives corresponding to the same non-auxiliary alternative. This implies that the non-perishability of the older versions of the auxiliary alternatives in the fictitious environment also plays no role. The same condition also guarantees that the policy π in the definition of the index of the non-auxiliary alternatives in the fictitious problem coincides with the one in (S.1) where the selection π is restricted to be over the exploration of the non-auxiliary 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

[3]If $M_\xi > 0$, the introduction of the auxiliary alternative as the result of the exploration of an alternative in state $\omega^P = (\xi, \theta)$ occurs only if $\theta = (\vartheta_1, ..., \vartheta_s)$ with $s \geq M_\xi$, that is, only if the alternative has been explored at least M_ξ times.

[4]The proof must be adjusted to accommodate for the auxiliary alternatives introduced as the result of the DM exploring the physical alternatives. Since all the steps are virtually the same, the proof is omitted.

that, in the fictitious environment, an auxiliary alternative is introduced into the CS only when its corresponding non-auxiliary alternative has been explored more than M_ξ times, with M_ξ possibly stochastic and learned over time (in this latter case, the time-varying component of an alternative's state, θ, may also contain information about M_ξ). ■

S.2 Pandora's boxes with an endogenous CS

As discussed in Section 4 of the main body, a special case of the setting described in the previous section arises when exploring an alternative immediately reveals its value, in which case the problem corresponds to the well-known Pandora's boxes problem, but here with an endogenous CS. Proposition S.1 below characterizes the optimal policy for this problem. The optimal policy prescribes when to search for an additional box, the order in which existing boxes should be opened, and when to stop and either recall an opened box or the outside option. The proof maps the Pandora's boxes problem with an endogenous set of boxes into an auxiliary problem that fits into the setting of Section 2 in the main body, and then uses Theorem 1 and Proposition 1 in the main body to identify the properties of the optimal policy in Proposition S.1 below.

Proposition S.1 (Pandora's-boxes with an endogenous set of boxes). *For any ξ-box that has not been opened yet (i.e., for which $\omega^P = (\xi, \emptyset)$ for some $\xi \in \Xi$) the reservation prize $\mathcal{I}^P(\xi, \emptyset)$ is given by the solution to:*

$$\mathcal{I}^P(\xi,\emptyset) = \frac{-\lambda^\xi + \delta \int_{\frac{\mathcal{I}^P(\xi,\emptyset)}{1-\delta}}^{\infty} v\, dF^\xi(v)}{1 + \frac{\delta}{1-\delta}\left(1 - F^\xi\left(\frac{\mathcal{I}^P(\xi,\emptyset)}{1-\delta}\right)\right)}. \tag{S.2}$$

For any $l \in \mathbb{R}$, let $\Xi(l) \equiv \left\{\xi \in \Xi : \mathcal{I}^P(\xi,\emptyset) > l\right\}$ denote the set of boxes whose reservation prize exceeds l. For any m, the reservation prize of search $\mathcal{I}^S(m)$ is given by the solution to:[5]

$$\mathcal{I}^S(m) = \frac{-c(m) + \delta \sum_{\xi\in\Xi(\mathcal{I}^S(m))} \rho^\xi(m)\left(-\lambda^\xi + \delta \int_{\frac{\mathcal{I}^S(m)}{1-\delta}}^{\infty} v\, dF^\xi(v)\right)}{1 + \sum_{\xi\in\Xi(\mathcal{I}^S(m))} \rho^\xi(m)\left[\delta + \frac{\delta^2}{1-\delta}\left(1 - F^\xi\left(\frac{\mathcal{I}^S(m)}{1-\delta}\right)\right)\right]}. \tag{S.3}$$

The solution to Pandora's-boxes problem with an endogenous CS takes the following form:

1. *If the highest reservation prize among all unopened boxes in the CS is greater than the reservation prize $\mathcal{I}^S(m)$ of search, and is greater than the flow value $(1-\delta)v$ of each opened box and the outside-option, the DM opens one of the boxes with the highest reservation prize.*

2. *If the reservation prize of search $\mathcal{I}^S(m)$ is higher than the reservation prize $\mathcal{I}^P(\xi,\emptyset)$ of any unopened box and of the flow value $(1-\delta)v$ of each opened box and the outside-option, the DM searches.*

3. *If neither of the above two situations applies, the DM stops. He then takes the prize of one of*

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

the opened boxes whose flow value $(1-\delta)v$ is the highest among the opened boxes if the latter value exceeds the outside-option, and takes the outside-option otherwise.

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

To see that the index of a ξ-box that has not been opened yet is given by (S.2), note that the index of an opened box is equal to $(1-\delta)v$. Because the optimal stopping time τ^* in the definition

$$\mathcal{I}^P(\xi, \emptyset) \equiv \sup_{\tau>0} \frac{\mathbb{E}\left[\sum_{s=0}^{\tau-1} \delta^s u_s | \xi, \emptyset\right]}{\mathbb{E}\left[\sum_{s=0}^{\tau-1} \delta^s | \xi, \emptyset\right]}, \tag{S.4}$$

of the index $\mathcal{I}^P(\xi, \emptyset)$ is the first time at which the value of the index drops below its value $\mathcal{I}^P(\xi, \emptyset)$ at the time the index is computed, we then have that $\tau^* = 1$ if $(1-\delta)v \le \mathcal{I}^P(\xi, \emptyset)$ and $\tau^* = \infty$ otherwise.

Turning to the index for search, the combination of the assumption that $c(m)$ is weakly increasing in m with the assumption that the distribution $\rho(m) \in \Delta(\Xi)$ from which the boxes are drawn "decreases" with m in a FOSD sense implies that the optimal stopping-time τ^* in

$$\mathcal{I}^S(m) \equiv \sup_{\pi,\tau} \frac{\mathbb{E}^\pi\left[\sum_{s=0}^{\tau-1} \delta^s U_s | m\right]}{\mathbb{E}^\pi\left[\sum_{s=0}^{\tau-1} \delta^s | m\right]}, \tag{S.5}$$

is equal to (a) $\tau^* = \infty$ if the box identified at the m-th search has a reservation prize $\mathcal{I}^P(\xi, \emptyset) > \mathcal{I}^S(m)$ and its realized flow payoff satisfies $v(1-\delta) > \mathcal{I}^S(m)$, (b) $\tau^* = 1$ if $\mathcal{I}^P(\xi, \emptyset) \le \mathcal{I}^S(m)$, and (c) $\tau^* = 2$ if $\mathcal{I}^P(\xi, \emptyset) > \mathcal{I}^S(m)$ and $v(1-\delta) \le \mathcal{I}^S(m)$. ■