Searching for “Arms”: Experimentation with Endogenous Consideration Sets

Online Supplementary Material

Daniel Fershtman        Alessandro Pavan

January 2022

Abstract

This document contains the formal analysis of the comparative statics for the applications described in the main text, as well as a few additional results. 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 contains the comparative statics analysis for the application to the administration of medical treatments. Section S.2 contains the comparative statics analysis for the clinical-trials application, whereas Section S.3 contains the comparative statics analysis for the online-consumer-search application. Section S.4 shows how the characterization of the optimal policy extends to certain problems with irreversible choice. Finally, Section S.5 discusses why an index policy need not be optimal in the presence of “meta” arms corresponding to independent sub-problems.

S.1 Comparative statics - administration of medical treatments

In this section, we use the characterization of the physician’s optimal policy in Section 4.1 in the main text to illustrate some of the novel comparative statics that may arise when accounting for the endogeneity of the CS. Consider an improvement in a category of treatments. Such an improvement may occur because of technological advancements, changes in pricing, or changes in the incentive schemes offered to the physician for the administration of the various treatments. With an exogenous CS, an improvement in a category of treatments always leads to an increase in their administration at the expense of other categories. While this may happen also with an endogenous CS, the opposite may also occur, as the following result shows.

Corollary S.1. Consider the environment in Subsections 2.1 and 4.1 in the main text. The ex-ante expected discounted number of times the $\alpha$-treatments are administered may decrease both with the ex-ante probability $p^\alpha(\emptyset)$ such treatments are effective and with the value $v^\alpha$ the physician assigns to good outcomes delivered through such treatments.
Proof of Corollary S.1. We first identify conditions under which the exploration and expansion dynamics take a particularly simple form. We then compute the ex-ante expected discounted number of times an α-treatment is administered prior to the improvement and after. Finally, we show that the administration of the α-treatments may be higher before the improvement. The conditions below are sufficient but not necessary for the result. While a complete characterization of the conditions for which the result holds is not easily attainable, the conditions below make clear that the result is not knife-edge.

Suppose that initially there are two treatments in the physician’s CS, one of each category, and suppose that the α-treatments improve. Such an improvement may take the form of (1) an increase in the ex-ante probability that each α-treatment is effective from \( p^\alpha(\emptyset) \) to \( \hat{p}^\alpha(\emptyset) = p^\alpha(\emptyset) + \varepsilon_p, \varepsilon_p > 0 \), or (2) an increase in the payoff the physician derives from a good outcome delivered through the administration of an α-treatment from \( v^\alpha \) to \( \hat{v}^\alpha = v^\alpha + \varepsilon_v \), with \( \varepsilon_v \geq 0 \). Let \( \Lambda^\xi(\emptyset) \equiv p^\xi(\emptyset)q_1^\xi \) denote the posterior probability that a treatment yields a good outcome. Note that the posterior belief \( \Lambda^\xi \) is effective and the index of a ξ-treatment that has never been tested, respectively. Likewise, we let \( \Lambda(\emptyset) \equiv p(\emptyset)q_1 \). We continue to denote by \( \hat{p}^\xi(\emptyset) \) and \( \hat{P}^\xi(\xi,\emptyset) \) the prior belief a ξ-treatment is effective and the index of a ξ-treatment that has never been tested, respectively.

Suppose that the following order applies

\[
\mathcal{I}^P(\alpha, \emptyset) > \mathcal{I}^P(\beta, \emptyset) > \mathcal{I}^P(\alpha, 1)
\]

\[
-1 - \delta^2 \{ \rho^\omega \delta^2 \Lambda^\omega(\emptyset) + q_1^\omega \Lambda^\omega(\emptyset) \}
\]

\[
\max\{\mathcal{I}^P(\beta, 1), \mathcal{I}^P(\alpha, 2)\}.
\]

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

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

The result follows from part (ii) of Theorem 1, along with Corollary 5, in the main text. In particular, the order in (S.1) implies that the optimal stopping time in the formula for the search index in Condition (6) in the main text is equal to: (a) \( \tau^\xi = \infty \) if either the new treatment is an α-treatment and a good outcome is observed in one of the treatment’s first two administrations, or the new treatment is a β-treatment and a good outcome is observed after the treatment’s first administration; (b) \( \tau^\xi = 2 \) if the new treatment is a β-treatment and the outcome of its first administration is bad; (c) \( \tau^\xi = 3 \) if the new treatment is an α-treatment and each of its first two
administrations yielded a bad outcome.¹

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

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

> \max \{ \hat{\mathcal{I}}^P(\alpha, 1), \mathcal{I}^P(\beta, 1) \},

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

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

The result follows again from part (ii) of Theorem 1 along with Corollary 5 in the main text. Given (S.2), the optimal stopping time in the definition of the index of search is

\[
\tau^{\xi^*} = \begin{cases} 2 & \text{if the new treatment brought to the CS by search yields a bad outcome after its first administration,} \\
\infty & \text{otherwise.} \end{cases}
\]

Next, compare the ex-ante expected discounted number of times the physician administers an α-treatment before the improvement and after. The ordering in (S.1) implies that, before the improvement, the physician starts by administering the α-treatment in the CS. If such a treatment yields a bad outcome, she then administers the β-treatment in the CS. If the latter also yields a bad outcome, the physician administers again the α-treatment in the CS that yielded the initial bad outcome. If this latter treatment yields a second bad outcome, the physician then searches for new treatments. If, at any point, the administered treatment yields a good outcome, because the treatment is revealed effective, the physician then administers it in all subsequent periods, thus bringing the experimentation de facto to a halt.

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

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

Solving for \( A_S \), we have that

\[
\frac{\rho^\alpha}{1-\delta^2} \frac{1-\delta^2 + \delta^2 \Lambda^\alpha(\emptyset)(2-q_1^\alpha)}{\rho^\alpha(1-\Lambda^\alpha(\emptyset)(2-q_1^\alpha))\delta + \rho^\beta(1-\Lambda^\beta(\emptyset))}.\]

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

¹Recall that search itself occupies one period.
administered is therefore equal to
\[
A = 1 + \frac{\delta \Lambda^\alpha (\emptyset)}{1 - \delta} + (1 - \Lambda^\alpha (\emptyset))(1 - \Lambda^\beta (\emptyset))\delta^2 \left[ 1 + \frac{\delta \Lambda^\alpha (1)}{1 - \delta} + (1 - \Lambda^\alpha (1))\delta^2 A_S \right],
\]
where \( \Lambda^\xi (1) = p^\xi (1)q^\xi_1 = (1 - q^\xi_1)\Lambda^\xi (\emptyset) / (1 - \Lambda^\xi (\emptyset)) \) is the probability of a good outcome from a \( \xi \)-treatment that yielded a bad outcome at its first administration.

Now let \( \hat{A}_S \) and \( \hat{A} \) be the analogs of \( A_S \) and \( A \), respectively, after the improvement in the \( \alpha \)-treatments. Under the order in (S.2), the physician first administers the \( \alpha \)-treatment in the CS. If the latter yields a bad outcome, the physician then administers the \( \beta \)-treatment in the CS. If the latter also yields a bad outcome, the physician then searches for new treatments.\(^2\) Then
\[
\hat{A}_S = \rho^\alpha \left( 1 + \frac{\delta \hat{\Lambda}^\alpha (\emptyset)}{1 - \delta} + \delta^2 \hat{A}_S (1 - \hat{\Lambda}^\alpha (\emptyset)) \right) + \rho^\beta (1 - \Lambda^\beta (\emptyset)) \delta^2 \hat{A}_S.
\]
Solving for \( \hat{A}_S \), we have that
\[
\hat{A}_S = \frac{\rho^\alpha \left( 1 + \frac{\delta \hat{\Lambda}^\alpha (\emptyset)}{1 - \delta} \right)}{1 - \delta^2 + \delta^2 \left( \rho^\alpha \hat{\Lambda}^\alpha (\emptyset) + \rho^\beta \Lambda^\beta (\emptyset) \right)}.
\]
Therefore, the ex-ante expected discounted number of times an \( \alpha \)-treatment is administered when the \( \alpha \)-treatments are improved is equal to
\[
\hat{A} = 1 + \frac{\delta \hat{\Lambda}^\alpha (\emptyset)}{1 - \delta} + (1 - \hat{\Lambda}^\alpha (\emptyset))(1 - \Lambda^\beta (\emptyset))\delta^3 \hat{A}_S.
\]
It is easily verified that Conditions (S.1) and (S.2) are consistent with
\[
A > \hat{A} \quad \text{(S.3)}
\]
over an open set of parameter values such that \( \varepsilon_p \geq 0 \) and \( \varepsilon_v \geq 0 \), with at least one inequality strict.

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

Interestingly, this phenomenon is not monotone in the parameters of the model (and, as a result,

\(^2\)Again, if at any point a treatment yields a good outcome, because it is revealed effective, it is then administered in each subsequent period. Furthermore, because the search technology is stationary, all treatments in the current CS are effectively discarded when search is carried out.
does not involve extreme parameter values). To see why this is the case, consider the role of $\rho^\beta$, the probability that search brings a $\beta$-treatment. When $\rho^\beta$ is low, because search delivers primarily $\alpha$-treatments, the changes in the composition of the CS due to search being carried out more often contribute to a boost in the administration of the $\alpha$-treatments. In this case, there is no crowding out. When, instead, $\rho^\beta$ is high, search is unlikely to bring $\alpha$-treatments. As a result, improvements in the $\alpha$-treatments have a small impact on the search index as the latter is driven primarily by the properties of the $\beta$-treatments (this is because the search index “averages” over the properties of the two categories). In this case, the improvement in the search index is small compared to the improvement in the indexes of the $\alpha$-treatments that delivered bad outcomes. The instances where search is carried out instead of further exploration of the $\alpha$-treatments are then rare. The direct effect of the improvement in the attractiveness of the $\alpha$-treatments then prevails over the indirect effect of the improvement in the attractiveness of search. Hence, in this case, too, there is no crowding out. The phenomenon in the corollary thus occurs only for intermediate values of $\rho^\beta$. Similar non-monotonicities apply to the other parameters of the model.

S.2 Comparative statics - clinical trials towards regulatory approval

In this section, we consider the effects of changes in approval standards on the approval of products from different categories. In particular, consider the setting in Subsection 4.2 in the main text and suppose that the regulator lowers the approval standard for one of the products’ category, possibly with the intent of favoring the approval of products from that category, while keeping the standards for the other categories fixed. As we show below, such regulatory changes are not guaranteed to deliver the desired results, when accounting for their effects on the firm’s incentives to invest in basic research. For simplicity, suppose that there are two products’ categories, $\alpha$ and $\beta$.

Corollary S.2. A reduction in the approval standard for the $\alpha$-products from $\Psi^\alpha$ to $\Psi^\alpha - \epsilon$, $\epsilon > 0$, may reduce the ex-ante probability that an $\alpha$-product is approved.

Proof of Corollary S.2. For simplicity, suppose that the firm’s initial CS contains only two products, one of each category. Further assume that $q^\xi_1 = q^\xi_0 = q^\xi$, $\xi = \alpha, \beta$, meaning that the probability that safe products deliver good outcomes is the same as the probability that unsafe products deliver bad outcomes, in which case the posterior belief $p^\xi(\theta)$ that a $\xi$-product is safe depends on the history $\theta$ of past trials only through the difference between good and bad outcomes. Denote by $D(\theta)$ the difference between good and bad outcomes recorded in $\theta$.

We first identify conditions under which the exploration and expansions dynamics take a particularly simple form, both under the original regulatory regime and the new one. We then compute the ex-ante probability an $\alpha$-product is approved under each of the two regimes. Finally, we show that reducing the approval standard for the $\alpha$-products without changing the approval standards for the $\beta$-products may reduce the ex-ante probability the $\alpha$-products are approved. Once again,
the conditions below are sufficient but not necessary for the result but reveal that the result is not
knife-edge.

For simplicity, suppose that initially the standards \((\Psi^\alpha, \Psi^\beta)\) are such that \(p^\alpha(\theta) \geq \Psi^\alpha\) if and
only if \(D(\theta) \geq 2\) and \(p^\beta(\theta) \geq \Psi^\beta\) if and only if \(D(\theta) \geq 1\). Furthermore, assume that, for any
\(\xi \in \{\alpha, \beta\}\), \(\lambda^\xi(\theta) > \nu^\xi\) for each \(\theta\) containing more than four elements, meaning that it is not worth
experimenting with each product more than four times.\(^3\) Finally, suppose that \(\lambda^\beta(\theta) = 0\) for all \(\theta\) of
dimensionality smaller than four. For \(\xi = \alpha\), instead, \(\lambda^\alpha(\theta) = 0\) for all \(\theta\) of dimensionality smaller
than four such that \(p^\alpha(\theta) \geq p^\alpha(\theta)\), whereas, for any \(\theta\) of dimensionality smaller than four such that
\(p^\alpha(\theta) < p^\alpha(\theta)\), \(\lambda^\alpha(\theta) = \lambda^\alpha\), for some \(\lambda^\alpha \in \mathbb{R}_+\). These assumptions are not crucial for the results,
butsimply the calculations below. Given these assumptions, hereafter we will confine attention to histories \(\theta\) that contain no more than 4 elements and no more than 1 bad outcome.

Denote by \(\xi^\alpha(p^\xi) \equiv p^\xi(1 - q^\xi) + (1 - p^\xi)q^\xi\), \(\xi^\beta(p^\xi) \equiv p^\xi q^\xi + (1 - p^\xi)(1 - q^\xi)\), \(\xi^\xi_{GG}(p^\xi) \equiv
p^\xi(q^\xi)^2 + (1 - p^\xi)(1 - q^\xi)^2\), and \(\xi^\xi_B(p^\xi) \equiv p^\xi(1 - q^\xi)^2 + (1 - p^\xi)(q^\xi)^2\) the probabilities that (starting
with belief \(p^\xi\) about the product’s safety) the first outcome is bad, the first outcome is good, the first two outcomes are
good, and the first two outcomes are bad, respectively.

The index of an \(\alpha\)-product for which \(\theta \in \{\emptyset, (B, G), (G, B)\}\) is equal to
\[
I^p(\alpha, \theta) = I^p(\alpha, (G, B)) = I^p(\alpha, (B, G)) = \frac{(1 - \delta)\delta^2u^\alpha\xi^\alpha_{GG}(p^\alpha(\theta))}{1 - \delta\xi^\beta_B(p^\alpha(\theta)) + \delta^2q^\alpha(1 - q^\alpha)}.
\]
To see this, recall that, by Theorem 1 in the main text, the optimal stopping time in the definition
of an index is the first time at which the value of the index drops weakly below its initial value. In
particular, at each of the three histories in (S.4), the optimal stopping time \(\tau^{\alpha*}\) for the \(\alpha\)-products satisfies:
\(\tau^{\alpha*} = 1\) if the next outcome is bad; \(\tau^{\alpha*} = 2\) if the next two outcomes are \((G, B)\), that is,
first a good outcome and then a bad one; and \(\tau^{\alpha*} = \infty\) if the next two outcomes are both good.
Similarly,
\[
I^p(\alpha, G) = I^p(\alpha, (G, B, G)) = I^p(\alpha, (B, G, G)) = \frac{(1 - \delta)\delta u^\alpha\xi^\alpha_{GG}(p^\alpha(G))}{1 - \delta\xi^\beta_B(p^\alpha(G))},
\]
where \(p^\xi(G) = p^\alpha(\theta)q^\xi/\xi^\xi_G(p^\xi(\theta))\) is the posterior probability that the product is safe when \(\theta\)
contains a single good outcome. This is because, at each of the three histories in (S.5), the optimal
stopping time \(\tau^{\alpha*}\) for the \(\alpha\)-products is \(\tau^{\alpha*} = 1\) if the next outcome is bad and \(\tau^{\alpha*} = \infty\) otherwise.
For an \(\alpha\)-product whose first realization is bad, instead,
\[
I^p(\alpha, B) = \frac{-\lambda^\alpha(1 - \delta) - (1 - \delta)\delta u^\alpha \left[p^\alpha(B)(q^\alpha)^3 + (1 - p^\alpha(B))(1 - q^\alpha)^3\right]}{1 - \delta\xi^\beta_B(p^\alpha(B)) + \delta q^\alpha(1 - q^\alpha)\left[1 + \delta\xi^\xi_B(p^\alpha(B))\right]},
\]
where \(p^\xi(B) = p^\xi(\theta)(1 - q^\xi)/\xi^\xi_B(p^\xi(\theta))\) is the posterior probability that the product is safe when \(\theta\)
contains a single bad outcome. This is because, when \(\theta = B\), the optimal stopping time for the
\(\alpha\)-products is: \(\tau^{\alpha*} = 1\) if the next outcome is bad; \(\tau^{\alpha*} = 2\) if the next two outcomes are \((G, B)\);
\(^3\)Recall that \(\lambda^\xi(\theta)\) is the cost of experimenting with a \(\xi\)-product when the history of past experimentation outcomes
is \(\theta\).
\( \tau^{\alpha*} = 3 \) if the next three outcomes are \((G, G, B)\); and \( \tau^{\alpha*} = \infty \) if the next three outcomes are \((G, G, G)\).

Next consider the \( \beta \)-products. First, observe that

\[
I^P(\beta, \emptyset) = I^P(\beta, (B, G)) = \frac{(1 - \delta) \delta v^2 \delta_S^\beta G(G^\beta(\emptyset))}{1 - \delta_S^\beta (p^\beta(\emptyset)) \delta}.
\]

(S.6)

This is because, at the histories in (S.6), the optimal stopping time \( \tau^{\beta*} \) for the \( \beta \)-products is \( \tau^{\beta*} = 1 \) if the next outcome is bad and \( \tau^{\beta*} = \infty \) otherwise. Similarly,

\[
I^P(\beta, B) = \frac{(1 - \delta) \delta \delta_S^\beta G(G^\beta(B))}{1 - \delta \left[ \delta_S^\beta(p^\beta(B)) + \delta (1 - q^\beta) q^\beta \right]}.
\]

This is because, when \( \theta = B \), the optimal stopping time for the \( \beta \)-products satisfies: \( \tau^{\beta*} = 1 \) if the next outcome is bad; \( \tau^{\beta*} = 2 \) if the next two outcomes are \((G, B)\); and \( \tau^{\beta*} = \infty \) if the next two outcomes are \((G, G)\).

Finally, note that the index of any \( \xi \)-product whose history \( \theta \) includes two or more bad outcomes is nonpositive, as is that of any product for which \( \theta \) contains more than four elements.

Now suppose that the ordering of the indexes prior to the relaxation of the standard for approval for the \( \alpha \)-products satisfies

\[
I^P(\alpha, \emptyset) = I^P(\alpha, (G, B)) = I^P(\alpha, (B, G)) > I^P(\beta, \emptyset) = I^P(\beta, (B, G)) > I^P(\alpha, B)
\]

\[
> \frac{1 - \delta}{1 - \delta^2} \left\{ \rho^\alpha \left( \nu^\alpha \delta_S^\alpha G_G(p^\alpha(\emptyset)) \delta \left( 1 + 2 \delta^2 (1 - q^\alpha) q^\alpha \right) \right) - \lambda^\alpha \delta_S^\alpha (p^\alpha(\emptyset)) + \rho^\alpha \delta \delta_S^\alpha G_G(p^\alpha(\emptyset)) \right\}
\]

\[
> I^P(\beta, B).
\]

(S.7)

Then, from part (ii) of Theorem 1 and Corollary 6 in the main text, we have that

\[
I^S = \frac{1 - \delta}{1 - \delta^2} \left\{ \rho^\alpha \left( \nu^\alpha \delta_S^\alpha G_G(p^\alpha(\emptyset)) \delta \left( 1 + 2 \delta^2 (1 - q^\alpha) q^\alpha \right) \right) - \lambda^\alpha \delta_S^\alpha (p^\alpha(\emptyset)) + \rho^\alpha \delta \delta_S^\alpha G_G(p^\alpha(\emptyset)) \right\}
\]

This is because, if the new product brought to the CS by search is an \( \alpha \)-product, then the optimal stopping time in the definition of the search index is given by: \( \tau^{\alpha*} = \infty \) after the outcome histories \((G, G)\), \((B, G, G, G)\), and \((G, B, G, G)\); \( \tau^{\alpha*} = 3 \) after the history \((B, B)\); \( \tau^{\alpha*} = 4 \) after the histories \((B, G, B)\) and \((G, B, B)\); and \( \tau^{\alpha*} = 5 \) after the histories \((B, G, G, B)\) and \((G, B, G, B)\). If, instead, the product identified by search is a \( \beta \)-product, then the optimal stopping time in the definition of the search index is \( \tau^{\beta*} = \infty \) if the first outcome is good and \( \tau^{\beta*} = 2 \) if the first outcome is bad.

Now suppose that the regulator relaxes the approval standard for all the \( \alpha \)-products from \( \Psi^\alpha \) to \( \tilde{\Psi}^\alpha = \Psi^\alpha - \varepsilon \), with \( \varepsilon > 0 \). Suppose that, as a result of the change, \( p^\alpha(\emptyset) \geq \tilde{\Psi}^\alpha \) if and only if \( D(\emptyset) \geq 1 \). Denote by \( I^P(\alpha, \theta) \) the indexes of the \( \alpha \)-products, after the reduction in the standard of
approval. Then observe that

$$\hat{I}_P(\alpha, \emptyset) = \hat{I}_P(\alpha, (B, G)) = \frac{(1 - \delta)\delta v^\alpha \varsigma_G^\alpha(p^\alpha(\emptyset))}{1 - \varsigma_B^\alpha(p^\alpha(B))}. \quad (S.8)$$

This is because the optimal stopping time for the $\alpha$-products at the histories in (S.8) is now given by $\tau^{\alpha*} = 1$ if the next outcome is bad and $\tau^{\alpha*} = \infty$ otherwise. Similarly,

$$\hat{I}_P(\alpha, B) = \left(1 - \delta\right) \frac{\left[-\lambda^\alpha + \delta^2 v^\alpha \varsigma_G^\alpha(p^\alpha(B))\right]}{1 - \delta \varsigma_B^\alpha(p^\alpha(B)) + \delta(1 - q^\alpha)q^\alpha}. \quad (S.8)$$

This is because the optimal stopping time for the $\alpha$-products after one bad outcome is now given by: \(\tau^{\alpha*} = 1\) if the next outcome is also bad; \(\tau^{\alpha*} = 2\) if the next two outcomes are \((G, B)\); and \(\tau^{\alpha*} = \infty\) if the next two outcomes are \((G, G)\).

Now suppose that \(\hat{\Psi}^\alpha\) is such that the following order of the indexes holds:

$$\hat{I}_P(\alpha, \emptyset) = \hat{I}_P(\alpha, (B, G)) > \hat{I}_P(\beta, \emptyset) = \hat{I}_P(\beta, (B, G))$$

$$\frac{(1 - \delta)\left[-\lambda^\alpha + \delta^2 v^\alpha \varsigma_G^\alpha(p^\alpha(B))\right]}{1 - \delta \varsigma_B^\alpha(p^\alpha(B)) + \delta(1 - q^\alpha)q^\alpha} > \max \left\{\hat{I}_P(\alpha, B), \hat{I}_P(\beta, B)\right\}. \quad (S.9)$$

Given the ordering in (S.9), the search index after the change in the standard of approval for the $\alpha$-products is then given by

$$\hat{I}_S = \frac{(1 - \delta)\left[-c + \delta^2 \sum_{\xi \in \{\alpha, \beta\}} \rho^\xi \varsigma_G^\xi(p^\xi(\emptyset))\right]}{1 - \delta^2 \sum_{\xi \in \{\alpha, \beta\}} \rho^\xi \varsigma_B^\xi(p^\xi(\emptyset))}. \quad (S.9)$$

This follows directly from part (ii) of Theorem 1 along with Corollary 6 in the main text (in particular, the optimal stopping in the definition of the search index is now equal to $\infty$ if the first outcome is good and is equal to 2 if it is bad, irrespective of the type of product brought to the CS).

Given the results above, we now compare the ex-ante probabilities of approving one of the $\alpha$-products before the relaxation of the standard and after.

**Selection of $\alpha$-products before reduction in approval standard.** The order of experimentation is given by (S.7). Because the search technology is stationary, by virtue of Corollary 5 in the main text, when search is launched, all products in the CS are discarded. Denote by $A_S$ the probability with which one of the $\alpha$-products is approved in the continuation immediately following search. We then have that

$$A_S = \rho^\alpha \left\{\varsigma_G^\alpha(p^\alpha(\emptyset)) (1 + 2(1 - q^\alpha)q^\alpha) + [1 - \varsigma_G^\alpha(p^\alpha(\emptyset)) (1 + 2(1 - q^\alpha)q^\alpha)] A_S \right\} + \rho^\beta \varsigma_B^\beta(p^\beta(\emptyset)) A_S.$$

This is because an $\alpha$-product is approved after the histories $(G, G), (G, B, G, G), (B, G, G, G)$, whereas
a $\beta$-product is approved if and only if its first outcome is good. Rearranging, we have that

$$A_S = \frac{\rho^\alpha \zeta^{\alpha}_G(p^\alpha(\emptyset)) [1 + 2(1 - q^\alpha)q^\alpha]}{1 - \left\{ \rho^\alpha [1 - \zeta^{\alpha}_G(p^\alpha(\emptyset)) (1 + 2(1 - q^\alpha)q^\alpha)] + \rho^\beta \zeta^{\beta}_B(p^\beta(\emptyset)) \right\}}.$$  

Therefore, the ex-ante probability one of the $\alpha$-products is approved is equal to

$$A = \zeta^{\alpha}_G(p^\alpha(\emptyset)) \left[ 1 + (1 - q^\alpha)q^\alpha \left( 1 + \zeta^{\beta}_B(p^\beta(\emptyset)) \right) \right] + \left[ \zeta^{\alpha}_B(p^\alpha(\emptyset)) + 2(1 - q^\alpha)^2(q^\alpha)^2 + 2q^\alpha(1 - q^\alpha)\zeta^{\beta}_B(p^\beta(\emptyset)) \right] \zeta^{\beta}_B(p^\beta(\emptyset)) A_S.$$  

This is because the $\beta$-product in the initial CS is explored if either the first exploration of the $\alpha$-product in the initial CS yields a bad outcome, or the history of outcomes of the $\alpha$-product in the initial CS contains two bad outcomes. Furthermore, the $\alpha$-product in the initial CS is approved in the following cases: (1) its first two explorations yield two good outcomes; (2) its first four explorations yield the outcomes $(G, B, G, G)$; (3) its first exploration yields a bad outcome, at which point the $\beta$-product is explored, yields a bad outcome and is abandoned, and the subsequent three explorations of the $\alpha$-product yield three good outcomes.

*Selection of $\alpha$-products after reduction in approval standard.* Denote by $\hat{A}_S$ and $\hat{A}$ the analogs of $A_S$ and $A$, respectively, after the reduction in the approval standard. Given the ordering in (S.9), we have that

$$\hat{A}_S = \rho^\alpha \left[ \zeta^{\alpha}_G(p^\alpha(\emptyset)) + \zeta^{\alpha}_B(p^\alpha(\emptyset)) \hat{A}_S \right] + \rho^\beta \zeta^{\beta}_B(p^\beta(\emptyset)) \hat{A}_S,$$

from which we obtain that

$$\hat{A}_S = \frac{\rho^\alpha \zeta^{\alpha}_G(p^\alpha(\emptyset))}{1 - \left\{ \rho^\alpha \zeta^{\alpha}_B(p^\alpha(\emptyset)) + \rho^\beta \zeta^{\beta}_B(p^\beta(\emptyset)) \right\}}.$$  

The ex-ante probability with which one of the $\alpha$-products is approved is thus equal to:

$$\hat{A} = \zeta^{\alpha}_G(p^\alpha(\emptyset)) + \zeta^{\alpha}_B(p^\alpha(\emptyset)) \zeta^{\beta}_B(p^\beta(\emptyset)) \hat{A}_S.$$  

Note that the result follows from the fact that the $\alpha$-product in the initial CS is explored first. If its exploration yields a good outcome, the $\alpha$-product is approved. If, instead, it yields a bad outcome, the $\beta$-product in the initial CS is explored next. If the exploration of the $\beta$-product yields a positive outcome, the $\beta$-product is approved, bringing an end to the exploration process. If, instead, it yields a negative outcome, search is launched, at which point the probability one of the $\alpha$-products is approved is $\hat{A}_S$.

*Comparison.* The claim in the Corollary follows from observing that Conditions (S.7) and (S.9) are consistent with $A > \hat{A}$ over an open set of parameter values.  

---

4Note that, in cases (1) and (2), the $\beta$-product in the initial CS is never explored.
Intuitively, the reduction in the approval standard for the $\alpha$-products unambiguously contributes to an increase in the value the firm attaches to experimenting with such products. However, it also increases the value the firm assigns to basic research, as the new $\alpha$-products that the firm expects to discover are more likely to be approved. Search may thus crowd out further experimentation with those $\alpha$-products whose earlier tests yielded negative outcomes. Because search also brings $\beta$-products, the changes in the composition of the CS due to search may then favor the $\beta$-products more than the $\alpha$-ones, possibly leading to a reduction in the ex-ante probability of approval of the $\alpha$-products. The proof above identifies conditions for which this happens. As in the application in the previous section, such conditions tend to involve intermediate values of the relevant parameters. The reason for the non-monotonicities is similar to the one discussed in the previous section.

### S.3 Comparative statics - online consumer search

Consider the setting described at the end of Subsection 4.4. The following result illustrates how the increase in the probability that search brings an additional product by firm $\xi$ may reduce the index of search, inducing the consumer to visit the website of one of firm $\xi$’s competitors before searching for new products. When strong enough, such an effect may reduce the probability that one of firm $\xi$’s product is selected, and hence firm $\xi$’s profits.

**Corollary S.3.** Consider the environment described at the end of Subsection 4.4. An increase in the probability $\rho^\xi$ that search brings an additional product from firm $\xi$ may reduce firm $\xi$’s ex-ante expected profits.

**Proof of Corollary S.3.** Suppose that each $F^\xi$ is a Bernoulli distribution assigning probability $p^\xi$ to $v = \hat{v}^\xi$ and $(1 - p^\xi)$ to $v = 0$, with $\hat{v}^\xi \in \mathbb{R}^+$. Each firm makes equal profits on each of its two products. Hence, each firm’s ex-ante total profits are equal to the total probability with which one of its two products is selected. To keep things simple, suppose the consumer incurs no cost for inspecting any product other than the time cost of postponing the final purchase: that is, $\lambda^\xi = 0$ for $\xi = A, B, C$. The consumer’s discount factor is $\delta$.

**Exogenous CS.** Suppose the identity of the firm receiving the additional slot is determined ex-ante, i.e., before the consumer starts the exploration process. Given the composition of the CS, the consumer then sequentially decides which product to inspect and when to stop, at which point she either chooses one of the inspected products or her outside option (whose value is normalized to zero). As shown in the main text, the reservation price for each $\xi$’s product, before the latter is inspected, is equal to

$$I(\xi, \emptyset) = \frac{(1 - \delta) p^\xi \hat{v}^\xi}{1 - \delta + p^\xi},$$

whereas the reservation price of each $\xi$’s product after it is inspected is equal to $I(\xi, v) = (1 - \delta) v$, with $v \in \{\hat{v}^\xi, 0\}$. The optimal policy is to inspect products in descending order of their reservation

\[^5\text{One can think of } \hat{v}^\xi \text{ as the value (net of price) to the consumer in case the product is a good match, and } p^\xi \text{ as the probability of such an event.} \]
prices, stopping when the remaining reservation prices are all smaller than the maximal realized value among the inspected products. Clearly, in this environment, each firm benefits from an increase in the probability it is given a second slot.

**Endogenous CS.** Now suppose that the consumer’s initial CS consists of three products, one from each firm $\xi = A, B, C$, and that the CS can be expanded only once, with the expansion bringing an additional product drawn from $\Xi$ according to $\rho$, with $\rho^\xi \geq 0$, $\xi = A, B, C$, and with $\sum_{\xi} \rho^\xi = 1$. The result in the corollary then follows from Claim S.1 below.

Claim S.1. Suppose that $\mathcal{I}(A, \emptyset) > \mathcal{I}(B, \emptyset) > \mathcal{I}(C, \emptyset)$. There exist parameter values consistent with the above inequalities such that an increase in $\rho^B$, together with a reduction by the same amount in $\rho^A$, leads to a decrease in the overall probability that one of firm B’s products is sold (and hence in its ex-ante expected profits).

Proof. We establish the claim above by showing that an increase in the probability that search brings an extra $B$-product (along with a reduction by the same amount in the probability that it brings an $A$-product) may reduce the attractiveness of search thus inducing the consumer to inspect firm C’s product before expanding the CS. We show that this effect may imply a drop in firm B’s ex-ante profits.

It is easy to verify that the index for search is equal to

$$
\mathcal{I}^S = \delta^2 \max_{k \in \{A, B, C\}} \left\{ \frac{\sum_{\xi \in \{\xi' \in \Xi | \mathcal{I}(\xi', \emptyset) \geq \mathcal{I}(k, \emptyset)\}} \rho^\xi \rho^* \delta \frac{\xi \delta}{1 + \sum_{\xi' \in \Xi, \mathcal{I}(\xi', \emptyset) \geq \mathcal{I}(k, \emptyset)} \rho^\xi \delta \left(1 + \frac{\rho^*}{1 - \delta}\right)} \right\}.
$$

(S.10)

For concreteness, let $\delta = 0.9$ and suppose that $(\hat{\nu}^A, p^A) = (10, \frac{1}{10})$, $(\hat{\nu}^B, p^B) = (3, \frac{1}{3})$, and $(\hat{\nu}^C, p^C) = (2, \frac{1}{2})$. Note that the distributions $F^\xi$ from which the consumer’s values for the firms’ products are drawn have the same mean, but are mean preserving spreads of one another; hence $\mathcal{I}(A, \emptyset) > \mathcal{I}(B, \emptyset) > \mathcal{I}(C, \emptyset)$. Suppose that, initially, $\rho^A = \rho^B = \frac{1}{4}$, and $\rho^C = \frac{1}{2}$. It is easily verified that $\mathcal{I}(A, \emptyset) = 0.473$, $\mathcal{I}(B, \emptyset) = 0.225$, $\mathcal{I}(C, \emptyset) = 0.163$, and $\mathcal{I}^S = 0.174$. Also note that $\mathcal{I}(C, \emptyset) < \mathcal{I}^S < \mathcal{I}(B, \emptyset)$, so that $\mathcal{I}^S$ does not take into account the benefits from inspecting firm C’s additional product, in case search brings a second product by firm C.

Now suppose $\rho^B$ is increased by $\zeta \in [0, 0.25]$ while $\rho^A$ is reduced by the same amount. Let $\phi(\zeta)$ denote the probability that one of firm B’s products is ultimately chosen when the probability that search brings a $B$-product is $\rho^B + \zeta$. Figure 1 depicts the change $\phi(\zeta) - \phi(0)$ in the probability that one of firm B’s products is selected as a function of $\zeta$, where $\phi(0) = (1 - p^A)(p^B + (1 - p^B)p^B) = 0.35$. The horizontal gray lines correspond to the indices $\mathcal{I}(A, \emptyset)$, $\mathcal{I}(B, \emptyset)$, and $\mathcal{I}(C, \emptyset)$, whereas the dark gray curve depicts $\mathcal{I}^S$, as a function of $\zeta$. Note that $\mathcal{I}^S$ is decreasing in $\zeta$, since $\mathcal{I}(A, \emptyset) > \mathcal{I}(B, \emptyset)$. Hence, an increase in $\zeta$ implies a lower index for search. $\mathcal{I}^S$ starts out above $\mathcal{I}(C, \emptyset)$, and intersects $\mathcal{I}(C, \emptyset)$ at an interior $\zeta$ (smaller than 0.25), denoted by $\zeta^*$ (the vertical dashed line). For $\zeta < \zeta^*$, $\mathcal{I}(C, \emptyset) < \mathcal{I}^S < \mathcal{I}(B, \emptyset)$, whereas for $\zeta > \zeta^*$, $\mathcal{I}^S < \mathcal{I}(C, \emptyset)$. The function $\mathcal{I}^S(\zeta)$ has a kink at $\zeta = \zeta^*$. For $\zeta \in [0, \zeta^*)$, the CS is expanded before firm C’s product is inspected,
whereas for $\zeta \in (\zeta^*, 0.25]$ the opposite is true. Therefore, the probability that one of firm B’s products is chosen is equal to $\phi(\zeta) = (1 - p_A)(p_B + (1 - p_B)(\rho_B + \zeta)p_B)$ for $\zeta \in [0, \zeta^*)$ and is equal to $\phi(\zeta) = (1 - p_A)(p_B + (1 - p_B)(1 - p^C)(\rho_B + \zeta)p_B)$ for $\zeta \in (\zeta^*, 0.25]$, with a downward discontinuity at $\zeta = \zeta^*$ equal to $(1 - p_A)(1 - p_B)p_Bp_C(\rho^2 + \zeta^*)$. Furthermore, the downward drop in $\phi(\zeta)$ at $\zeta = \zeta^*$ makes $\phi(\zeta) - \phi(0)$ negative over $(\zeta^*, 0.25]$, thus establishing the claim above.

S.4 Irreversible Choice Among Alternatives

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 $M_\xi$ times (with $\xi$ denoting the alternative’s category).\footnote{If $M_\xi = 0$, the DM can irreversibly commit to any $\xi$-alternative without first exploring it.} 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 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” $\omega^P$ if and only if $\hat{\xi} = \xi$, $\theta = (\hat{\vartheta}_1, \ldots, \hat{\vartheta}_m)$, for some $m$, and $\hat{\theta} = (\vartheta_1, \ldots, \vartheta_m, \ldots, \vartheta_{\hat{m}})$ for some $\hat{m} \geq m$. Denote this relation by $\hat{\omega}^P \succeq \omega^P$.

Condition 1. A category-$\xi$ alternative has the better-later-than-sooner property if, for any $\omega^P =
such that $\theta = (\vartheta_1, \ldots, \vartheta_m)$, 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-$\xi$ box requires $M_\xi$ 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 $M_\xi$ explorations, where $M_\xi$ 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 $M_\xi$-th exploration) remains constant after the $M_\xi$-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 $\xi$ in state $\omega^P = (\xi, \theta)$ is a function of the information $\theta = (\vartheta_1, \ldots, \vartheta_m)$ 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 $\theta = (\vartheta_1, \ldots, \vartheta_m)$ and $\hat{\theta} = (\vartheta_1, \ldots, \vartheta_m, \ldots, \vartheta_{\hat{m}})$ 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:

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

where $\tau$ is a stopping time, and where $\pi$ is a rule specifying whether the DM explores the alternative, or irreversibly commits to it. Similarly, modify the index of search $I^S(\omega^S)$ by letting the rule $\pi$ 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 $\chi^*$ as follows. At each period $t \geq 0$, given the state $S_t$ of the decision problem, the policy specifies to (a) search if $T^S$ is greater than the index $I^P$ 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 $\omega^P$ if its index $I^P$ 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 $\omega^P$ if its retirement value $R$ is greater than its index $I^P$, 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 $\tau^*$ 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 $T^S$, all the indexes of the alternatives brought to the CS by search, and all retirement values of such alternatives fall below the value $T^S(\omega^S)$ 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 $\omega^P$, 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 $\chi^*$ does not hinge on this assumption. Consider first an environment where $M_\xi = 0$ for all $\xi$. 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 $M_\xi$ is stochastic and learned over time.

Consider the following fictitious environment, where all choices are reversible. Whenever an alternative of category $\xi$ is brought to the CS, an additional auxiliary alternative is also introduced into the CS, yielding a fixed flow payoff of $R(\xi, \emptyset)$.\(^7\) Furthermore, whenever a non-auxiliary alternative in state $\omega^P$ 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 $H_{\omega^P}$, as in the baseline model.\(^8\) We say that an auxiliary alternative corresponds to a (non-auxiliary) alternative

---

\(^7\)Recall that $R(\xi, \emptyset)$ is the retirement value of a physical alternative of category $\xi$ that has never been explored.

\(^8\)If $M_\xi > 0$, the introduction of the auxiliary alternative as the result of the exploration of an alternative in state $\omega^P = (\xi, \emptyset)$ occurs only if $\theta = (\theta_1, ..., \theta_s)$ with $s \geq M_\xi$, that is, only if the alternative has been explored at least $M_\xi$
in state $\omega^P$ 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 $\omega^P$. In this auxiliary environment, define the index of search as in the main text, with the rule $\pi$ 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 $\omega^P$, define its new index as in (S.11), with the rule $\pi$ 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, $\pi$ 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’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.\(^9\) 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 $\pi$ in the definition of the index of the non-auxiliary alternatives in the fictitious problem coincides with the one in (S.11) where the selection $\pi$ 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 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 $M_\xi$ times, with $M_\xi$ possibly stochastic and learned over time (in this latter case, the time-varying component of an alternative’s state, $\theta$, may also contain information about $M_\xi$).

\(^9\)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.
S.5 Suboptimality of Index Policies with “Meta-Arms”

In this section, we briefly illustrate, by means of an example, why multi-armed 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 main text. 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 body of the 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 need not be an index policy. The following example illustrates.

**Example S.3.** There are two arms. Arm 1 yields a reward of 1,000 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.