

The Pandora's Box Problem with Sequential Inspections

Ali Aouad

Sloan School of Management, Massachusetts Institute of Technology, maouad@mit.edu

Jingwei Ji

Management Science and Engineering, Stanford University, jingwei.ji@stanford.edu

Yaron Shaposhnik

Simon Business School, University of Rochester, yaron.shaposhnik@simon.rochester.edu

Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and are not intended to be a true representation of the article's final published form. Use of this template to distribute papers in print or online or to submit papers to another non-INFORM publication is prohibited.

Abstract. The Pandora's box problem (Weitzman 1979) is a core model in economic theory that captures an agent's (Pandora's) search for the best alternative (box). We study an important generalization of the problem where the agent can either *fully open* boxes for a certain fee to reveal their exact values or *partially open* them at a reduced cost. This introduces a new tradeoff between information acquisition and cost efficiency. We establish a hardness result and employ an array of techniques in stochastic optimization to provide a comprehensive analysis of this model. This includes (1) the identification of structural properties of the optimal policy that provide insights about optimal decisions; (2) the derivation of problem relaxations and provably near-optimal solutions; (3) the characterization of the optimal policy in special yet non-trivial cases; and (4) an extensive numerical study that compares the performance of various policies, and which provides additional insights about the optimal policy. Throughout, we show that intuitive threshold-based policies that extend the Pandora's box optimal solution can effectively guide search decisions.

Key words: Pandora's box problem, dynamic programming, approximate algorithms

1. Introduction

Sequential search problems refer to a class of optimization problems that have long been studied in the economics, operations research (OR), and computer science (CS) literature. In his seminal work, Weitzman (1979) studies the Pandora's box sequential search problem,¹ in which Pandora is given a collection of boxes containing random prizes drawn from known and independent probability distributions. She can adaptively pay box-dependent fees to open boxes and observe their realizations until she decides to select the prize held within one of the open boxes, or stop exploring without

selecting any box. Her objective is to maximize the expected profit, that is, the expected value of the collected prize minus the expected combined box-opening fees. This fundamental model, which captures an agent's search for the best alternative, found applications in consumer search, project management, job recruitment, and asset-selling problems, to name a few.

In this paper, we consider a generalization of this problem in which Pandora can also exercise a cheaper and possibly less accurate option of *partially opening* boxes. When partially opening a box, Pandora observes its *type*, an informative signal based on which she updates her belief about the prize inside the box. We refer to this variant as *Pandora's Box with Sequential Inspection* (PSI for short). Similarly to the standard problem, she may only select a prize contained in a box that was *fully opened*, that is, a box whose exact prize has been observed. However, the ability to partially open boxes introduces a new tradeoff. On the one hand, partially opening a box reveals certain information about the prize value that allows Pandora to potentially detect "bad outcomes" at a lower cost, and thus, avoid paying the full opening fee. On the other hand, while the cost of partial opening is cheaper than the cost of full opening, it may not be negligible, in which case overusing the option to partially open may lead to excessive costs.

To put this model in concrete terms, consider job recruitment, an application discussed in Weitzman (1979). Recent technological advances transform job recruitment processes in various industries. For example, video conferencing is often used for remote (offsite) interviews. More recently, AI systems have been introduced to conduct such initial remote interviews without human intervention; the transcriptions are then reviewed by human evaluators. Remote interviews provide convenient ways for companies to learn more about candidates at reduced costs, compared with onsite interviews. Nevertheless, remote interviews cannot fully replace onsite visits, which are often still necessary to eventually make job offers. Moreover, while cheaper than onsite visits, substantive remote interviews – beyond the preliminary screenings based on CVs – still require time from human interviewers or evaluators to assess the candidates' fit for the position. In selecting a candidate for a vacant position, companies must therefore decide how to synthesize the different mediums to identify and select qualified candidates while minimizing costs. Put into the framework of our model, Pandora represents the company, boxes correspond to candidates whose prior evaluations are based on screenings and automated assessments, and box opening actions correspond to conducting interviews – offsite or onsite. The goal is to effectively administer the hiring process by deciding which candidates should be interviewed and in which format: cheaper and less accurate offsite interviews, or more expensive and informative onsite visits.

More generally and beyond the particular application, our work tries to study how to integrate different *sequential information acquisition* methods in the context of stochastic optimization problems. To our knowledge, this is the first work that examines this feature, which is quite common in various practical sequential search problems (a more detailed discussion can be found in Section 1.2 that surveys related work).

Weitzman (1979) showed that there exist optimal policies to the standard Pandora's problem that admit an elegant and simple threshold-based structure. A threshold is defined for each box, indicating the order in which boxes should be considered for opening, reducing the problem to an optimal stopping time problem. The decision about when to stop is determined using the same thresholds by comparing the largest prize amongst open boxes to the largest threshold amongst closed boxes; if the incumbent prize is larger, Pandora should stop; otherwise, she should proceed to the next box, and so forth. However, it has been observed in the literature (see Section 1.2) that seemingly small changes in the model assumptions break the structure of the optimal policy and lead to problems whose optimal solution cannot be analytically characterized. Our goal is to study this fundamental sequential search process when partial information acquisition is possible. Specifically, we strive to answer the following questions: How should partial opening and full opening be prioritized? Could threshold-based policies, in the same vein as those defined in the standard Pandora's box problem, describe the optimal policy in settings with sequential inspections? Beyond the characterization of optimal policies, we are also interested in developing simple and intuitive policies with provably good performance guarantees.

1.1. Results and contributions

We summarize below the main findings and contributions of our work.

Modeling. We introduce a new model for sequential search by an agent, which incorporates sequential information acquisition into stochastic optimization problems. Such information acquisition process can be represented by a directed acyclic graph describing the information states of each box (see Figure 1). Our graph layout is unique in that boxes can be inspected repeatedly and there is more than one way to inspect boxes. To our knowledge, this generalized graph layout has not been previously studied. Moreover, it raises a fundamental technical challenge, which has been a long-lasting open question in the literature (we further discuss this in Section 1.2).

Hardness. We prove that computing optimal policies for PSI is NP-hard, contrary to the standard Pandora's box problem. The proof is based on a reduction defined using PSI instances that consist of boxes with binary prizes and where partial opening indicates one of two types.

Characterization of an optimal policy. Despite the hardness result, we identify threshold-based rules that almost fully describe the optimal policies. Similarly to the Pandora's box problem, one type of threshold is associated with decisions to fully open boxes. To capture the added value of partial inspections, we define two additional types of thresholds associated with decisions to partially open boxes. We show that, in most cases, the optimal actions can be prescribed using the newly defined thresholds. In particular, we characterize the optimal stopping criterion (i.e., when to stop searching and collect the best prize presently available), and identify sufficient conditions under which it is optimal to fully or partially open a box. We identify special non-trivial cases where our characterization fully describes an optimal policy, and show that a simple policy, which is defined using these thresholds, is asymptotically optimal. Numerical experiments show that optimal actions are identified by our theoretical characterization in more than 92% of all systems states, on average, and this percentage increases in the number of boxes of the problem instance.

Simple robust policies. Next, we focus on the class of *committing policies*.² These semi-adaptive policies specify a priori whether each box should be partially or fully opened the first time it is inspected. Interestingly, each committing policy induces a corresponding multi-armed bandit problem (MAB) in which the optimal indices coincide with our proposed thresholds. By leveraging fundamental results on the adaptivity gap of stochastic submodular optimization problems, we argue that the optimal a committing policy can be efficiently approximated within a ratio of $1 - 1/e$.

Numerical experiments. We conduct extensive numerical experiments to evaluate the performance of various policies on a diverse set of synthetic instances and obtain additional insights about the optimal policy:

- Solving the problem to optimality using exact dynamic programming methods is computationally prohibitive, even when the number of boxes is around 10, the prize distributions have a small support, and the number of types is small. These computational challenges warrant the development of simple suboptimal policies.
- The number of times an optimal policy exercises the partial opening option increases as (i) there are more boxes, (ii) the partial opening thresholds tend to be larger than the corresponding full opening thresholds, and (iii) the distributions of the prizes are similar across boxes. In contrast, when boxes tend to be different, much about the order by which boxes are inspected is determined without partial opening.
- The newly introduced thresholds are key to the development of effective policies. Specifically, the class of committing policies, which utilize our thresholds, achieves a robust performance. In

particular, the best committing policy is very close to the optimum, both in terms of average and worst-case performance. In line with the asymptotic analysis, we observe that an intuitive threshold-based committing policy³ is near-optimal when the number of boxes is sufficiently large.

Our proofs combine several methods used in recent literature, including sample path coupling and interchange arguments (Attias et al. 2017), reductions to submodular maximization problems (Beyhaghi and Kleinberg 2019), and results based on the information relaxation duality approach (Balseiro and Brown 2019). In this context, our main technical contribution lies in analyzing threshold-based rules using coupling arguments. We also reveal an unnoticed connection between approximation methods independently proposed in the literature. Specifically, we show that the cost accounting technique of Beyhaghi and Kleinberg (2019) amounts to a relaxation of Whittle's integral, a more general upper bound for multi-armed bandit superprocesses.

Overall, our results suggest that generalized Weitzman's thresholds play an important role in characterizing both optimal and near-optimal inspection and selection decisions when the decision-maker can sequentially inspect boxes. This aspect of selection processes has not been thoroughly studied in previous literature, as we discuss next.

1.2. Related Work

Our work relates to several streams of literature in economics, operations research, and computer science. We divide them imperfectly into three parts: generalization of the standard problem, Markovian MAB problems, and applications of sequential search.

The Pandora's box problem and its variants. The seminal work of Weitzman (1979) led to substantial research on more general and realistic settings. For example, Vishwanath (1992) consider a generalization with parallel searches, where Pandora can open several boxes simultaneously at a discounted cost. The authors identify distributional assumptions under which it is optimal to open boxes according to Weitzman's thresholds. Olszewski and Weber (2015) examine a generalization where the objective function depends on all discovered prizes.⁴ They show that if a threshold-based rule is unconditionally optimal, then the problem is reducible to the standard Pandora's box setting. Boodaghians et al. (2020) study a variant of the Pandora's box problem in which precedence constraints are imposed on the inspections of boxes. They show that when the precedence constraints are tree-shaped, the optimal policy can still be described as a threshold-based rule. [Gibbard \(2022\)](#) extend the Pandora's box problem by assuming that the prize is additive and that a two-step sequential inspection reveals the exact value of each part. Conceptually, such a model studies

how information acquisition should be prioritized when the acquisition process can be broken into multiple steps. The model is a special case of the MAB problem (Bertsimas and Nino-Mora 1996) that inherits the indexability structure of the optimal policy. In contrast, our model attempts to capture a different tradeoff between two information acquisition methods (a cheaper and less precise versus a more expensive and accurate), a characteristic which we show to be challenging analytically and computationally.

Recently, there has been a growing interest in non-obligatory inspections (Attias et al. 2017, Doval 2018, Beyhaghi and Kleinberg 2019) where Pandora may select any closed box without inspecting it (in which case the prize is random at selection). Building on a clever cost accounting method, Beyhaghi and Kleinberg (2019) show that this setting is a special case of maximizing a stochastic non-negative submodular function under a matroid constraint, a breakthrough that paved the way to the design of approximation algorithms. Our algorithmic results in Section 4 draw from this line of reasoning. Closer to our work, Beyhaghi (2019) studies a setting in which Pandora can choose from multiple opening actions for each box. However, unlike our work, the search process in that scenario does not feature the notion of sequential inspections. Therefore, such a model does not capture the key tradeoff which we aim to study between reducing uncertainty significantly at a higher cost, and reducing it only partially at a lower cost with an option to further reduce the uncertainty later at an overall higher cost. This very tradeoff is present in the applications that motivate our work. Finally, we note the work of Clarkson et al. (2020) that study a problem where an object is hidden at one of several locations according to some known distribution. The goal is to find it in the minimum expected number of attempts using two modes of inspection, fast or slow. This can be seen as a variant of the Pandora's box problem, where the reward is binary, and boxes are correlated and can be inspected indefinitely.

Markovian multi-armed bandits. From a technical perspective, the standard Pandora's box problem is a special case of the Markovian MAB (Gittins 1979), in which an agent sequentially activates (pulls) one out of many bandits (arms). Arms correspond to independent Markov reward processes, which generate a reward and transition to a new state upon activation. The optimal solution admits an index structure, known as the Gittins index⁵, where every state of every bandit is associated with an index, and it is always optimal to activate the arm whose current state's index is maximal. However, the optimality of the Gittins index breaks under small variations of the model assumptions, such as correlated arms (Pandey et al. 2007), switching costs (Asawa and Teneketzis 1996),

or delayed feedback (Eick 1988). We similarly observe that index policies are suboptimal in our setting (see Example 2 in Section 2.2).

The problem considered in our paper can be viewed as a special case of MAB superprocesses (Gittins et al. 2011), in which there exist multiple ways to pull each arm. That is, after an arm is selected, one has to choose the mode of activation, which affects transitions and rewards. General knowledge about this class of problems is rather limited. Unlike the standard MAB problem, MAB superprocesses need not admit an optimal index policy. Much of the existing literature explores special cases where index policies are “unambiguously” optimal (Whittle 1980), as well as heuristic solution methods and upper bounds (Brown and Smith 2013, Balseiro and Brown 2019). The fact that we identify that our problem is NP-Hard suggests that a complete characterization of the optimal policy is not possible, not even through the MAB framework.

Sequential search and applications. The stylized sequential search process introduced by Weitzman (1979) has been widely used in economics, marketing, and operations research. This includes empirical work (De los Santos et al. 2012, Ursu 2018), behavioral research (Gabaix et al. 2006), and economic models that use the Pandora’s box as a discrete-choice framework (Miller 1984, Robert and Stahl 1993, Choi et al. 2018, Derakhshan et al. 2020, Chu et al. 2020, Immorlica et al. 2020). Our work offers a richer modeling structure to represent decision-making processes where alternatives can be inspected in stages with varying costs, although our current contribution remains theoretical.

We note in passing about the existence of a literature on sequential search problems modeled using a Bayesian framework where there is a single box that can be sampled repeatedly for a fee and the decision maker tries to decide when to stop sampling with the goal of maximizing the expected value of the highest observed sample minus the sampling costs (see, e.g., the recent work of Baucells and Zorc 2023, and the references therein). Although the idea of sequential inspection is similar to our work, the key characteristics of our model, a tradeoff between cheap and costly inspections across multiple alternatives, are not captured by this literature.

2. Problem Formulation

Pandora (the decision-maker) is given $N \in \mathbb{N}^*$ boxes containing random prizes. Each box $i \in [N]$ is associated with a priori unknown prize V_i and type T_i drawn from a known joint probability distribution \mathcal{D}_i . We assume that V_i is either a continuous or discrete scalar random variable and that T_i is a discrete scalar random variable with support Γ_i .⁶ We assume that the support of \mathcal{D}_i is

contained in $\mathbb{R}^+ \times \Gamma_i$, and that the random variables $(V_1, T_1), \dots, (V_N, T_N)$ are mutually independent (for each box i , the random variables V_i and T_i are dependent; otherwise the type would not be informative about the prize). Pandora can open a box in one of two inspection modes: full opening (F-opening) and partial opening (P-opening). The realization of type T_i is revealed immediately by P-opening box $i \in [N]$, while the corresponding F-opening action reveals the realization of (V_i, T_i) . These opening actions incur constant costs, respectively denoted by c_i^F and c_i^P , which are also known to Pandora.

Figure 1 illustrates the state evolution of each box following various actions. We call the initial state of a box *closed*; here, the realizations of V_i and T_i are unknown. P-opening a closed box changes its state to *partially open*. In this case, Pandora observes the realization of T_i , which we denote by $t_i \in \Gamma_i$, and going forward, she updates her belief about V_i to the conditional distribution $\Pr[V_i \in \cdot \mid T_i = t_i]$. We sometimes refer to this conditional distribution as $V_i \mid t_i$ for short. We say that a box is *fully open* once F-opening takes place. Here, Pandora observes the realizations of V_i and T_i , which we denote by v_i, t_i , respectively. Lastly, the problem terminates when a fully open box is selected. We note that full opening is a prerequisite for selection in a way similar to the standard Pandora's box problem. The state of the system is jointly described by the collective individual state of all boxes.

An adaptive policy in our generalized Pandora's box problem is a mapping between system states and feasible actions. There are three types of feasible actions: (1) P-opening a closed box, (2) F-opening a closed or partially open box, and (3) selecting a fully open box. Alternatively, this can be seen as a cascade process of (1) deciding whether to select a box (stop) or keep inspecting (continue), (2) choosing a box to inspect, and (3) deciding on a box-opening mode. Our goal is to develop a policy that maximizes the expected profit, defined as the selected prize minus the box-opening costs.

2.1. Dynamic programming (DP) formulation

The problem naturally lends itself to a DP formulation. While the state space is exponentially large (and therefore, the DP cannot be approached using standard solution methods), we exploit this formulation to derive structural properties of the optimal policy.

State space. Each state of the DP can be described by a tuple (C, \mathcal{P}, y) , where $C \subseteq [N]$ is the set of closed boxes, \mathcal{P} is formed by all pairs (i, t_i) , describing each P-opened box i and its corresponding type t_i (revealed by P-opening), and y is the maximal prize found so far (it is easy to see that all fully

open boxes whose prize is smaller than y can be discarded, since at most one prize can be claimed). We observe that the total number of states is at least $\prod_{i=1}^N (1 + |\Gamma_i|)$, which is exponentially large.

Actions. At each state (C, \mathcal{P}, y) , we can fully open a closed or partially open box, partially open a closed box, or select a fully open box and stop. After F-opening a box $i \in C$ and observing the realization v_i of its prize V_i , we transition to state $(C \setminus \{i\}, \mathcal{P}, \max\{y, v_i\})$. Hereinafter, we slightly abuse the notation by using $i \in \mathcal{P}$ in place of $(i, t_i) \in \mathcal{P}$. If we choose to F-open a box $i \in \mathcal{P}$, we transition to state $(C, \mathcal{P} \setminus \{(i, t_i)\}, \max\{y, v_i\})$, where v_i is the realization of $V_i|t_i$ revealed by our action. Likewise, if we choose to P-open a box $i \in C$, we transition to state $(C \setminus \{i\}, \mathcal{P} \cup \{(i, t_i)\}, y)$, where t_i is the type realization of T_i revealed by P-opening box i . Finally, when we select the fully open box whose value is equal to y , the decision process terminates.

Bellman's Equation. Let $J(C, \mathcal{P}, y)$ denote the value function in state (C, \mathcal{P}, y) under an optimal policy. Then, $J(C, \mathcal{P}, y)$ satisfies the Bellman's Equation:

$$J(C, \mathcal{P}, y) = \max \begin{cases} y & \text{select (stop)} \\ -c_i^F + \mathbb{E}_{V_i} [J(C \setminus \{i\}, \mathcal{P}, \max\{V_i, y\})] & \text{F-open box } i \in C \\ -c_i^F + \mathbb{E}_{V_i|T_i=t_i} [J(C, \mathcal{P} \setminus \{(i, t_i)\}, \max\{V_i, y\})] & \text{F-open box } i \in \mathcal{P} \\ -c_i^P + \mathbb{E}_{T_i} [J(C \setminus \{i\}, \mathcal{P} \cup \{(i, T_i)\}, y)] & \text{P-open box } i \in C \end{cases} \quad (1)$$

$$J(\emptyset, \emptyset, y) = y.$$

In what follows, we use the shorthand $s_0 = ([N], \emptyset, 0)$ to designate the initial system state.

Policies. A feasible policy π is a mapping from each state to a corresponding action, which is selected amongst all those that are feasible in the state in question. For simplicity, we assume that there exists a unique optimal policy that satisfies Bellman's Equation.

2.2. Illustrative examples

To provide some intuition about why P-opening could be beneficial, consider the following simple examples.

EXAMPLE 1. There is a large number of i.i.d. boxes; each contains a prize of 110 with probability $\frac{1}{2}$ and zero otherwise. The box-opening costs are $c_i^F = 100$ and $c_i^P = \epsilon$, and P-opening acquires the same information as F-opening (nevertheless, F-opening is still required to collect the prize). It is easy to see that given only the F-opening inspection mode, it is not worthwhile to F-open any box and the expected profit of any policy that opens boxes is negative. However, given the option to P-open, one could identify a box that contains the prize prior to F-opening it. In that case, the

expected profit is $110 - 100 = 10$ minus ϵ times the expected number of opened boxes, which is equal to two.

This example shows how the option to P-open could transform a problem where opening boxes is not worthwhile into one in which it is. Hence, the optimality gap for the naive policy that ignores P-opening (i.e., Weitzman's policy, which is a natural candidate for Pandora's box-like problems) is unbounded.

EXAMPLE 2. Consider the state $(\{1\}, \emptyset, y)$ where there is one box left with two types: 'exceptional' and 'average,' which can be revealed by P-opening. Box 1 is 'exceptional' with probability $\delta = 0.01$ and is 'average' with probability $1 - \delta$. If the box is exceptional, the prize is known to be 2 almost surely, but if it is average, the prize is uniformly distributed in $[0, 1]$. Assuming that the F-opening cost is $c_1^F = \frac{1}{3}$ and the P-opening cost is $c_1^P = \frac{1}{3}\delta$, it can be easily seen that when $y = 0$, both inspection modes are preferred over stopping (which yields 0). Moreover, F-opening dominates P-opening because F-opening is inevitable in this case. However, when $y = 1$, F-opening would be no longer worthwhile compared to stopping, because the cost $c_1^F = \frac{1}{3}$ exceeds the expected gain $\delta \cdot (2 - y) = \delta$. By contrast, P-opening (followed by the F-opening) of an exceptional box is advantageous because the expected cost is $c_1^P + \delta c_1^F = \frac{1}{3}\delta + \frac{1}{3}\delta = \frac{2}{3}\delta$, whereas the expected gain is δ .

Intuitively, in the context of job hiring, the instance captures a candidate who is likely 'average' but has a small chance of being 'exceptional.' A low-cost online interview can reveal this. If no good candidate has been found ($y = 0$), it is optimal to proceed directly with an onsite interview. However, if a strong candidate has already been identified ($y = 1$), the onsite visit should be reserved for an 'exceptional' applicant, identified via a relatively short online screening.

Technically, the example shows that the optimal policy is not indexable. If it were, each action would receive a fixed index, and the best action—corresponding to the highest index—would be independent of y , which is not the case in Example 2. In MAB terms, the optimal activation of one arm depends on the state of others. This illustrates the added complexity of bandit superprocesses, which do not inherit the indexability of classic MABs (see also Section 1.2).

3. Characterizations of Optimal Policies

We begin by studying structural properties of optimal policies, which can be exploited to identify optimal actions using well-designed *opening thresholds*. In Section 3.1, we show that it is optimal to stop when the largest available prize exceeds all opening thresholds. In Section 3.2, we establish

that whenever the largest opening threshold corresponds to an F-opening action, it is optimal to F-open the corresponding box. In contrast, we show that the complementary case, in which the largest opening threshold corresponds to a P-opening action, is more complex to analyze, as there exists no threshold-based decision rule that is universally optimal. Nevertheless, in Section 3.3, we identify sufficient conditions for which P-opening the box with the largest threshold is optimal. In Section 3.4, we provide a complete characterization of the optimal policy for a special case of a PSI with binary prizes where partial opening reveals two types.

3.1. Opening thresholds and optimal stopping rule

We start by defining *opening thresholds*, a type of index associated with each potential action. Although there may not exist an optimal index policy (as shown in Section 2.2), we show that such indices explain to a large degree the optimal policy. While applicable to general system states, the thresholds can be most easily illustrated by considering the special case where there is a single box that is not fully open. This setting also allows us to express the optimal stopping criterion using our thresholds. We let $x^+ = \max\{0, x\}$ for every $x \in \mathbb{R}$.

DEFINITION 1 (THRESHOLDS). For every box $i \in [N]$, we define the opening thresholds:

- If box i is closed, the F-threshold σ_i^F is a solution to

$$c_i^F = \mathbb{E}_{V_i} \left[(V_i - \sigma_i^F)^+ \right]. \quad (2)$$

- If box i is partially opened with type t_i , the (conditional) F-threshold $\sigma_i^{F|t_i}$ is a solution to

$$c_i^F = \mathbb{E}_{V_i|T_i=t_i} \left[(V_i - \sigma_i^{F|t_i})^+ \right]. \quad (3)$$

- If box i is closed, the P-threshold σ_i^P is a solution to

$$c_i^P = \mathbb{E}_{T_i} \left[\max\{0, -c_i^F + \mathbb{E}_{V_i|T_i} [(V_i - \sigma_i^P)^+]\} \right]. \quad (4)$$

- If box i is closed, the FP-threshold $\sigma_i^{F/P}$ is a solution to

$$c_i^P = \mathbb{E}_{T_i} \left[\max\{0, c_i^F - \mathbb{E}_{V_i|T_i} [(V_i - \sigma_i^{F/P})^+]\} \right]. \quad (5)$$

Observe that the threshold is well-defined even in the case that $c_i^P \geq c_i^F$, in which case P-opening is not beneficial.

Assuming the prize distributions are non-atomic, it is not difficult to see that the above thresholds σ_i^F , $\sigma_i^{F|t_i}$, σ_i^P , and $\sigma_i^{F/P}$ exist and are uniquely defined. Moreover, there is a natural ordering between threshold values:

Observation 1 (Thresholds order)

1. For any box i , either $\sigma_i^{F/P} = \sigma_i^F = \sigma_i^P$, $\sigma_i^{F/P} < \sigma_i^F \leq \sigma_i^P$, or $\sigma_i^P \leq \sigma_i^F < \sigma_i^{F/P}$;
2. For any box i , we have $\min_{t_i \in \Gamma_i} \sigma_i^{F|t_i} \leq \sigma_i^{F/P}$ and $\max_{t_i \in \Gamma_i} \sigma_i^{F|t_i} \geq \sigma_i^P$.

Jingwei: — To be removed below this line — **Yaron:** remove completely or move to the appendix?

Economic interpretation. The set of F-thresholds $\{\sigma_i^F\}$ in Equation (2) describes the optimal index policy for the standard Pandora's box problem (Weitzman 1979). The conventional economic interpretation for this opening threshold is that of a fair strike price of a hypothetical option contract. That is, if c_i^F is the cost of a call option to buy prize V_i with strike price σ_i^F , then σ_i^F is the “fair” strike price that equalizes the cost and the expected profit of buying this option.

Much of this intuition carries over to our problem where inspections are sequential. Specifically, using the notion of compound option or, informally, a “two-stage call option.” The latter refers a more complex financial contract that allows the buyer to purchase an option at some future date. That is, for the price c_i^P the decision-maker can buy the initial (overlying) option, which allows her at a future date (when more information is available) to purchase the second (underlying) option for a price of c_i^F . The additional information that is available at the time of purchasing the second option is represented by the realization t_i . The conditional F-threshold represents the fair strike price for purchasing V_i , given the information that is available at the future date, while the P-threshold represents the strike price for buying the initial option. This type of option is referred to as Call on Call. On the other hand, the FP-threshold corresponds to a compound option which is referred to as Call on Put. It allows the buyer to short an option at a future time. That is, in the first stage, the buyer pays c_i^P to observe the type t_i , and then decides whether or not to enter the second stage. In the second stage, she has the right to choose whether to short a call option with prize V_i and a strike price $\sigma_i^{F/P}$. Clearly, if the strike price $\sigma_i^{F/P}$ is overpriced, i.e., greater than $\sigma_i^{F|t_i}$, a risk-neutral player will sell the second-stage option and otherwise not. Therefore, we can interpret $\sigma_i^{F/P}$ as the fair strike price to equalize the expected costs and expected profits of this two-stage option.

The above interpretation motivates a policy in which the decision maker chooses the highest option strike price—that is, according to the box and opening mode with highest threshold. We

show next that this index policy is indeed optimal for the single-box case, and we further expand its sufficient optimality conditions in the next sections.

Jingwei: — To be removed above this line —

Single-box setting and optimal stopping rule. We justify our definition of the opening thresholds through their ability to express optimal policies in single closed-box settings, and to prescribe an optimal stopping rule in general. Namely, the *single closed-box setting* is the collection of states (C, \mathcal{P}, y) where there is only one closed box that can be inspected, i.e., $C = \{i\}$ for some box $i \in [N]$ and $\mathcal{P} = \emptyset$. The difference between the expected profit from F-opening and from stopping is exactly

$$-c_i^F + \mathbb{E} \left[\max \{V_i, y\} \right] - y = -c_i^F + \mathbb{E} \left[\max \{V_i - y, 0\} \right] = -\mathbb{E} \left[\max \{V_i - \sigma_i^F, 0\} \right] + \mathbb{E} \left[\max \{V_i - y, 0\} \right]. \quad (6)$$

The F-threshold can be therefore interpreted as a break-even point for the value of y under which F-opening is preferred to stopping, and over which stopping is preferred. Given the fair strike price interpretation of the F-threshold, there is a gain from F-opening if and only if $y \leq \sigma_i^F$, meaning the best prize at hand is lower than the fair strike price. Otherwise, it is preferable to stop searching. Similarly, in this single-box case, the difference between the expected profit from P-opening (and following the optimal action thereafter) and from stopping is

$$\begin{aligned} & -c_i^P + \mathbb{E}_{T_i} \left[\max \{y, -c_i^F + \mathbb{E}_{V_i|T_i} [\max \{V_i, y\}] \} \right] - y \\ &= -c_i^P + \mathbb{E}_{T_i} \left[\max \{0, -c_i^F + \mathbb{E}_{V_i|T_i} [\max \{V_i - y, 0\}] \} \right] \\ &= -\mathbb{E}_{T_i} \left[\max \{0, -c_i^F + \mathbb{E}_{V_i|T_i} [(V_i - \sigma_i^P)^+] \} \right] + \mathbb{E}_{T_i} \left[\max \{0, -c_i^F + \mathbb{E}_{V_i|T_i} [\max \{V_i - y, 0\}] \} \right]. \end{aligned} \quad (7)$$

Similarly, the P-threshold can be interpreted as a break-even point for the value of y under which P-opening is preferred to stopping, and over which stopping is preferred. Given the fair strike price interpretation of the P-threshold, there is a gain from P-opening if and only if $y \leq \sigma_i^P$, meaning the best prize at hand y is lower than the fair strike price σ_i^P . Otherwise, it is preferable to stop searching.

Regarding the FP-threshold $\sigma_i^{F/P}$, we note that the difference between the expected profits from the optimal course of actions starting from P-opening box i and that from the optimal course of actions starting from F-opening box i is

$$\begin{aligned} & \left(-c_i^P + \mathbb{E}_{T_i} \left[\max \{y, -c_i^F + \mathbb{E}_{V_i|T_i} [\max \{V_i, y\}] \} \right] \right) - \left(-c_i^F + \mathbb{E} [\max \{V_i, y\}] \right) \\ &= -c_i^P + \mathbb{E}_{T_i} \left[\max \{0, c_i^F - \mathbb{E}_{V_i|T_i} [(V_i - y)^+] \} \right]. \end{aligned} \quad (8)$$

One can verify that the threshold $\sigma_i^{F/P}$ is a break-even point on the value of y that determines whether P-opening or F-opening achieves a higher expected profit. Intuitively, in the single-box case, the definition for the set of thresholds σ_i^F , $\sigma_i^{F|t_i}$, σ_i^P , and $\sigma_i^{F/P}$ exploits the single-crossing property between pairs of action-value functions. In the more general case, the definition can be viewed as using a myopic approximation for the action-value functions that assumes that only one box may be opened. ~~With the fair strike price interpretation of the FP-threshold in mind, P-opening is better if and only if $y > \sigma_i^{F/P}$, meaning that we already have secured a prize y higher than the fair strike price $\sigma_i^{F/P}$. Otherwise, it is preferable to F-open box i .~~

Based on the above discussion, we construct a natural *threshold-based policy* for the single closed box setting. We assign the index y to the action of stopping, the index σ_i^P to the action of P-opening, and the index σ_i^F to the action of F-opening. Our policy prescribes which action to take by comparing their corresponding indices. If the largest threshold is y , then we stop and choose the largest prize at hand. Otherwise, if the largest threshold is σ_i^F , then we F-open box i and choose the largest prize at hand. Finally, if the highest threshold is σ_i^P , we compare y to $\sigma_i^{F/P}$: if $y \leq \sigma_i^{F/P}$, we F-open, otherwise we P-open box i . We refer the reader to Figure 2 for a pictorial illustration; the policy is also summarized as Algorithm 2 in Appendix B.

The next theorem states the optimality of our threshold-based rules in the single close box setting. Moreover, in the general case (where there is more than one box), the same threshold-based logic yields an optimal stopping criterion. For any event A , we let $\mathbf{I}\{A\}$ denote the indicator function that equals 1 if A occurs and 0 otherwise.

We define the largest opening threshold $\sigma_M(C, \mathcal{P})$:

$$\sigma_M(C, \mathcal{P}) = \max \left\{ \max_{j \in C} \left\{ \sigma_j^F, \sigma_j^P \right\}, \max_{j \in \mathcal{P}} \left\{ \sigma_j^{F|t_j} \right\} \right\}. \quad (9)$$

We refer to the box attaining the largest threshold $\sigma_M(C, \mathcal{P})$ as the *leading box*.

THEOREM 1. *The threshold-based policy is optimal in the single closed box setting $(C, \mathcal{P}, y) = (\{i\}, \emptyset, y)$, and it achieves the following expected profit:*

$$J(\{i\}, \emptyset, y) = \mathbb{E} \left[\max \left\{ y, \mathbf{I}\{y > \sigma_i^{F/P}\} \cdot \min\{V_i, \sigma_i^{F|t_i}, \sigma_i^P\}, \mathbf{I}\{y \leq \sigma_i^{F/P}\} \cdot \min\{V_i, \sigma_i^F\} \right\} \right].$$

In general, for every state (C, \mathcal{P}, y) , it is optimal to stop if and only if

$$y \geq \sigma_M(C, \mathcal{P}). \quad (10)$$

The formal proof of the stopping rule is based on a coupling argument, which is presented in Appendix B. We argue that the stopping decision regarding a box (i.e., never to open this box in the future) only depends on y . Hence, the optimal stopping rule shown for the single box setting holds for the general case.

3.2. Sufficient conditions for optimal F-opening

Having characterized the optimal stopping criterion, we now focus on states (C, \mathcal{P}, y) in which stopping is not optimal. Here, there are two possibilities: $\sigma_M(C, \mathcal{P})$ corresponds either to an F-threshold or a P-threshold. The next theorem establishes that in the former case, it is optimal to F-open the leading box. This condition holds regardless of whether the leading box is closed or partially-opened.

THEOREM 2. *There exists an optimal policy that immediately F-opens the leading box in every state (C, \mathcal{P}, y) where stopping is suboptimal and $\sigma_M(C, \mathcal{P})$ is an F-threshold.*

The proof of this result appears in Appendix B.4. The crux of it is constructing a policy that F-opens the leading box, and then showing that this policy achieves a larger expected profit than any alternative one using interchange and coupling arguments. Compared to the standard Pandora's box problem, the main difficulty here is that P-opening introduces more courses of action to consider (e.g., P-open the leading box or alternative ones). Hence, our proof carefully distinguishes between different cases, where the leading box is P-opened and those where it is closed, and proceeds by induction. We need to quantify the value of information gathered by P-opening in order to eventually rule out this option. To this end, we upper bound this quantity using a probabilistic coupling, which exploits the dynamic programming tree locally.

We mention in passing that as part of the proof of the theorem we also show that boxes whose F-threshold is higher than their P-threshold should never be P-opened. The opposite direction does not hold, as our illustrative examples suggest.

3.3. Sufficient conditions for optimal P-opening

We consider the alternative case where $\sigma_M(C, \mathcal{P})$ is a P-threshold. Not surprisingly, in this case, there is no straightforward decision rule that prescribes the optimal action as a function of the opening thresholds. This was illustrated in Section 2.2, even in the simplest case of a single closed-box where the optimal inspection mode could well be an F-opening or P-opening action, depending on the value of y (we provide another example in Appendix A that shows the intricacy of optimal policies). We identify a sufficient condition under which P-opening the leading box is optimal.

DEFINITION 2. Given a state (C, \mathcal{P}, y) and a closed box $i \in C$, we define $\sigma_{-i} = \sigma_M(C \setminus \{i\}, \mathcal{P})$. Suppose $\sigma_i^P > \sigma_i^F$. We say that box i is *well-classified* if $y > \sigma_i^{F/P}$ and for every type $t_i \in \Gamma_i$, one of the following conditions holds: (1) $\sigma_i^{F|t_i} \geq \sigma_{-i}$ or (2) $y > \sigma_i^{F|t_i}$.

To unpack this notion, we remark that, when dealing with a well-classified box i , there is no ambiguity about the action that follows P-opening. Put simply, upon P-opening, the box can be classified as “good” or “bad.” After observing its type $t_i \in \Gamma_i$, our definition delineates two cases. In case (1), we infer from Theorem 2 that it is optimal to immediately F-open the same box (good case), whereas in case (2), we infer from Theorem 1 that box i will never be F-opened (bad case). Importantly, our notions of “good” and “bad” types are preserved in later stages of the decision process, regardless of which actions are taken. As a result, for a well-classified box, it is relatively easier to quantify the value of information afforded by the P-opening.

THEOREM 3. *There exists an optimal policy that P-opens the leading box in every state (C, \mathcal{P}, y) where stopping is not optimal, $\sigma_M(C, \mathcal{P})$ is a P-threshold, and the leading box is well-classified.*

The proof, presented in Appendix B.5, is based on induction using interchange arguments. Similarly to Theorem 2, the analysis requires a probabilistic coupling of different cases.

Unfortunately, combining Theorems 1, 2, and 3 does not determine an optimal policy in every state of the DP. Unlike the optimal stopping condition, Theorems 2-3 provide sufficient but not necessary conditions. Nevertheless, in Section 5.2, we numerically solve synthetic instances and show that these theorems capture optimal decisions in most system states. Theorem 3 is relatively more useful in “later stages” of the decision process, where y tends to be larger, and condition (2) of Definition 2 becomes easier to satisfy. Moreover, our theorems speed up the DP computation. When the conditions of Theorems 1–3 hold, there is no need to apply the DP recursion.

That said, we identify a special case in which Theorems 1–3 fully characterize an optimal policy. This happens when every box i such that $\sigma_i^P > \sigma_i^F$, has only two types (i.e., $|\Gamma_i| = 2$), and the FP-threshold $\sigma_i^{F/P}$ is negative (i.e., $\sigma_i^{F/P} < 0$). To see this, note that whenever stopping is suboptimal, we have two cases: If the largest opening threshold is an F-threshold, it is optimal to F-open the leading box (Theorem 2). If the largest threshold is a P-threshold, then the leading box i is well-classified. Indeed, in view of the discussion that follows Definition 1, we know that $\min_{t_i \in \Gamma_i} \sigma_i^{F|t_i} \leq \sigma_i^{F/P} < 0 \leq y$ and $\max_{t_i \in \Gamma_i} \sigma_i^{F|t_i} \geq \sigma_i^P > \sigma_{-i}$. Intuitively, this setting captures scenarios in which alternatives can be either very good or very bad. If partial opening is cheap enough, it should always be exercised. We summarize this fact in the next corollary.

COROLLARY 1. *Suppose that every box $i \in [N]$ such that $\sigma_i^P > \sigma_i^F$ satisfies $|\Gamma_i| = 2$ and $\sigma_i^{F/P} < 0$. Then, Theorems 1, 2 and 3 fully determine an optimal policy.*

3.4. An optimal policy for a problem with binary prizes

Next, we study a special case of PSIs whereby boxes are statistically identical and independent, contain binary Bernoulli prizes, and belong to one of two types—“good” or “bad.” We refer to this problem as PSI-B2I (Bernoulli, 2-type, identical and independent). This setting applies to situations where a decision maker is searching for a satisfactory alternative among options that are initially perceived as equivalent. Partial inspection reveals whether an alternative is promising (“good”) or has low potential (“bad”), while full inspection discloses the exact outcome. Several variants of this problem in which the decision maker cannot inspect boxes sequentially have been studied in the literature (see, e.g. Doval 2018, and the references therein). Our setting is also different from Clarkson et al. (2020) who assume that there is exactly one prize hidden in the boxes and that opening can be done repeatedly (which does not fit our motivating applications).

In striking contrast to the problem class with binary prizes which is shown to be NP-hard (Section 4.1), the optimal policy to PSI-B2I admits a simple and intuitive structure. To formally describe it, we first remark that the equivalence of boxes implies that their opening thresholds are equal; hence, we will refer to them as σ^F and σ^P . If $\sigma^F \geq \sigma^P$, the optimal policy is fully characterized by Theorems 1 and 2: the optimal policy F-opens closed boxes until a prize is found.

In the difficult case $\sigma^P > \sigma^F$, we denote by n_C, n_G, n_B the number of closed, good and bad boxes, respectively, in the current state. Our main result in this section shows that there exists a threshold N_C on the number of closed boxes so that when $n_C > N_C$, an optimal policy exclusively P-opens closed boxes until stopping or reaching a state in which $n_C \leq N_C$, at which point it exclusively F-opens closed boxes. We note that the threshold N_C is fixed and independent of the state. Moreover, an optimal policy immediately F-opens good-type boxes upon discovery, while postponing bad-type boxes until all closed boxes have been at least partially opened. There are exactly N such policies—corresponding to each value of N_C —that can be easily enumerated and evaluated for the purpose of optimization.

~~This above policy structure~~The structure of the policy above captures the case $\sigma^P \leq \sigma^F$, using the convention $N_C = \infty$. We also observe that the special case where $\sigma^F < \sigma^P$ and $\sigma^{F/P} \leq 0$ is subsumed by Theorem 3 (see Observation 1), in which case $N_C = 0$. Therefore, to ease our analysis, we make **without loss of generality** the following simplifying assumption:

ASSUMPTION 1. *The boxes of PSI-B2I satisfy: $0 < \sigma^{F/P} < \sigma^F < \sigma^P$.*

THEOREM 4. *For a PSI-B2I instance satisfying Assumption 1, there exists an optimal policy π that satisfies the following properties:*

1. *Policy π assigns the highest priority to F-opening good-type boxes.*
2. *Policy π either discards all bad-type boxes, or assigns the lowest priority to F-opening them.*
3. *There exists a finite threshold N_C such that policy π prioritizes P-opening closed boxes over F-opening them iff the number of closed boxes $n_C > N_C$.*

The proof can be found in Appendix B.6. The theorem is presented as an algorithm in Appendix B.7.

Yaron: Thanks Jingwei. What do you think about adding two lines to the algorithm describing the parameters N_C and OpenBad, and how they are computed?

The structure of the optimal policy suggests that the relation between the opening thresholds is the primary factor that determines whether or not P-opening should be considered. If $\sigma^P > \sigma^F$ then P-opening should be considered to identify “promising” boxes. This option is exerted whenever there are sufficiently many a priori equivalent alternatives (as captured by the threshold rule $n_C > N_C$). Intuitively, due to the uncertainty about finding good-type boxes, multiple boxes must be P-opened to obtain with high probability a satisfactory outcome. By contrast, when the total number of boxes is small, the decision-maker will most likely end up opening most of them and therefore P-opening is not beneficial.

Observe that this policy satisfies the optimal characterization given in Theorems 1-3; indeed, the proof in Appendix B.6 directly builds on these properties. Interestingly, although PSI-B2I is a special case, much of the structure exhibited by the optimal policy for this instance extends naturally to the broader class of PSI problems. For example, it is useful to consider a class of policies that commit ex ante to fully or partially opening each closed box before execution; this notion of a committing policy proves useful in the general problem and is discussed in detail in Section 4.3. As another illustration, the instance highlights how an increase in the number of boxes makes partial opening more advantageous, a phenomenon that we further validate numerically in Section 5.4.

Yaron: Jingwei - are there additional connection? To Whittle's integral? to the hardness result?

4. Simple Approximately Optimal Policies

We establish that computing optimal policies for PSI is NP-hard (Section 4.1). We therefore turn our attention to the development and analysis of simple approximate solutions that utilize relaxations that approximate the value function. In Section 4.2, we propose two approximation methods: *Whittle's integral* and the *free-info relaxation*, which adapt tools from related literature. Interestingly, we show that Whittle's integral provides a tighter upper bound, and establish the asymptotic optimality of the index policy based on the opening thresholds. In general, we exploit these problem relaxations in Section 4.3 to devise approximation algorithms, in particular achieving a $(1 - 1/e)$ -approximation ratio, through a reduction to the maximization of stochastic submodular functions (Asadpour and Nazerzadeh 2015, Singla 2018, Beyhaghi and Kleinberg 2019).

4.1. Hardness of the problem

THEOREM 5. *Computing optimal policies for PSI is NP-hard.*

The proof strategy involves a reduction from a family of subset sum problems. Specifically, we construct a hard instance in which every box i has two types $|\Gamma_i| = 2$: “good” and “bad.” The good type is sufficiently valuable that an optimal policy F -opens the box upon discovering it, and bad type boxes are discarded. Hence, determining the optimal policy is essentially equivalent to deciding how to partition the N boxes, i.e., deciding the opening mode for each box. We carefully design this PSI instance so that the partition exactly corresponds to the solution of the subset sum problem. The proof is intricate and appears in Appendix C.6. In particular, formalizing the preceding argument requires controlling the encoding error due to the bit precision of the input.

It is interesting to note that the hard instances we constructed for the proof of Theorem 5 are very close to the one subsumed by Corollary 1, for which we were able to provide a polynomial-time algorithm. Specifically, our family of hard instances satisfies that, every box $i \in [N]$ has $\sigma_i^P > \sigma_i^F$ and $|\Gamma_i| = 2$. The only difference, however, is that $\sigma_i^{F/P} > 0$, contrary to Corollary 1. This suggests that Corollary 1 is “tight” in some sense.

[Yaron: Place-holder for a discussion about similarity to and differences from Fu et al.](#)

4.2. Relaxations and upper bounds

A common method to solve high-dimensional DPs consists in approximating the value function (see, e.g., Chapter 9 in Powell 2007). For example, effective policies are often obtained by solving the approximate Bellman's equation, where the true (unknown) value function is replaced by its

approximate counterpart. In what follows, we construct two natural upper bounds on $J(C, \mathcal{P}, y)$ and compare them analytically.

Whittle's integral. This is a known upper bound on the value function of MAB superprocesses (see, e.g., Brown and Smith 2013). Consider a problem with a single non-fully-open box i where the current maximum prize is y . Let $\omega_i(y)$ be the probability that the current maximum prize y is eventually selected by an optimal policy. In other words, we can think of $\omega_i(y)$ as the probability that the effort of opening box i is in vain, provided that only box i is left. Given the characterization of the optimal policy for the single-box case in Theorem 1, it is not hard to see that, if box $i \in C$,

$$\omega_i(y) = \mathbf{I}\{y > \sigma_i^{F/P}\} \Pr\left[\min\{V_i, \sigma_i^{F|T_i}, \sigma_i^P\} \leq y\right] + \mathbf{I}\{y \leq \sigma_i^{F/P}\} \Pr\left[\min\{V_i, \sigma_i^F\} \leq y\right].$$

Otherwise, if box $i \in \mathcal{P}$,

$$\omega_i(y) = \Pr\left[\min\{V_i, \sigma_i^{F|t_i}\} \leq y \mid T_i = t_i\right].$$

Consequently, Whittle's integral corresponds to the following quantity:

$$J^W(C, \mathcal{P}, y) = \sigma_M(C, \mathcal{P}) - \int_y^{\sigma_M(C, \mathcal{P})} \prod_{i \in C \cup \mathcal{P}} \omega_i(u) du. \quad (11)$$

We will establish in Lemma 1 that this function provides an upper bound on the $J(C, \mathcal{P}, y)$. Conversely, when $y = \sigma_M(C, \mathcal{P})$, we straightforwardly obtain $J(C, \mathcal{P}, y) \geq y \geq \sigma_M(C, \mathcal{P}) \geq J^W(C, \mathcal{P}, y)$, implying that the upper bound (11) can be tight in certain states.

Free-info relaxation. We now consider an alternative upper-bounding method that exploits the interpretation of indices as fair strike prices of two-stage call options. To this end, we introduce indicator random variables that describe the actions chosen by any given policy $\pi \in \Pi$ from state (C, \mathcal{P}, y) .⁷ Let $\mathbb{S}_i^\pi, \tilde{\mathbb{S}}_i^\pi$ be the binary random variables that indicate whether box i is eventually selected without a partial opening and after a partial opening, respectively. (Note that we necessarily have $\mathbb{S}_i^\pi + \tilde{\mathbb{S}}_i^\pi \leq 1$). Similarly, let $\mathbb{F}_i^\pi, \tilde{\mathbb{F}}_i^\pi$ be the binary random variables that indicate whether box i is F-opened without any prior P-opening or after a P-opening, respectively. Finally, let \mathbb{P}_i^π be the binary random variable that indicates whether box i is P-opened. While these decision variables depend on the policy π , for ease of notation, we sometimes suppress the superscript π . With this notation at hand, the expected value of policy π in state (C, \mathcal{P}, y) can be written as

$$J^\pi(C, \mathcal{P}, y) = \mathbb{E}\left[\mathbb{S}_0^\pi \cdot y + \sum_{i \in C} \left(\mathbb{S}_i^\pi V_i - \mathbb{F}_i^\pi c_i^F - \mathbb{P}_i^\pi c_i^P\right) + \sum_{i \in C \cup \mathcal{P}} \left(\tilde{\mathbb{S}}_i^\pi V_i - \tilde{\mathbb{F}}_i^\pi c_i^F\right) \mid (C, \mathcal{P}, y)\right], \quad (12)$$

where, by a slight abuse of notation, the conditional expectation indicates that policy π is executed starting from state (C, \mathcal{P}, y) , and thus, we only accrue the costs of future opening decisions and the rewards of future selections decisions. Hence, the Pandora's box problem with sequential inspections can be formulated as the following mathematical program:

$$J(C, \mathcal{P}, y) = \max_{\pi \in \Pi} J^\pi(C, \mathcal{P}, y) . \quad (13)$$

The free-info relaxation considers an alternative setting, in which there is no cost of P-opening and F-opening, but instead, the rewards of the boxes are capped according to the threshold values. The notion of a feasible policy does not change in this new formulation, only the cost accounting is modified. Specifically, for every box $i \in C$, we define the capped value of F-opening to be $K_i = \min\{V_i, \sigma_i^F\}$ and the capped value of P-opening to be $\tilde{K}_i = \min\{V_i, \sigma_i^{F|T_i}, \sigma_i^P\}$. Similarly, for every box $i \in \mathcal{P}$, we define the capped value of type t_i as $\tilde{K}_i^{t_i} = \min\{\sigma_i^{F|t_i}, V_i\}$. Consequently, in the free-info formulation, the expected reward of policy π in state (C, \mathcal{P}, y) is

$$J^{\pi, K}(C, \mathcal{P}, y) = \mathbb{E} \left[\mathbb{S}_0^\pi \cdot y + \sum_{i \in C} \left(\mathbb{S}_i^\pi K_i + \tilde{\mathbb{S}}_i^\pi \tilde{K}_i \right) + \sum_{i \in \mathcal{P}} \tilde{\mathbb{S}}_i^\pi \tilde{K}_i^{t_i} \middle| (C, \mathcal{P}, y) \right] . \quad (14)$$

Recall that we interpret the caps as fair strike prices of call options on the box values, whose expected payoff offsets the opening costs. The formulation (14) can be viewed as an alternative cost accounting where Pandora opens boxes free of charge (there are no negative terms), and in exchange, there is a short position on a call option that exactly compensates the opening costs. However, Pandora has the advantage of being subjected to the short contract only if the box is ultimately selected; otherwise, the short contract is void. Under this optimistic assumption, Pandora essentially collects one of the capped rewards, K_i , \tilde{K}_i , or $\tilde{K}_i^{t_i}$, depending on the chosen opening mode.

Going one step further, the free-info relaxation in state (C, \mathcal{P}, y) corresponds to the following quantity:

$$J^K(C, \mathcal{P}, y) = \mathbb{E} \left[\max \left\{ y, \max_{i \in C} \left\{ \max \{K_i, \tilde{K}_i\}\right\}, \max_{i \in \mathcal{P}} \left\{ \tilde{K}_i^{t_i} \right\} \right\} \right] . \quad (15)$$

Similarly to Whittle's integral, we eliminate the dependence on policy π by choosing the box with the highest capped reward. Therefore, $J^K(C, \mathcal{P}, y)$ is an upper bound on (14), which might not be attainable through a specific policy π .

We are now ready to state our main result with respect to Whittle's integral and the free-info relaxation, showing that both methods provide valid upper bounds on the value function.

LEMMA 1. *For every state (C, \mathcal{P}, y) , we have*

$$J(C, \mathcal{P}, y) \leq J^W(C, \mathcal{P}, y) \leq J^K(C, \mathcal{P}, y) . \quad (16)$$

The proof appears in Appendix C.1. These upper bounds have been independently derived in the literature for related optimization settings. To our knowledge, this result is the first to formally connect these methods and show that Whittle's integral provides a tighter upper bound. The proof relies on results by Brown and Smith (2013) who exhumed an unnoticed technical claim in Whittle (1980) to derive Whittle's integral (for the cost-discounted case), and Beyhaghi and Kleinberg (2019) who proposed the free-info relaxation; a technical difference in our setting is that information can be acquired sequentially over multiple stages.

We note in passing that Lemma 1 can be used to establish the asymptotic optimality of the index policy that prioritizes the highest threshold inspection. For brevity, we relegate this result to Appendix C.2.

4.3. Near-optimal algorithms

In this section, we leverage the upper bounds of Section 4.2 to construct simple near-optimal policies. Specifically, we develop an algorithm with a worst-case approximation ratio of $(1 - 1/e)$ using a reduction to the stochastic submodular optimization problem.

Our approach is an application of the framework by Singla (2018) and Beyhaghi and Kleinberg (2019). In this context, the main idea is to focus on policies for which the free-info upper bound $J^{\pi, K}$ of equation (14) is exactly equal to the value function J^{π} of equation (13). It turns out there is a simple recipe for constructing such policies: (i) commit before opening any box on the selection mode of each box (i.e., whether we start by F-opening or P-opening it), and (ii) strictly follow the order of decreasing opening thresholds. That is, we partition the set of boxes into two subsets, $[N] = F \cup P$, such that boxes $i \in F$ are initially *F*-opened, whereas boxes $i \in P$ are *P*-opened, if inspected. Next, in each state, we select the action with the highest associated threshold. The resulting class of policies $\pi^{F, P}$ is termed *committing* policies. The reader is referred to the Algorithm 1 for a complete description.

Algorithm 1 Committing policy $\pi^{F,P}$.

```

1: Initialize state  $(C, \mathcal{P}, y) \leftarrow ([N], \emptyset, 0)$ ;
2: while  $y < \max \left\{ \max_{j \in C \cap F} \{\sigma_j^F\}, \max_{j \in C \cap P} \{\sigma_j^P\}, \max_{j \in \mathcal{P}} \{\sigma_j^{F|t_j}\} \right\}$  do
3:   Let  $\sigma_i = \max \left\{ \max_{j \in C \cap F} \{\sigma_j^F\}, \max_{j \in C \cap P} \{\sigma_j^P\}, \max_{j \in \mathcal{P}} \{\sigma_j^{F|t_j}\} \right\}$ ;
4:   if  $i \in F$  then
5:     F-open box  $i$ ;  $y \leftarrow \max\{y, v_i\}$ ;  $C \leftarrow C \setminus \{i\}$ ;
6:   else if  $i \in P$  then
7:     if  $i \in C$  then
8:       P-open box  $i$ ;  $C \leftarrow C \setminus \{i\}$ ;  $\mathcal{P} \leftarrow \mathcal{P} \cup \{(i, t_i)\}$ ;
9:     else if  $i \in \mathcal{P}$  then
10:      F-open box  $i$ ;  $\mathcal{P} \leftarrow \mathcal{P} \setminus \{(i, t_i)\}$ ;  $y \leftarrow \max\{y, v_i\}$ ;
11:    end if
12:  end if
13: end while
14: Select  $y$  and stop;

```

The next lemma shows that the expected profit of any committing policy indeed coincides with its free-info upper bound, which can be expressed in a simple form.

LEMMA 2. *Given a partition (F, P) of the set of boxes $[N]$, the committing policy $\pi^{F,P}$ generates an expected profit of*

$$J^{\pi^{F,P}}(s_0) = J^{\pi^{F,P},K}(s_0) = \mathbb{E} \left[\max \left\{ \max_{i \in F} \{K_i\}, \max_{i \in P} \{\tilde{K}_i\} \right\} \right]. \quad (17)$$

Intuitively, committing policies are not exposed to any cost accounting error, in comparing the opening costs to the corresponding short positions on call options. Recall that this free-info cost accounting is optimistic because Pandora is subjected to the short contract only if the box is ultimately selected. In the proof of Lemma 2, we identify necessary and sufficient conditions such that Pandora always selects any F-opened box (or F-opens any P-opened box) whenever the call option (or two-stage call option) is in the money. This way, Pandora meets the contractual payment of the option, which is equal to the opening costs in expectation. In particular, we show these necessary and sufficient conditions are met by committing policies.

Now, this introduction of such committing policies raises the following question: how should we design the partition (F, P) ? Can we construct committing policies that achieve strong (constant-factor) approximation ratios? Based on the expected profit characterization of Lemma 2, we can compare committing policy $\pi^{F,P}$'s payoff $\mathbb{E}[\max_{i \in F} \{K_i\}, \max_{i \in P} \{\tilde{K}_i\}]$ to the free-info upper bound $J^K(s_0) = \mathbb{E}[\max_{i \in [N]} \max\{K_i, \tilde{K}_i\}]$ (see Equation (15)). The former ex-ante chooses either K_i

or \tilde{K}_i , before observing their realizations, depending on whether $i \in F$ or $i \in P$. By contrast, the latter chooses the maximum reward $\max \{K_i, \tilde{K}_i\}$ out of the two opening modes after seeing their realizations. This informal comparison leads to the design of simple approximation algorithms.

First, note that the following upper bound on the free-info setting always holds:

$$\begin{aligned} J^K(s_0) &= \mathbb{E} \left[\max_{i \in [N]} \max \{K_i, \tilde{K}_i\} \right] \\ &\leq \mathbb{E} \left[\max_{i \in [N]} \left(\mathbb{I}[i \in F] \cdot K_i + \mathbb{I}[i \in P] \cdot \tilde{K}_i \right) + \max_{i \in [N]} \left(\mathbb{I}[i \in P] \cdot K_i + \mathbb{I}[i \in F] \cdot \tilde{K}_i \right) \right] \\ &= J^{\pi^{F,P}}(s_0) + J^{\pi^{P,F}}(s_0), \end{aligned}$$

where the inequality holds since $\max\{a, b\} \leq \max\{c, d\} + \max\{e, f\}$ for every scalars $a \leq \max\{c, e\}$ and $b \leq \max\{d, f\}$, and the last equality proceeds from Lemma 2. Importantly, the above inequality implies that, for any partition of (F, P) of $[N]$, either the committing policy $\pi^{F,P}$ or the one $\pi^{P,F}$ that flips the opening modes achieves a factor- $\frac{1}{2}$ approximation of the optimum.

More broadly, the performance of a well-designed committing policy is related to the adaptivity gap in stochastic submodular maximization; the question can be reformulated as optimizing a certain submodular function over random entries and the goal is to bound the gap between non-adaptive policies, that select those entries before their realizations are revealed, and adaptive ones. By leveraging fundamental results in this area, we obtain the following theorem.

THEOREM 6. *There exists a committing policy that achieves at least $1 - 1/e \approx 0.63$ of the optimum and can be approached via a polynomial-time approximation scheme.*

In Appendix C.5, we devise a reduction from Pandora's box problem to the problem of maximizing a stochastic submodular function under additional constraints. In this context, Asadpour and Nazerzadeh (2015) showed, among other results, that the adaptivity gap (i.e., the performance gap between non-adaptive and adaptive policies) is lower bounded by $1 - 1/e$, and that an approximate non-adaptive policy can be efficiently constructed using a continuous-greedy algorithm.

5. Numerical Study

We conduct numerical experiments to further explore the characteristics of optimal policies and assess the effectiveness of natural candidate policies. Section 5.1 describes the process for generating a diverse set of problem instances. Section 5.2 quantifies the extent to which our theoretical characterization captures optimal decisions. Section 5.3 compares the performance of various policies inspired by theoretical analysis. Finally, in Section 5.4, we examine factors that increase the tendency of the optimal policies to P-open.

5.1. Problem instances

Our design of numerical experiments is based on the observation that the interaction between opening thresholds significantly impacts the behavior of optimal policies. We therefore create instances that consist of boxes with diverse values of opening thresholds.

We begin by generating a collection of “prototypical” boxes. To create a prototypical box i , we sample 5 values for the support of V_i from the uniform distribution $\mathcal{U}(0, 10)$; every type T_i can realize up to at most three types (that is, $|\Gamma_i| \leq 3$); for each type we uniformly sample the conditional distribution $V_i | T_i = t$ over the support; we sample the F-opening and P-opening costs from the uniform distributions $\mathcal{U}(0, 5)$ and $\mathcal{U}(0, 3)$, respectively. We then generate a candidate set of prototypical boxes out of which we select a small representative set (Figure 3).

Each problem instance consists of randomized copies of the prototypical boxes, in varying compositions with $N = 2$ to $N = 16$ boxes. Due to running time limitations, there are 1000 instances whose size is 2 to 9 (“small instances” which can be solved to optimality using DP), and there are 300 instances whose size is 10 to 16 (“large instances” which can only be solved using heuristics). To avoid trivial cases, we only generate boxes where $c_i^F \geq c_i^P$ (which is why in Figure 3 most prototypical boxes have greater P-thresholds).

5.2. Quantifying the coverage of analytical results

The analysis of Section 3 revealed various insights about the nature of the optimal policy. Since the analysis includes sufficient but not always necessary conditions, we are interested in understanding to what extent the sufficient conditions of Theorems 1–3 describe an optimal policy.

To this end, we conduct two types of experiments. First, we examine optimal solutions to the “small instances” and compute the percentage of states where the sufficient conditions of Theorems 1–3 hold (i.e., where our theorems prescribe optimal actions). We implement the DP recursion in (1) and compute *coverage rates*, defined as the percentages of system states where the optimal action can be determined by our sufficient conditions.

The results are summarized in Table 1. The overall coverage rate is presented in the 5th column, and is divided into specific conditions in the 2nd to 4th columns. Looking at the 5th column of the table, we see that as the problem increases in size, the sufficient conditions of Theorems 1–3 capture optimal actions in all but a relatively few states. The last two columns present the recall of Theorems 2 and 3. These reflect the percentage of states in which the sufficient conditions of the theorems hold out of all states where F-opening (column 6) and P-opening (column 7) are optimal.

For example, for instances of size 5 (4th row in the table), Theorem 2 captures 97% of states where F-opening is optimal, and Theorem 3 captures 77% of all states where P-opening is optimal. We observe that identifying P-opening is more challenging than identifying when F-opening is optimal, but overall, the experiment suggests that our characterization explains to a large degree the behavior of the optimal policy. Note that the recall of Theorem 1 is 100% as the theorem perfectly describes when stopping is optimal.

We conduct a second experiment where we compare two DP implementations that recursively compute the optimal policy. The first policy “Naïve DP” is implemented using the DP recursion in Equation (1), which does not leverage any structure of the optimal policy. The second implementation, on the other hand, utilizes the sufficient conditions of Theorems 1–3 which determine optimal actions and do not require calling the DP recursion when the conditions are satisfied. Table 2 compares the two approaches in terms of the total number of states created, the total number of visits to system states, and the runtime. We see that the characterization of the optimal policy not only provides insights about the optimal policy, but also bears computational benefits, allowing to solve problems at significantly shorter times. That said, the “curse of dimensionality” cannot be avoided, as can be observed by the rate at which the size of the state space grows.

5.3. Policy benchmark and computational study

We evaluate and compare various policies that arise in our analysis:

- π^{OPT} – the optimal policy, computed by solving the DP naively.
- π^{F^*, P^*} – the best committing policy, obtained through complete enumeration.
- π^{index} – an index policy defined using the opening thresholds associated with every box and action. In every state, the policy acts according to the largest index, where the thresholds σ_i^F , $\sigma_i^{F|t_i}$, and σ_i^P serve as indices for the F- and P-opening actions, while y is the index for selecting the best opened box. This policy is asymptotically optimal (Corollary EC.6) and satisfies the sufficient optimality conditions.
- π^{W} – a one-step lookahead policy (Bertsekas 2019) based on approximate value functions using Whittle’s integral.⁸
- π^{STP} – a policy that determines its next action based on the maximal myopic improvement (using a single “test,” hence Single Test Policy) in comparison to stopping. Computing the improvement from each non-open box becomes tractable by ignoring other non-open boxes, and follows the optimal policy for the single box problem (see Figure 2).

To evaluate performance, we measure the value function and runtime of each problem instance under different policies, starting from the initial state. As the expected profit can vary significantly among different instances, we normalize the expected profit of each policy relative to the optimal value obtained by policy π^{OPT} . For large instances where the optimal policy cannot be computed, we use the best-performing policy to normalize the expected profit. Thus, the performance of each policy is reported as a number between 0 and 1, where larger values indicate better performance. For each policy and a set of instances, we report on performance using the average across instances and the worst instance tested.

Average performance. All heuristics perform quite well in terms of their average performance, often exceeding 90% of OPT. However, π^{index} and π^{F^*, P^*} are superior, with average performances that are close to 100%.

Worst-case performance. We observe that π^{W} can perform poorly, at times nearing zero profit. This suggests that Whittle's integral approach may not be a suitable approach to deriving tight bounds for our problem, and perhaps to other Pandora's box problems as well. We also see that π^{index} is superior to π^{STP} and that its relative performance improves as the number of boxes increases. We observe that π^{index} approaches optimality as N increases, which is aligned with its asymptotic optimality. Outperforming all others is policy π^{F^*, P^*} , which achieves the best worst-case performance, with a loss of at most 7% for instances of size 2. This suggests that index policies based on our opening thresholds are effective for solving the problem. We note, however, that this policy is not optimal, and finding the optimal committing policy is computationally challenging (an approximation algorithm is provided in Asadpour and Nazerzadeh 2015).

Runtime. Table EC.1 in Appendix D summarizes the runtimes. As expected, the time for solving instances to optimality grows exponentially fast, making it challenging to solve instances with more than 9 boxes. We also see that applying policy π^{W} becomes more challenging as the number of boxes increases, due to the calculation of Whittle's integral. Similarly, the exhaustive enumeration carried out by policy π^{F^*, P^*} could result in long runtimes. We note that the runtime also depends on the size of the support of the distribution (which we assumed to be 5) and the number of box types (which we assumed to be at most 3). Note that these runtimes include the entire calculation of the value function and that simply applying these policies is very fast.

Optimality. Table 4 shows the percentage of instances for which different policies attain the optimal expected objective value. In contrast to the near-optimal performance displayed in Table 3, policy π^{W} is almost never optimal. Additionally, even policies $\pi^{F, P}$ and π^{index} , which perform better,

in many cases do not achieve the optimal objective value. We notice that overall, the performance of the latter two index policies improves with the number of boxes, which aligns with the asymptotic optimality result (Corollary EC.6).

5.4. When is P-opening worthwhile?

We next conduct experiments that shed light on the structure of optimal policies and further characterize when P-opening is worthwhile.

To this end, we define the notion of *P-ratio* as the average percentage of P-openings conducted by the optimal policy of all opening actions. A higher P-ratio implies that the optimal policy applies P-opening more frequently. We compute the P-ratio for a set of problem instances and plot it against various problem characteristics. For brevity, we describe the main results below and postpone the complete details of the experiment and the figures to Appendix E.

One of our main findings is that greater homogeneity of boxes in terms of their opening thresholds (i.e., boxes i and j are “homogeneous” when the tuples (σ_i^F, σ_i^P) and (σ_j^F, σ_j^P) are close) is correlated with more frequent partial opening. This aligns with intuition, since when boxes are very different, their relative value can be determined without partial inspection; hence, F-opening can simply be conducted as a necessity prior to selection.

Additionally, in line with our theoretical results, we observe an increased tendency to partially open when there is a higher proportion of boxes whose P-thresholds exceed their corresponding F-thresholds. This finding underscores the role of opening thresholds in capturing the value of corresponding inspection modes.

Finally, in the case of identical boxes (homogeneity is maximal), the tendency to partially open increases in the number of boxes. This suggests that when the pool of alternatives is larger, partial opening becomes more significant—assuming partial opening is worthwhile to begin with.

6. Conclusions and Future Directions

The work studies a generalized Pandora’s box problem that captures a tradeoff related to acquiring accurate and costly versus inaccurate but cheaper information about the value of competing alternatives. We proved that the problem is NP-hard, but nevertheless, using a synthesis of techniques, we identified structural properties of the optimal policy and provided insights into the key drivers of optimal information acquisition decisions. We showed that intuitive threshold-based policies that extend the Pandora’s box solution describe optimal decisions and form the basis of approximate

solutions. Extensive numerical experiments show that this characterization almost entirely explains the optimal policy, and that the approximate policies perform extremely well.

As our work focused on the simplest yet sufficiently rich context where the aforementioned tradeoff occurs, it would be interesting to consider more general settings. For example, a richer information model that allows multiple types of repeated inspections, inspection costs that decrease after inspections take place, and the option to select partially opened boxes, all could be interesting directions for future work.

Notes

¹We refer to this model as the *standard Pandora's problem*.

²Our approach is inspired by the class of committing algorithms studied by Beyhaghi and Kleinberg (2019), with the notable difference that our setting involves sequential inspections.

³Here, the policy commits to partial opening vs. full opening simply by comparing the corresponding thresholds for each box.

⁴In the standard Pandora's box problem, the objective function is simply the maximum observed prize.

⁵Gittins indices and thresholds in Weitzman's policy are equivalent in the special case of the Pandora's box problem (see, e.g., an early version of Olszewski and Weber 2015).

⁶Throughout, we use capital letters to denote random variables and lowercase letters to denote their realizations.

⁷Recall that, in the state description, y is the reward of a best F -opened box. Without loss of optimality, we restrict attention to policies that discard other F -opened boxes.

⁸Another natural policy to consider is the one-step lookahead policy that uses J^K to approximate the value function. However, given that J^W provides a tighter upper bound (Lemma 1), we only test π^W in the experiments.

References

- Asadpour, A. and Nazerzadeh, H. (2015). Maximizing stochastic monotone submodular functions. *Management Science*, 62(8):2374–2391.
- Asawa, M. and Teneketzis, D. (1996). Multi-armed bandits with switching penalties. *IEEE Transactions on Automatic Control*, 41(3):328–348.
- Attias, C., Krauthgamer, R., Levi, R., and Shaposhnik, Y. (2017). Stochastic selection problems with testing. *Available at SSRN 3076956*.
- Balseiro, S. R. and Brown, D. B. (2019). Approximations to stochastic dynamic programs via information relaxation duality. *Operations Research*, 67(2):577–597.
- Baucells, M. and Zorc, S. (2023). Search in the dark: The normal case.
- Bertsekas, D. (2019). *Reinforcement learning and optimal control*. Athena Scientific.
- Bertsimas, D. and Nino-Mora, J. (1996). Conservation laws, extended polymatroids and multiarmed bandit problems; a polyhedral approach to indexable systems. *Mathematics of Operations Research*, 21(2):257–306.
- Beyhaghi, H. (2019). *Approximately-Optimal Mechanisms in Auction Design, Search Theory, and Matching Markets*. PhD thesis, Cornell University.
- Beyhaghi, H. and Kleinberg, R. (2019). Pandora's problem with nonobligatory inspection. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 131–132.
- Boodaghians, S., Fusco, F., Lazos, P., and Leonardi, S. (2020). Pandora's box problem with order constraints. In *ACM Conference on Economics and Computation (EC'20)*, page 439–458. Association for Computing Machinery.

- Brent, R. P. (1976). Fast multiple-precision evaluation of elementary functions. *Journal of the ACM (JACM)*, 23(2):242–251.
- Brown, D. B. and Smith, J. E. (2013). Optimal sequential exploration: Bandits, clairvoyants, and wildcats. *Operations Research*, 61(3):644–665.
- Choi, M., Dai, A. Y., and Kim, K. (2018). Consumer search and price competition. *Econometrica*, 86(4):1257–1281.
- Chu, L. Y., Nazerzadeh, H., and Zhang, H. (2020). Position ranking and auctions for online marketplaces. *Management Science*.
- Clarkson, J., Glazebrook, K. D., and Lin, K. Y. (2020). Fast or slow: Search in discrete locations with two search modes. *Operations Research*.
- De los Santos, B., Hortacısu, A., and Wildenbeest, M. R. (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, 102(6):2955–80.
- Derakhshan, M., Golrezaei, N., Manshadi, V., and Mirrokni, V. (2020). Product ranking on online platforms. In *Proceedings of the 21st ACM Conference on Economics and Computation*, EC '20, page 459. Association for Computing Machinery.
- Doval, L. (2018). Whether or not to open pandora's box. *Journal of Economic Theory*, 175:127–158.
- Eick, S. G. (1988). Gittins procedures for bandits with delayed responses. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(1):125–132.
- Gabaix, X., Laibson, D., Moloche, G., and Weinberg, S. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review*, 96(4):1043–1068.
- Gibbard, P. (2022). A model of search with two stages of information acquisition and additive learning. *Management science*, 68(2):1212–1217.
- Gittins, J., Glazebrook, K., and Weber, R. (2011). *Multi-armed bandit allocation indices*. John Wiley & Sons.
- Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):148–164.
- Immorlica, N., Leshno, J., Lo, I., and Lucier, B. (2020). Information acquisition in matching markets: The role of price discovery. *Available at SSRN*.
- Kleinberg, R., Waggoner, B., and Weyl, E. G. (2016). Descending price optimally coordinates search. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 23–24.
- Miller, R. A. (1984). Job matching and occupational choice. *Journal of Political Economy*, 92(6):1086–1120.
- Olszewski, W. and Weber, R. (2015). A more general pandora rule? *Journal of Economic Theory*, 160:429–437.
- Pandey, S., Chakrabarti, D., and Agarwal, D. (2007). Multi-armed bandit problems with dependent arms. In *Proceedings of the 24th International Conference on Machine Learning*, pages 721–728.
- Powell, W. B. (2007). *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. John Wiley & Sons.
- Robert, J. and Stahl, D. O. (1993). Informative price advertising in a sequential search model. *Econometrica: Journal of the Econometric Society*, pages 657–686.
- Singla, S. (2018). The price of information in combinatorial optimization. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2523–2532. SIAM.
- Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, 37(4):530–552.
- Vishwanath, T. (1992). Parallel search for the best alternative. *Economic Theory*, pages 495–507.
- Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society*, pages 641–654.
- Whittle, P. (1980). Multi-armed bandits and the gittins index. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):143–149.

Wolinsky, A. (1986). True monopolistic competition as a result of imperfect information. *The Quarterly Journal of Economics*, 101(3):493–511.

Tables

Table 1 Average weighted coverage rates of Theorems 1–3 and recall rates of Theorems 2 and 3. Theorem 1 describes the optimal stopping criterion. Theorems 2 and 3 provide sufficient condition for F-opening and P-opening, respectively.

N	Coverage rate				Recall	
	Thm. 1	Thm. 2	Thm. 3	Overall	Thm. 2	Thm. 3
2	0.48	0.38	0.06	0.92	0.90	0.53
3	0.47	0.38	0.09	0.94	0.95	0.66
4	0.46	0.39	0.10	0.95	0.97	0.72
5	0.46	0.39	0.11	0.95	0.97	0.77
6	0.45	0.40	0.11	0.96	0.97	0.81
7	0.45	0.40	0.11	0.96	0.97	0.83
8	0.44	0.41	0.11	0.96	0.96	0.86
9	0.44	0.41	0.11	0.96	0.96	0.87

Table 2 A comparison between a DP implementation that utilizes the sufficient optimality conditions and one that does not, in terms of the average number of states created and revisited, and average runtimes (in seconds).

N	Naïve DP			DP that utilizes Theorems 1–3			runtime ratio
	# states created	# states revisited	runtime (s)	# states created	# states revisited	runtime (s)	
2	38.4	145.0	0.04	24.8	32.37	0.03	0.731
3	209.0	1384.0	0.01	115.3	195.09	0.00	0.225
4	1101.4	10513.9	0.10	575.1	1204.03	0.02	0.180
5	5567.0	69796.4	0.64	2798.7	6703.55	0.10	0.156
6	28158.6	440669.7	3.99	12943.3	35306.54	0.51	0.128
7	137652.0	2570720.3	23.59	65826.3	208193.38	3.00	0.121
8	680310.1	14825094.6	143.80	298534.1	1067723.38	15.93	0.106
9	3238243.1	80398698.6	810.50	1368893.5	5450797.57	82.81	0.094

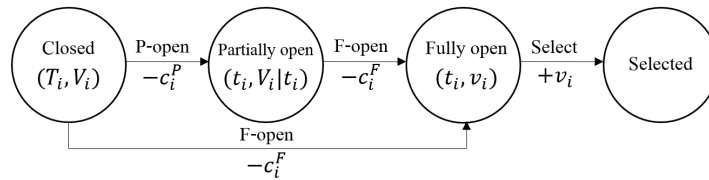
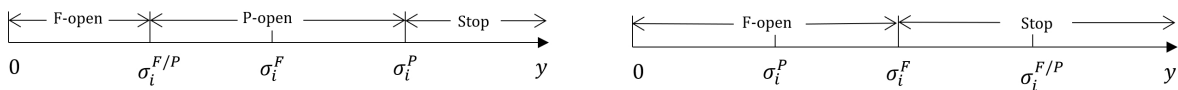
Table 3 Normalized performance of different policies.

N	π^{OPT}			π^{index}			π^{W}			π^{F^*, P^*}			π^{STP}		
	mean	std	worst	mean	std	worst	mean	std	worst	mean	std	worst	mean	std	worst
2	1.000	0.000	1.000	0.987	0.036	0.605	0.922	0.123	0.033	0.999	0.005	0.933	0.963	0.057	0.709
3	1.000	0.000	1.000	0.995	0.017	0.804	0.901	0.101	0.298	1.000	0.002	0.960	0.947	0.062	0.601
4	1.000	0.000	1.000	0.998	0.007	0.882	0.893	0.088	0.403	1.000	0.001	0.976	0.943	0.063	0.611
5	1.000	0.000	1.000	0.999	0.003	0.958	0.890	0.082	0.373	1.000	0.000	0.993	0.938	0.062	0.652
6	1.000	0.000	1.000	0.999	0.002	0.966	0.890	0.085	0.438	1.000	0.000	0.993	0.940	0.059	0.691
7	1.000	0.000	1.000	0.999	0.002	0.984	0.888	0.084	0.503	1.000	0.000	0.996	0.933	0.063	0.575
8	1.000	0.000	1.000	0.999	0.002	0.989	0.879	0.088	0.494	1.000	0.000	0.996	0.937	0.061	0.564
9	1.000	0.000	1.000	0.999	0.001	0.991	0.877	0.090	0.522	1.000	0.000	0.997	0.935	0.061	0.653
10	-	-	-	0.999	0.001	0.992	0.876	0.099	0.515	1.000	0.000	1.000	0.941	0.061	0.709
11	-	-	-	0.999	0.001	0.993	0.871	0.110	0.463	1.000	0.000	1.000	0.939	0.063	0.686
12	-	-	-	0.999	0.001	0.994	0.870	0.104	0.451	1.000	0.000	1.000	0.940	0.061	0.693
13	-	-	-	0.999	0.001	0.992	0.869	0.099	0.568	1.000	0.000	1.000	0.945	0.055	0.710
14	-	-	-	0.999	0.001	0.996	0.877	0.102	0.464	1.000	0.000	1.000	0.940	0.058	0.691
15	-	-	-	0.999	0.001	0.996	0.847	0.136	0.512	1.000	0.000	1.000	0.957	0.064	0.779
16	-	-	-	0.999	0.001	0.997	0.884	0.096	0.633	1.000	0.000	1.000	0.944	0.055	0.780

Table 4 Percentage of instances in which the policies attained the optimal objective value.

N	π^{OPT}	π^{index}	π^{W}	π^{F^*, P^*}
2	1.000	0.543	0.238	0.835
3	1.000	0.614	0.065	0.850
4	1.000	0.622	0.023	0.853
5	1.000	0.661	0.002	0.892
6	1.000	0.686	0.001	0.904
7	1.000	0.701	0.000	0.927
8	1.000	0.715	0.001	0.944
9	1.000	0.733	0.000	0.943

Figures

**Figure 1** Illustration of the dynamics of a single box.**Figure 2** Illustration of the optimal policy for the single closed box case. Here, we present the case where $\sigma_i^{F/P} \geq 0, \sigma_i^P \geq 0$.

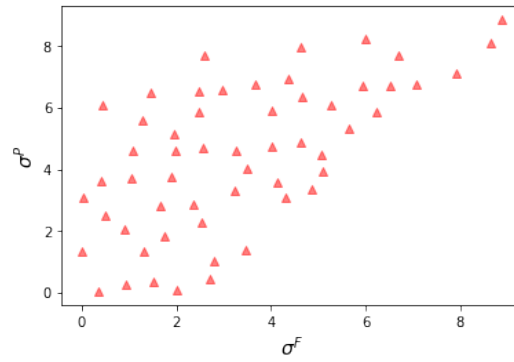


Figure 3 A visualization of how the opening thresholds of the prototypical boxes are distributed.

Appendix A: Example EC.1

We construct an additional example which provides further evidence about the intricacy of the optimal policy in states where the leading threshold corresponds to a P-threshold (see Figure EC.1 for a pictorial illustration).

EXAMPLE EC.1. We consider the following three boxes:

- Box 1: The opening costs are $c_1^F = 1$ and $c_1^P = 0.5$. There are two types $\Gamma_1 = \{A, B\}$ and $\Pr[T_1 = A] = 0.2$. The joint distribution of (V_1, T_1) is given by: $\Pr[V_1 = 1 | T_1 = A] = 0.1$, $\Pr[V_1 = 10 | T_1 = A] = 0.9$ and $\Pr[V_1 = 1 | T_1 = B] = 0.05$, $\Pr[V_1 = 10 | T_1 = B] = 0.95$.
- Box 2: The opening costs are $c_2^F = 0.99$ and $c_2^P = 500$. There are two types $\Gamma_2 = \{A, B\}$ and $\Pr[T_2 = A] = 0.2$. The joint distribution of (V_2, T_2) is given by: $\Pr[V_2 = 1 | T_2 = A] = 0.1$, $\Pr[V_2 = 10 | T_2 = A] = 0.9$ and $\Pr[V_2 = 1 | T_2 = B] = 0.05$, $\Pr[V_2 = 10 | T_2 = B] = 0.95$.
- Box 3: The opening costs are $c_3^F = 1$ and $c_3^P = 0.5$. There are two types $\Gamma_3 = \{A, B\}$ and $\Pr[T_3 = A] = 0.2$. The joint distribution of (V_3, T_3) is given by: $\Pr[V_3 = 1 | T_3 = A] = 0.1$, $\Pr[V_3 = 10 | T_3 = A] = 0.9$ and $\Pr[V_3 = 1 | T_3 = B] = 0.05$, $\Pr[V_3 = 10 | T_3 = B] = 0.95$.

The thresholds of the boxes are computed numerically and are equal to: $\sigma_1^F = \sigma_3^F \approx 5.45$, $\sigma_1^P = \sigma_3^P \approx 6.11$, $\sigma_2^F \approx 5.50$, $\sigma_2^P < 0$. The thresholds are visualized in Figure EC.1.

Depending on the initial state, the unique optimal policy proceeds as follows:

- In the initial state of the problem where the only box is box 1, i.e., in state $(\{1\}, \emptyset, 0)$, the optimal policy F-opens box 1.
- In the initial state of the problem where there are two boxes, box 1 and box 2, i.e., in state $(\{1, 2\}, \emptyset, 0)$, the optimal policy F-opens box 2.
- In the initial state of the problem where there are three boxes: box 1, box 2, and box 3, i.e., in state $(\{1, 2, 3\}, \emptyset, 0)$, the optimal policy P-opens box 1.

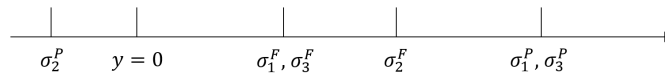


Figure EC.1 Illustration of thresholds in Example EC.1

As evidenced by Example EC.1, the unique optimal policy is not indexable. More importantly, and unlike Example 2, it shows that the opening priorities may change when boxes are added,

even when the value of y remains unchanged: in state $(\{1\}, \emptyset, 0)$, the action of F-opening box 1 is optimal; after adding box 2 (i.e., in state $(\{1, 2\}, \emptyset, 0)$), F-opening box 2 becomes optimal, even if the P-threshold of box 1 remains the largest; however, after adding box 3 (which is statistically identical to box 1), the optimal action is to P-open box 1. In other words, this example shows three states in which σ_1^P is the largest threshold but three distinct actions are in fact optimal.

Appendix B: Proofs for Section 3

Algorithm 2 below summarizes the optimal policy for the single closed-box setting, which is described in Section 3.1.

Algorithm 2 The optimal policy for the single closed-box setting.

```

1: Initialize state  $(C, \mathcal{P}, y) \leftarrow (\{i\}, \emptyset, y)$ ;
2: if  $y > \max\{\sigma_i^F, \sigma_i^P\}$  then
3:   Stop and select  $y$ ;
4: else if  $\sigma_i^F > \max\{y, \sigma_i^P\}$  then
5:   F-open box  $i$ , and select the best prize at hand;
6: else if  $\sigma_i^P > \max\{y, \sigma_i^F\}$  then
7:   if  $y > \sigma_i^{F/P}$  then
8:     P-open box  $i$ ;
9:     Then further F-open box  $i$  if and only if the realized type  $t_i$  is such that  $y < \sigma_i^{F|t_i}$  or else stop;
10:  else
11:    F-open box  $i$ , and select the best prize at hand;
12:  end if
13: end if

```

In what follows we first prove the optimality of the threshold-based policy for the single closed-box case (Section B.1), then prove the expression for the expected profit of the optimal policy in the single closed-box case (Section B.2), and finally, prove the optimal stopping condition for the general case (Section B.3).

B.1. The optimality of the threshold-based policy for the single closed-box case

In state $(C, \mathcal{P}, y) = (\{i\}, \emptyset, y)$, there are three possible actions that we can take: stop, F-open box i or P-open box i . The optimal stopping rule is argued in Section 3.1. We proceed to compare F-opening and P-opening. If we F-open the box, then the expected value function is

$$J^F(\{i\}, \emptyset, y) = -c_i^F + \mathbb{E}_{V_i}[\max\{y, V_i\}] . \quad (\text{EC.1})$$

If we P-open the box, we then need to choose between further F-opening and stopping. Therefore, the value function of this course of action is

$$J^P(\{i\}, \emptyset, y) = -c_i^P + \mathbb{E}_{T_i} \left[\max \left\{ y, -c_i^F + \mathbb{E}_{V_i|T_i}[\max\{y, V_i\}] \right\} \right] . \quad (\text{EC.2})$$

Let $\Lambda_i(y)$ be the set of “promising” types of box i where F-opening is warranted, that is

$$\Lambda_i(y) := \{t_i \in \Gamma_i \mid \sigma_i^{F|t_i} > y\} = \{t_i \in \Gamma_i \mid -c_i^F + \mathbb{E}[(V_i - y)^+ \mid T_i = t_i] > 0\} .$$

The difference between F-opening (EC.1) and P-opening and (EC.2) can be written as

$$\begin{aligned} & J^F(\{i\}, \emptyset, y) - J^P(\{i\}, \emptyset, y) \\ &= -c_i^F + \mathbb{E}[\max\{y, V_i\}] - \left(-c_i^P + \sum_{T_i \in \Lambda_i(y)} \Pr[T_i = t_i] \left(-c_i^F + \mathbb{E}[\max\{y, V_i\} \mid T_i = t_i] \right) + \sum_{T_i \notin \Lambda_i(y)} \Pr[T_i = t_i] y \right) \\ &= \sum_{T_i \in \Lambda_i(y)} \Pr[T_i = t_i] \left(-c_i^F + \mathbb{E}[\max\{y, V_i\} \mid T_i = t_i] \right) + \sum_{T_i \notin \Lambda_i(y)} \Pr[T_i = t_i] \left(-c_i^F + \mathbb{E}[\max\{y, V_i\} \mid T_i = t_i] \right) \\ &+ c_i^P + \sum_{T_i \in \Lambda_i(y)} \Pr[T_i = t_i] \left(c_i^F - \mathbb{E}[\max\{y, V_i\} \mid T_i = t_i] \right) - \sum_{T_i \notin \Lambda_i(y)} \Pr[T_i = t_i] y \\ &= c_i^P + \sum_{T_i \notin \Lambda_i(y)} \Pr[T_i = t_i] \left(\underbrace{-c_i^F + \mathbb{E}[(V_i - y)^+ \mid T_i = t_i]}_{\leq 0} \right) . \end{aligned} \quad (\text{EC.3})$$

Using simple arithmetics, one can show that $J^F(\{i\}, \emptyset, y) - J^P(\{i\}, \emptyset, y)$ is strictly decreasing in y . Moreover, the solution to $J^F(\{i\}, \emptyset, y) - J^P(\{i\}, \emptyset, y) = 0$ is exactly $y = \sigma_i^{F/P}$. The optimality of the threshold-based policy for the single-box case then follows from the optimal stopping and the condition about the dominance between F-opening and P-opening, which is determined by the relation between y and $\sigma_i^{F/P}$.

B.2. The expected profit of the optimal policy in the single closed-box case

In what follows, we calculate the expected profit of this optimal policy in closed form.

1. First, if the policy does not stop and F-opens box i , the expected profit is

$$\begin{aligned} & -c_i^F + \Pr[V_i > \sigma_i^F] \mathbb{E}[V_i | V_i > \sigma_i^F] + \Pr[V_i \leq \sigma_i^F] \mathbb{E}[\max\{y, V_i\} | V_i \leq \sigma_i^F] \\ &= \Pr[V_i > \sigma_i^F] \sigma_i^F + \Pr[V_i \leq \sigma_i^F] \mathbb{E}[\max\{y, V_i\} | V_i \leq \sigma_i^F] \\ &= \mathbb{E}[\max\{y, \min\{V_i, \sigma_i^F\}\}], \end{aligned}$$

where the first equation follows from the definition of F-threshold: $c_i^F = \mathbb{E}[(V_i - \sigma_i^F)^+] = \Pr[V_i > \sigma_i^F] (\mathbb{E}[V_i | V_i > \sigma_i^F] - \sigma_i^F)$. Note the terms can be rearranged so that

$$-c_i^F + \Pr[V_i > \sigma_i^F] \mathbb{E}[V_i | V_i > \sigma_i^F] = \Pr[V_i > \sigma_i^F] \sigma_i^F.$$

2. By the same reasoning, following from the definition of the P-threshold σ_i^P ,

$$\begin{aligned} c_i^P &= \mathbb{E}[\max\{0, -c_i^F + \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i]\}] \\ &= \sum_{t_i: \sigma_i^{F|t_i} > \sigma_i^P} \Pr[T_i = t_i] \left(-c_i^F + \Pr[V_i > \sigma_i^P | T_i = t_i] (\mathbb{E}[V_i | V_i > \sigma_i^P, T_i = t_i] - \sigma_i^P) \right). \end{aligned}$$

Rearranging terms on the right side to the left of the latter equation yields that

$$\begin{aligned} & -c_i^P + \sum_{t_i: \sigma_i^{F|t_i} > \sigma_i^P} \Pr[T_i = t_i] \left(-c_i^F + \Pr[V_i > \sigma_i^P | T_i = t_i] \mathbb{E}[V_i | V_i > \sigma_i^P, T_i = t_i] \right) \\ &= \sum_{t_i: \sigma_i^{F|t_i} > \sigma_i^P} \Pr[T_i = t_i] \Pr[V_i > \sigma_i^P | T_i = t_i] \sigma_i^P. \end{aligned} \tag{EC.4}$$

Therefore, if the policy does not stop and P-opens box i first, the expected profit is

$$\begin{aligned} & -c_i^P + \sum_{t_i: \sigma_i^{F|t_i} > \sigma_i^P} \Pr[T_i = t_i] \left(-c_i^F + \Pr[V_i > \sigma_i^P | T_i = t_i] \mathbb{E}[V_i | V_i > \sigma_i^P, T_i = t_i] \right. \\ & \quad \left. + \Pr[V_i \leq \sigma_i^P | T_i = t_i] \mathbb{E}[\max\{y, V_i\} | V_i \leq \sigma_i^P, \sigma_i^{F|t_i} > \sigma_i^P] \right) \\ & + \sum_{t_i: \sigma_i^{F|t_i} \leq \sigma_i^P} \Pr[T_i = t_i] \left(\mathbf{I}\{y > \sigma_i^{F|t_i}\} \cdot y + \mathbf{I}\{y \leq \sigma_i^{F|t_i}\} \left(-c_i^F + \mathbb{E}[\max\{y, V_i\} | y \leq \sigma_i^{F|t_i}, \sigma_i^{F|t_i} \leq \sigma_i^P] \right) \right) \\ &= \sum_{t_i: \sigma_i^{F|t_i} > \sigma_i^P} \Pr[T_i = t_i] \left(\Pr[V_i > \sigma_i^P | T_i = t_i] \sigma_i^P + \Pr[V_i \leq \sigma_i^P | T_i = t_i] \mathbb{E}[\max\{y, V_i\} | V_i \leq \sigma_i^P, \sigma_i^{F|t_i} > \sigma_i^P] \right) \\ & + \sum_{t_i: \sigma_i^{F|t_i} \leq \sigma_i^P} \Pr[T_i = t_i] \left(\mathbf{I}\{y > \sigma_i^{F|t_i}\} \cdot y + \mathbf{I}\{y \leq \sigma_i^{F|t_i}\} \mathbb{E}[\max\{y, \min\{V_i, \sigma_i^{F|t_i}\}\} | y \leq \sigma_i^{F|t_i}, \sigma_i^{F|t_i} \leq \sigma_i^P] \right) \\ &= \mathbb{E}[\max\{y, \min\{V_i, \sigma_i^{F|T_i}, \sigma_i^P\}\}]. \end{aligned}$$

Here, the first equality follows by introducing (EC.4) and using the definition of F-threshold again. Indeed, for a type t_i such that $y \leq \sigma_i^{F|t_i} \leq \sigma_i^P$, $c_i^F = \Pr \left[V_i > \sigma_i^{F|t_i} \right] \mathbb{E} \left[V_i - \sigma_i^{F|t_i} \mid V_i > \sigma_i^{F|t_i}, y \leq \sigma_i^{F|t_i} \leq \sigma_i^P \right]$. Therefore, we have

$$\begin{aligned}
& -c_i^F + \mathbb{E} \left[\max\{y, V_i\} \mid y \leq \sigma_i^{F|t_i}, \sigma_i^{F|t_i} \leq \sigma_i^P \right] \\
&= -\Pr \left[V_i > \sigma_i^{F|t_i} \right] \mathbb{E} \left[V_i - \sigma_i^{F|t_i} \mid V_i > \sigma_i^{F|t_i}, y \leq \sigma_i^{F|t_i} \leq \sigma_i^P \right] \\
&\quad + \Pr \left[V_i > \sigma_i^{F|t_i} \right] \mathbb{E} \left[V_i \mid V_i > \sigma_i^{F|t_i}, y \leq \sigma_i^{F|t_i}, \sigma_i^{F|t_i} \leq \sigma_i^P \right] + \Pr \left[V_i \leq \sigma_i^{F|t_i} \right] \mathbb{E} \left[\max\{y, V_i\} \mid V_i \leq \sigma_i^{F|t_i}, y \leq \sigma_i^{F|t_i}, \sigma_i^{F|t_i} \leq \sigma_i^P \right] \\
&= \Pr \left[V_i > \sigma_i^{F|t_i} \right] \mathbb{E} \left[\sigma_i^{F|t_i} \mid V_i > \sigma_i^{F|t_i}, y \leq \sigma_i^{F|t_i} \leq \sigma_i^P \right] + \Pr \left[V_i \leq \sigma_i^{F|t_i} \right] \mathbb{E} \left[\max\{y, V_i\} \mid V_i \leq \sigma_i^{F|t_i}, y \leq \sigma_i^{F|t_i}, \sigma_i^{F|t_i} \leq \sigma_i^P \right] \\
&= \mathbb{E} \left[\max\{y, \min\{V_i, \sigma_i^{F|t_i}\}\} \mid y \leq \sigma_i^{F|t_i}, \sigma_i^{F|t_i} \leq \sigma_i^P \right].
\end{aligned}$$

B.3. Optimal stopping

We proceed to prove the optimal stopping rule in the general case. The overall proof strategy is to show that the stopping decision regarding a box (i.e., never open this box in the future) is independent of the existence of other boxes. Then, the optimal stopping rule shown in Section 3.1 for the single box setting can be applied to the general case.

We call a closed box i “F-expired” if $y \geq \max \sigma_i^F$. We call a partially open box i “F-expired” if $y \geq \sigma_i^{F|t_i}$. We call a closed box i P-expired if $y \geq \sigma_i^P$. In what follows, we show that there exists an optimal policy that never F-opens an F-expired box and P-opens a P-expired box.

Case 1: F-expired box. First, we show that an optimal policy never F-opens an F-expired box. The proof proceeds by contradiction. Suppose that in state (C, \mathcal{P}, y) the optimal policy π F-opens an F-expired box i . We define the information set I_i of box i to be $I_i \triangleq \emptyset$ if box i is closed and $I_i \triangleq \{T_i = t_i\}$ if box i is partially opened. By the Bellman equation, we have

$$J^\pi(C, \mathcal{P}, y) = -c_i^F + \mathbb{E}[J^\pi(C', \mathcal{P}', \max\{y, V_i\}) \mid I_i],$$

where $C' = C \setminus \{i\}$, $\mathcal{P}' = \mathcal{P}$ if $i \in C$ and $C' = C$, $\mathcal{P}' = \mathcal{P} \setminus \{i\}$ if $i \in \mathcal{P}$.

We construct an alternative policy π' that does not open box i ; instead, it “simulates” a value for V_i , discards box i for the remainder of the inspection process, and imitates the policy π after that, with the only difference that π' selects y whenever policy π selects box i . With this definition at hand, it is not difficult to show that the expected profit under policy π' is given by

$$J^{\pi'}(C, \mathcal{P}, y) = \mathbb{E}[J^\pi(C', \mathcal{P}', y) \mid I_i].$$

The remainder of the proof proceeds by coupling the realizations of V_i under the two policies to compare their expected profits. For any sample path ρ , we denote the realization of V_i by V_i^ρ . Then, the difference in the value functions of the two policies is

$$\begin{aligned} J_\rho^\pi(C, \mathcal{P}, y) - J_\rho^{\pi'}(C, \mathcal{P}, y) &= \begin{cases} -c_i^F & \text{if box } i \text{ is not selected} \\ -c_i^F + \max\{y, V_i^\rho\} - y & \text{if box } i \text{ is selected} \end{cases} \\ &\leq \begin{cases} -c_i^F + \max\{y, V_i^\rho\} - y & \text{if box } i \text{ is not selected} \\ -c_i^F + \max\{y, V_i^\rho\} - y & \text{if box } i \text{ is selected} \end{cases}. \end{aligned}$$

Therefore, by taking expectations on both sides, we obtain

$$J^\pi(C, \mathcal{P}, y) - J^{\pi'}(C, \mathcal{P}, y) = \mathbb{E}_\rho [J_\rho^\pi(C, \mathcal{P}, y) - J_\rho^{\pi'}(C, \mathcal{P}, y)] \leq \mathbb{E} [-c_i^F + \max\{y, V_i\} - y \mid I_i] .$$

Hence, other boxes do not affect the stopping decision regarding box i . Recalling the optimal stopping policy in the single box setting, one can easily verify that $\mathbb{E} [-c_i^F + \max\{y, V_i\} - y \mid I_i] \leq 0$ if box i is F-expired. We conclude that there exists an optimal policy that does not F-open an F-expired box.

Case 2: P-expired box. Next, we prove that there exists an optimal policy that never P-opens a P-expired box. Suppose that in state (C, \mathcal{P}, y) , the optimal policy π P-opens a box i with $y \geq \sigma_i^P$. We denote its value function by $J^\pi(C, \mathcal{P}, y)$. Hence, we have

$$J^\pi(C, \mathcal{P}, y) = -c_i^P + \mathbb{E}[J^\pi(C \setminus \{i\}, \mathcal{P} \cup \{(i, T_i)\}, y)].$$

Similar to Case 1, we construct an alternative policy π' that “simulates” a type for T_i , discards box i for the remainder of the inspection process, and imitates the policy π after that, with the only difference that π' selects y whenever policy π selects the box i . For any sample path ρ , we denote the realization of T_i and V_i by t_i^ρ and V_i^ρ , respectively. Then the difference between the expected profits of the two policies is

$$\begin{aligned} &J_\rho^\pi(C, \mathcal{P}, y) - J_\rho^{\pi'}(C, \mathcal{P}, y) \\ &= \begin{cases} -c_i^P & \text{if box } i \text{ is not F-opened by } \pi \\ -c_i^P - c_i^F & \text{if box } i \text{ is F-opened, but not selected by } \pi \\ -c_i^P - c_i^F + \max\{V_i^\rho, y\} - y & \text{if box } i \text{ is F-opened, and selected by } \pi \end{cases} \\ &\leq \begin{cases} -c_i^P & \text{if box } i \text{ is not F-opened by } \pi \\ -c_i^P - c_i^F + \max\{V_i^\rho, y\} - y & \text{if box } i \text{ is F-opened, but not selected by } \pi \\ -c_i^P - c_i^F + \max\{V_i^\rho, y\} - y & \text{if box } i \text{ is F-opened, and selected by } \pi \end{cases} . \quad (\text{EC.5}) \end{aligned}$$

We denote by $q_i^\pi(t_i^\rho) \in [0, 1]$ the probability that policy π F-opens box i when the revealed type is t_i^ρ , and we define the set of “promising” types $\Lambda_i(y) = \{t_i \in \Gamma_i \mid y < \sigma_i^{F|t_i}\}$ which warrant F-opening.

We next write an explicit expression for the expectation of the difference, splitting to cases based on (EC.5) and obtain the following:

$$\begin{aligned}
 J^\pi(C, \mathcal{P}, y) - J^{\pi'}(C, \mathcal{P}, y) &= \mathbb{E}_\rho [J_\rho^\pi(C, \mathcal{P}, y) - J_\rho^{\pi'}(C, \mathcal{P}, y)] \\
 &\leq -c_i^P + \sum_{t_i^\rho \in \Lambda_i(y)} \Pr[T_i = t_i^\rho] q_i^\pi(t_i^\rho) \left(\underbrace{-c_i^F + \mathbb{E}[(V_i - y)^+ \mid T_i = t_i^\rho]}_{>0} \right) \\
 &\quad + \left(\sum_{t_i^\rho \notin \Lambda_i(y)} \Pr[T_i = t_i^\rho] + \sum_{t_i^\rho \in \Lambda_i(y)} \Pr[T_i = t_i^\rho] (1 - q_i^\pi(t_i^\rho)) \right) \cdot 0 \quad (\text{EC.6}) \\
 &\leq -c_i^P + \sum_{t_i^\rho \in \Lambda_i(y)} \Pr[T_i = t_i^\rho] \left(\underbrace{-c_i^F + \mathbb{E}[(V_i - y)^+ \mid T_i = t_i^\rho]}_{>0} \right). \quad (\text{EC.7})
 \end{aligned}$$

Inequality (EC.6) is based on (EC.5), multiplied by the conditional probability of type T_i .

We showed that there exists an optimal policy that stops when all boxes F-expire and P-expire. On the other hand, if there is even a single box that did not F-expire or P-expire, opening such box is better than stopping. The latter follows from the single-box case. Therefore, the stopping condition is both sufficient and necessary.

B.4. Proof of Theorem 2

The proof consists of two parts, dealing with cases where the box that attains the largest F-opening threshold is either partially open or closed. We prove these cases as separate lemmas.

LEMMA EC.1. *For every state (C, \mathcal{P}, y) where it is suboptimal to stop, if the F-threshold of a partially opened box k is the largest opening threshold among boxes in that state, then there exists an optimal policy that F-opens box k at that state.*

Proof: The lemma applies to cases in which there is at least one partially opened box, implying $|\mathcal{P}| \geq 1$. We prove the lemma by induction on the maximal potential number of box openings from a given state, which is captured by the expression: $2|C| + |\mathcal{P}|$ (each partially opened box can be opened at most once, while closed boxes can be opened at most twice, first being P-opened and then F-opened). We denote by k the index of the partially open box whose F-threshold is the

largest opening threshold in state (C, \mathcal{P}, y) . With a slight abuse of notation, we denote by σ_k the conditional F-threshold of box k :

$$\sigma_k \triangleq \sigma_k^{F|I_k} = \sigma_M(C, \mathcal{P}).$$

Since stopping is assumed to be sub-optimal, the optimal stopping condition implies that $y < \sigma_k$.

As base cases, consider states where $2 \mid C \mid + \mid \mathcal{P} \mid \in \{1, 2\}$. Since $\mid \mathcal{P} \mid \geq 1$, this implies that $\mid C \mid = 0$. We observe that whenever $\mid C \mid = 0$ the problem reduces to the standard Pandora's box problem, from which the optimality of the lemma follows.

For the induction step, suppose that the lemma holds in states where $2 \mid C \mid + \mid \mathcal{P} \mid = 3, \dots, n$, where $n \geq 3$. We will show that the property still holds when $2 \mid C \mid + \mid \mathcal{P} \mid = n + 1$. Assume by contradiction that at a state (C, \mathcal{P}, y) where $2 \mid C \mid + \mid \mathcal{P} \mid = n + 1$ any optimal policy does not F-open box k . We will construct an alternative policy π' that does F-open box k at that state and which performs at least as well as policy π . Since stopping is suboptimal in state (C, \mathcal{P}, y) , there are two possibilities for the next action under policy π ; it either F-opens or P-opens another box.

Case 1: Policy π F-opens a box $i \neq k$. Suppose first that policy π F-opens a box $i \in C$. Then the system transitions to a state (C', \mathcal{P}', y') where $2 \mid C' \mid + \mid \mathcal{P}' \mid = n - 1$. From Theorem 1, in the next state it is optimal to stop if $V_i > \sigma_k$, since σ_k remains the largest threshold. Therefore, if $V_i > \sigma_k$, policy π stops, otherwise, using the induction hypothesis, we know that policy π will F-open box k and stop afterward if $V_k > \sigma_k$. Defining $p_s = \Pr[V_s > \sigma_k]$ for each box $s \in \{i, k\}$, we can write the expected profit under policy π as

$$\begin{aligned} J^\pi(C, \mathcal{P}, y) = & -c_i^F + p_i \cdot \mathbb{E}[V_i \mid V_i > \sigma_k] \\ & + (1 - p_i) \cdot \left(-c_k^F + p_k \cdot \mathbb{E}[V_k \mid V_k > \sigma_k] \right. \\ & \left. + (1 - p_k) \cdot \mathbb{E}[J^\pi(C', \mathcal{P}', \max\{y, V_k, V_i\}) \mid V_k \leq \sigma_k, V_i \leq \sigma_k] \right), \end{aligned}$$

where $C' = C \setminus \{i\}$ and $\mathcal{P}' = \mathcal{P} \setminus \{k\}$.

We construct an alternative policy π' that first F-opens box k and selects box k if $V_k > \sigma_k$; otherwise, policy π' F-opens box i , and proceeds to selecting box i if $V_i > \sigma_k$. In the event that $V_i \leq \sigma_k$ (and $V_k \leq \sigma_k$), policy π' imitates policy π thereafter. Therefore, the expected profit of policy π' is

$$\begin{aligned} J^{\pi'}(C, \mathcal{P}, y) = & -c_k^F + p_k \cdot \mathbb{E}[V_k \mid V_k > \sigma_k] \\ & + (1 - p_k) \cdot \left(-c_i^F + p_i \cdot \mathbb{E}[V_i \mid V_i > \sigma_k] \right. \\ & \left. + (1 - p_i) \cdot \mathbb{E}[J^\pi(C', \mathcal{P}', \max\{y, V_k, V_i\}) \mid V_k \leq \sigma_k, V_i \leq \sigma_k] \right). \end{aligned}$$

Using simple arithmetics, we can write the difference between $J^\pi(C, \mathcal{P}, y)$ and $J^{\pi'}(C, \mathcal{P}, y)$ as follows:

$$\begin{aligned}
J^{\pi'}(C, \mathcal{P}, y) - J^\pi(C, \mathcal{P}, y) &= -c_k^F + p_k \cdot \mathbb{E}[V_k | V_k > \sigma_k] + (1 - p_k) \cdot \left(-c_i^F + p_i \cdot \mathbb{E}[V_i | V_i > \sigma_k] \right) \\
&\quad + c_i^F - p_i \cdot \mathbb{E}[V_i | V_i > \sigma_k] - (1 - p_i) \cdot \left(-c_k^F + p_k \cdot \mathbb{E}[V_k | V_k > \sigma_k] \right) \\
&= p_k p_i \cdot (\mathbb{E}[V_k | V_k > \sigma_k] - \mathbb{E}[V_i | V_i > \sigma_k]) + p_k c_i^F - p_i c_k^F \\
&= p_k p_i \cdot (\mathbb{E}[V_k - \sigma_k | V_k > \sigma_k] - \mathbb{E}[V_i - \sigma_k | V_i > \sigma_k]) + p_k c_i^F - p_i c_k^F \\
&= p_i \cdot \mathbb{E}[(V_k - \sigma_k)^+] - p_k \cdot \mathbb{E}[(V_i - \sigma_k)^+] + p_k c_i^F - p_i c_k^F \\
&= p_i c_k^F - p_k \cdot \mathbb{E}[(V_i - \sigma_k)^+] + p_k c_i^F - p_i c_k^F \\
&= -p_k \cdot \mathbb{E}[(V_i - \sigma_k)^+] + p_k \cdot \mathbb{E}[(V_i - \sigma_i^F)^+].
\end{aligned}$$

The latter is nonnegative whenever $\sigma_k \geq \sigma_i^F$, which is the case considered in the lemma. Therefore, policy π' achieves at least the same expected profit as policy π , which concludes the proof of this case.

We briefly note that the case in which policy π F-opens a box $i \in \mathcal{P}$ can be proved similarly and, for brevity, is omitted.

Case 2: Policy π starts by P-opening a box $i \in C$. Suppose policy π P-opens a box $i \in C$. Then the system transitions to a state (C', \mathcal{P}', y) for which $2 | C' | + | \mathcal{P}' | = n$. By our induction hypothesis, after this P-opening action, if the revealed type t_i is such that $\sigma_i^{F|t_i} > \sigma_k$, policy π will immediately F-open box i . Consequently, by the induction hypothesis, after box i is F-opened, if it is suboptimal to stop, policy π will then F-open box k . Otherwise, if the revealed type t_i is such that $\sigma_i^{F|t_i} \leq \sigma_k$, then, by the induction hypothesis, policy π will F-open box k . Policy π is illustrated Figure EC.2a. The expected profit of policy π is

$$\begin{aligned}
J^\pi(C, \mathcal{P}, y) &= -c_i^P + \Pr[\sigma_i^{F|T_i} > \sigma_k] \left(-c_i^F + \Pr[V_i > \sigma_k | \sigma_i^{F|T_i} > \sigma_k] \mathbb{E}[V_i | V_i > \sigma_k, \sigma_i^{F|T_i} > \sigma_k] \right. \\
&\quad \left. + \Pr[V_i \leq \sigma_k | \sigma_i^{F|T_i} > \sigma_k] \left(-c_k^F + \Pr[V_k > \sigma_k] \cdot \mathbb{E}[V_k | V_k > \sigma_k] + \Pr[V_k \leq \sigma_k] \cdot J_1 \right) \right) \\
&\quad + \Pr[\sigma_i^{F|T_i} \leq \sigma_k] \left(-c_k^F + \Pr[V_k > \sigma_k] \mathbb{E}[V_k | V_k > \sigma_k] + \Pr[V_k \leq \sigma_k] \cdot J_2 \right),
\end{aligned}$$

where $J_1 = \mathbb{E} \left[J^\pi(C \setminus \{i\}, \mathcal{P} \setminus \{k\}, \max\{y, V_i, V_k\}) \mid \sigma_i^{F|T_i} > \sigma_k, V_i \leq \sigma_k, V_k \leq \sigma_k \right]$, and $J_2 = \mathbb{E} \left[J^\pi(C \setminus \{i\}, \mathcal{P} \setminus \{k\} \cup \{i\}, \max\{y, V_k\}) \mid \sigma_i^{F|T_i} \leq \sigma_k, V_k \leq \sigma_k \right]$.

Similarly to Case 1, we construct an alternative policy π' that first F-opens box k , as shown in Figure EC.2b. If $V_k \leq \sigma_k$, policy π' proceeds to P-opening box i , and it then F-opens box i if the

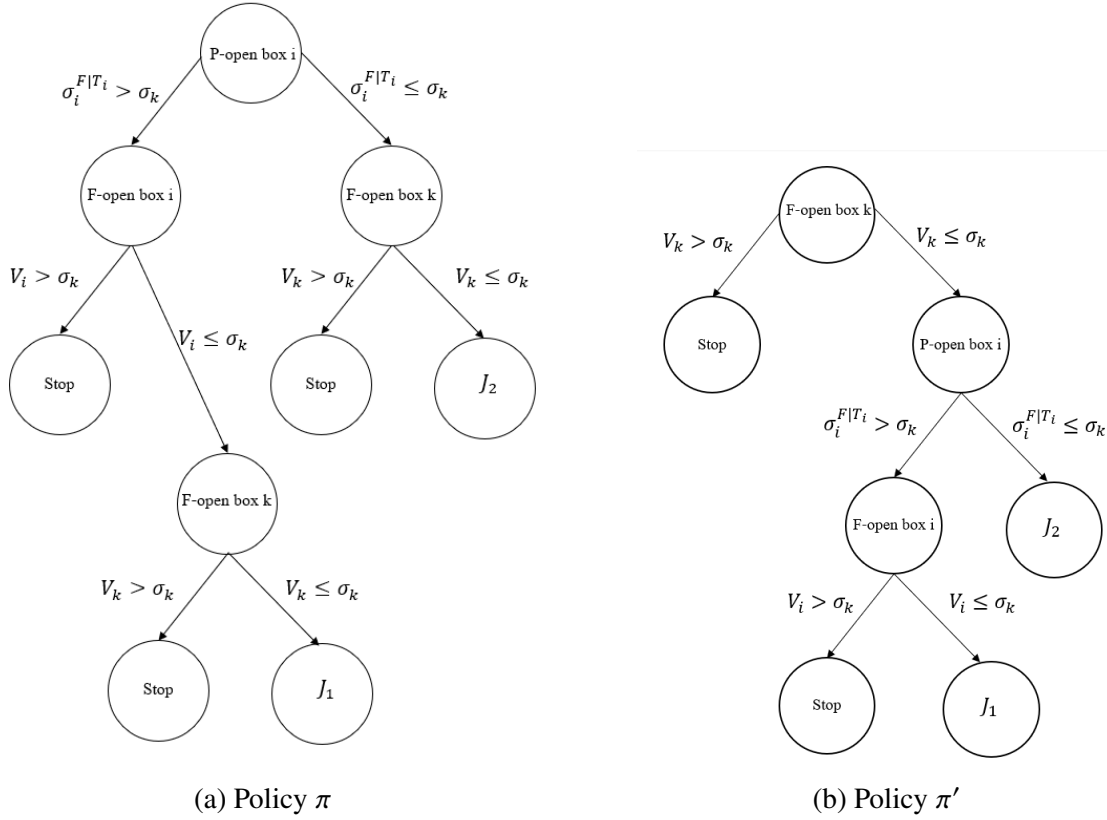


Figure EC.2 A graphical illustration of policy π and policy π' for Case 2, in the Proof of Lemma EC.1.

type T_i is such that $\sigma_i^{F|T_i} > \sigma_k$. Otherwise, π' imitates policy π in all subsequent states. Therefore, the expected profit of policy π' can be written as follows:

$$\begin{aligned}
 & J^{\pi'}(C, \mathcal{P}, y) \\
 &= -c_k^F + \Pr[V_k > \sigma_k] \mathbb{E}[V_k \mid V_k > \sigma_k] \\
 &+ \Pr[V_k \leq \sigma_k] \left(-c_i^P + \Pr[\sigma_i^{F|T_i} > \sigma_k] \left(-c_i^F + \Pr[V_i > \sigma_k \mid \sigma_i^{F|T_i} > \sigma_k] \mathbb{E}[V_i \mid V_i > \sigma_k, \sigma_i^{F|T_i} > \sigma_k] \right. \right. \\
 &\left. \left. + \Pr[V_i \leq \sigma_k \mid \sigma_i^{F|T_i} > \sigma_k] \cdot J_1 \right) + \Pr[\sigma_i^{F|T_i} \leq \sigma_k] \cdot J_2 \right).
 \end{aligned}$$

Let $p_k = \Pr[V_k > \sigma_k]$, $q_i = \Pr[\sigma_i^{F|T_i} > \sigma_k]$, and $p_i = \Pr[V_i > \sigma_k \mid \sigma_i^{F|T_i} > \sigma_k]$. The difference between $J^\pi(C, \mathcal{P}, y)$ and $J^{\pi'}(C, \mathcal{P}, y)$ can be written as

$$\begin{aligned}
 & J^{\pi'}(C, \mathcal{P}, y) - J^\pi(C, \mathcal{P}, y) \\
 &= \left(-c_k^F + q_i(1 - p_i)c_k^F + (1 - q_i)c_k^F \right) + \left(-(1 - p_k)c_i^P + c_i^P \right) + \left(-(1 - p_k)q_i c_i^F + q_i c_i^F \right) \\
 &\quad + p_k q_i p_i \cdot \mathbb{E}[V_k \mid V_k > \sigma_k] - p_k q_i p_i \cdot \mathbb{E}[V_i \mid V_i > \sigma_k, \sigma_i^{F|T_i} > \sigma_k]
 \end{aligned}$$

$$\begin{aligned}
&= -q_i p_i c_k^F + p_k c_i^P + p_k q_i c_i^F \\
&\quad + p_k q_i p_i \cdot (\mathbb{E}[V_k \mid V_k > \sigma_k] - \sigma_k) - p_k q_i p_i \cdot \left(\mathbb{E}[V_i \mid V_i > \sigma_k, \sigma_i^{F|T_i} > \sigma_k] - \sigma_k \right) \\
&= -q_i p_i c_k^F + p_k c_i^P + p_k q_i c_i^F \\
&\quad + q_i p_i \cdot \mathbb{E}[(V_k - \sigma_k)^+] - p_k q_i \cdot \mathbb{E}[(V_i - \sigma_k)^+ \mid \sigma_i^{F|T_i} > \sigma_k] \\
&= p_k \cdot \left(c_i^P + q_i c_i^F - q_i \cdot \mathbb{E}[(V_i - \sigma_k)^+ \mid \sigma_i^{F|T_i} > \sigma_k] \right) \\
&= p_k \cdot \left(c_i^P - \Pr[\sigma_i^{F|T_i} > \sigma_k] \left(-c_i^F + \mathbb{E}[(V_i - \sigma_k)^+ \mid \sigma_i^{F|T_i} > \sigma_k] \right) \right) \\
&= p_k \cdot \left(\mathbb{E}[\max\{0, -c_i^F + \mathbb{E}[(V_i - \sigma_i^P)^+]\}] - \mathbb{E}[\max\{0, -c_i^F + \mathbb{E}[(V_i - \sigma_k)^+]\}] \right).
\end{aligned}$$

The latter is nonnegative whenever $\sigma_k \geq \sigma_i^P$, which is the case considered in the lemma. This shows that policy π' achieves at least the same expected profit as policy π , thereby concluding the proof of Case 2 and the lemma. \square

We next turn to the second part of the theorem. To this end, we first make a notable observation.

LEMMA EC.2. *For a closed box $i \in C$, P-opening is dominated if $\sigma_i^F > \sigma_i^P$.*

Given this claim, it is straightforward to establish the second part of Theorem 2.

LEMMA EC.3. *For every state (C, \mathcal{P}, y) such that it is suboptimal to stop, if the F-threshold of a closed box i with $\sigma_i^F > \sigma_i^P$ is the largest opening threshold, then there exists an optimal policy that F-opens box i immediately.*

Proof of Lemma EC.3: According to Lemma EC.2, P-opening box i is a dominated action. Therefore, we can view box i as a partially opened box that can only be F-opened. From Lemma EC.1 and the fact that σ_i^F is the leading threshold, it follows that an optimal policy exists that F-opens box i immediately. \square

Proof of Lemma EC.2: Let m be the index of the box with the largest opening threshold excluding box i , and let σ_m be its opening threshold. i.e.,

$$\sigma_m = \max \left\{ \max_{j \in C \setminus \{i\}} \left\{ \sigma_j^F, \sigma_j^P \right\}, \max_{j \in \mathcal{P}} \left\{ \sigma_j^{F|t_j} \right\} \right\}.$$

If $y > \sigma_m$, then boxes other than box i should not be opened (such boxes F-expire and P-expire, see the proof of Theorem 1), and the problem reduces to the single box case for which Theorem 2 is a direct consequence of Theorem 1. As such, in what follows, we consider the case where $y \leq \sigma_m$.

Suppose an optimal policy π P-opens the box i in state (C, \mathcal{P}, y) . We will construct an alternative policy π' that F-opens box i instead at that state, and show that it achieves an expected profit at least

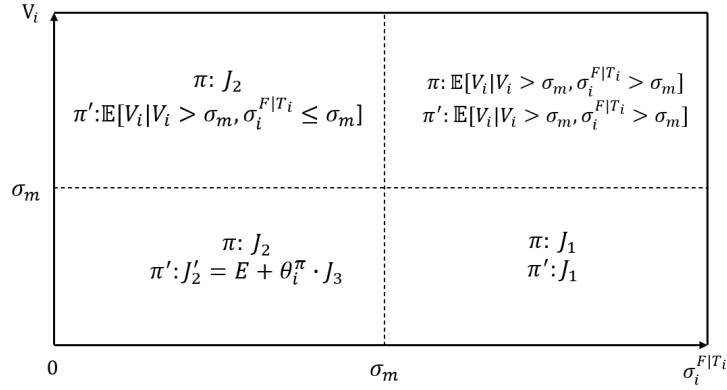


Figure EC.3 Illustration of payoffs of two policies in different scenarios. The y-axis denotes the realization of the prize, and the x-axis denotes the realization of the conditional threshold.

as high as that of policy π . According to Lemma EC.1, after P-opening box i , policy π proceeds to F-opening box i if $\sigma_i^{F|T_i} > \sigma_m$, and it will then stop if $V_i > \sigma_m$. We denote by G_i the set of types associated with box i in which policy π is guaranteed to F-open box i immediately upon P-opening, i.e., $T_i \in G_i \Leftrightarrow \sigma_i^{F|T_i} > \sigma_m$.

Therefore, the expected profit of policy π is

$$J^\pi(C, \mathcal{P}, y) = -c_i^P + \Pr[T_i \in G_i] \left(-c_i^F + \Pr[V_i > \sigma_m | T_i \in G_i] \mathbb{E}[V_i | V_i > \sigma_m, T_i \in G_i] \right. \\ \left. + \Pr[V_i \leq \sigma_m | T_i \in G_i] \cdot J_1 \right) + \Pr[T_i \notin G_i] \cdot J_2, \quad (\text{EC.8})$$

where $J_1 = \mathbb{E}[J^\pi(C \setminus \{i\}, \mathcal{P}, \max\{y, V_i\}) | V_i \leq \sigma_m, T_i \in G_i]$ and $J_2 = \mathbb{E}[J^\pi(C \setminus \{i\}, \mathcal{P} \cup \{i, t_i\}, y) | T_i \notin G_i]$.

Constructing an alternative policy π' . Let π' be the policy that F-opens box i in state (C, \mathcal{P}, y) , and selects box i if $V_i > \sigma_m$; otherwise, it further imitates policy π with the only difference that, whenever policy π F-opens box i , policy π' who already observed the value of box i , does nothing and no longer incurs the cost c_i^F . Therefore, the expected profit of policy π' is given by

$$J^{\pi'}(C, \mathcal{P}, y) = -c_i^F + \Pr[T_i \in G_i] \Pr[V_i > \sigma_m | T_i \in G_i] \mathbb{E}[V_i | V_i > \sigma_m, T_i \in G_i] \\ + \Pr[T_i \notin G_i] \Pr[V_i > \sigma_m | T_i \notin G_i] \mathbb{E}[V_i | V_i > \sigma_m, T_i \notin G_i] \\ + \Pr[T_i \in G_i] \Pr[V_i \leq \sigma_m | T_i \in G_i] \cdot J_1 \\ + \Pr[T_i \notin G_i] \Pr[V_i \leq \sigma_m | T_i \notin G_i] \cdot J'_2, \quad (\text{EC.9})$$

where $J'_2 = \mathbb{E}[J^{\pi'}(C \setminus \{i\}, \mathcal{P}, \max\{y, V_i\}) | V_i \leq \sigma_m, T_i \notin G_i]$, which is the expected profit of policy π' starting with the imitation of policy π , after observing that $T_i \notin G_i$ and $V_i \leq \sigma_m$.

In what follows, we argue that $J^{\pi'}(C, \mathcal{P}, y) \geq J^{\pi}(C, \mathcal{P}, y)$. The argument is established by first decomposing J_2 and J'_2 in a way that facilitates the comparison between J^{π} and $J^{\pi'}$.

Decomposition of J_2 and J'_2 . In order to compare the expected profits J^{π} and $J^{\pi'}$, we decompose the terms J_2 and J'_2 appearing in equations (EC.8) and (EC.9), respectively. For this purpose, we can think of a policy as a decision tree consisting of nodes and edges. Each node represents a state-action pair and each edge captures a transition. We call the nodes where the policy does not stop and opens a box *interior* nodes, and the nodes where the policy stops are called *terminal* nodes. Let $\text{box}(\tau)$ be the box that the policy selects at a terminal node τ . Let $c(\eta)$ be the opening cost paid at an interior node η . Then, we have

$$J_2 = \sum_{\text{terminal node } \tau} \Pr[\pi \text{ reaches } \tau \mid T_i \notin G_i] \cdot \mathbb{E}[V_{\text{box}(\tau)} \mid \text{info regarding box}(\tau)] \\ - \sum_{\text{interior node } \eta} \Pr[\pi \text{ reaches } \eta \mid T_i \notin G_i] \cdot c(\eta).$$

For simplicity, we assume that there is only one node in the decision tree of policy π that F-opens box i when $T_i \notin G_i$. By a slight abuse of language, we refer to this state-action pair as node i . If there are multiple such nodes, the same reasoning applies. Let $\text{tree}(i)$ be the subtree rooted at the node i . Then, on all paths outside $\text{tree}(i)$, policy π does not F-open box i . According to Lemma EC.1, if a conditional threshold greater than σ_m is revealed by a P-opening inspection, the corresponding box is immediately F-opened in an optimal policy. Therefore, when node i is reached, box i will be F-opened, and it will be selected if $V_i > \sigma_m$. We let $\theta_i^{\pi} = \Pr[\pi \text{ reaches node } i \mid T_i \notin G_i]$. Based on the preceding discussion, J_2 can be expressed as

$$J_2 = E + \theta_i^{\pi} \cdot \left(-c_i^F + \Pr[V_i > \sigma_m \mid T_i \notin G_i] \mathbb{E}[V_i \mid V_i > \sigma_m, T_i \notin G_i] + \Pr[V_i \leq \sigma_m \mid T_i \notin G_i] \cdot J_3 \right),$$

where we denote, for brevity,

$$E \triangleq \sum_{\substack{\text{terminal node } \tau \\ \text{outside } \text{tree}(i)}} \Pr[\pi \text{ reaches } \tau \mid T_i \notin G_i] \cdot \mathbb{E}[V_{\text{box}(\tau)} \mid \text{info regarding box}(\tau)] \\ - \sum_{\substack{\text{interior node } \eta \\ \text{outside } \text{tree}(i)}} \Pr[\pi \text{ reaches } \eta \mid T_i \notin G_i] \cdot c(\eta).$$

We now turn our attention to the decomposition of J'_2 . Let J_3 be the expected profit conditional on $T_i \notin G_i$ over all the sample paths that reach node i in the decision tree associated with policy π . After F-opening box i , policy π' generates the same total profit as policy π if node i is not reached.

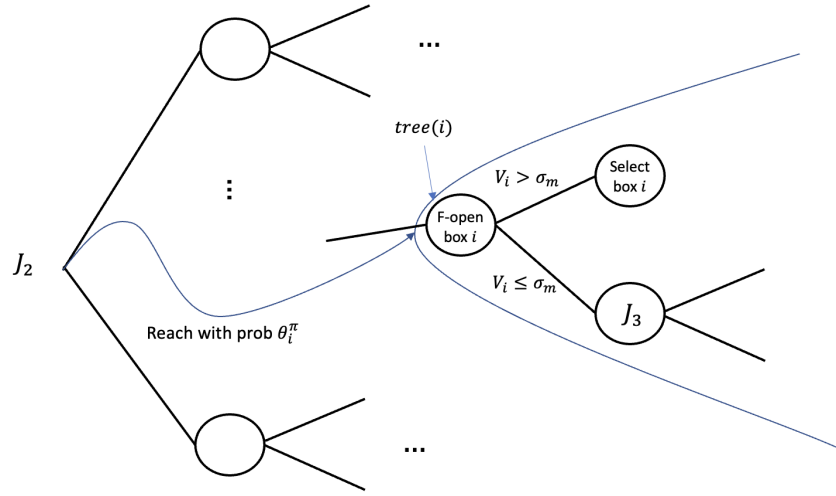


Figure EC.4 Illustration of J_2 via decision-tree-based coupling.

If node i is reached, policy π' does not need to pay an additional cost for inspection. Since policy π' has already observed that $T_i \notin G_i$ and $V_i \leq \sigma_m$, we obtain $J'_2 = E + \theta_i^\pi \cdot J_3$. In the next claim, we provide an upper bound on E using the above decision-tree decomposition of costs. The proof appears at the end of this section.

CLAIM EC.1. $E \leq (1 - \theta_i^\pi) \cdot \sigma_m$.

Bounding the difference between $J^{\pi'}$ and J^π . Following the preceding discussion, we have

$$\begin{aligned}
 & J^{\pi'}(C, \mathcal{P}, y) - J^\pi(C, \mathcal{P}, y) \\
 &= c_i^P + \Pr[T_i \notin G_i] \left(- (1 - \theta_i^\pi) c_i^F + (1 - \theta_i^\pi) \Pr[V_i > \sigma_m \mid T_i \notin G_i] \mathbb{E}[V_i \mid V_i > \sigma_m, T_i \notin G_i] \right. \\
 &\quad \left. - \Pr[V_i > \sigma_m \mid T_i \notin G_i] \cdot E \right) \\
 &\geq c_i^P + \Pr[T_i \notin G_i] \left(- (1 - \theta_i^\pi) c_i^F + (1 - \theta_i^\pi) \Pr[V_i > \sigma_m \mid T_i \notin G_i] \mathbb{E}[V_i \mid V_i > \sigma_m, T_i \notin G_i] \right. \\
 &\quad \left. - \Pr[V_i > \sigma_m \mid T_i \notin G_i] \cdot (1 - \theta_i^\pi) \sigma_m \right) \tag{EC.10}
 \end{aligned}$$

$$\begin{aligned}
 &= c_i^P + (1 - \theta_i^\pi) \Pr[T_i \notin G_i] \left(- c_i^F + \Pr[V_i > \sigma_m \mid T_i \notin G_i] (\mathbb{E}[V_i \mid V_i > \sigma_m, T_i \notin G_i] - \sigma_m) \right) \\
 &= c_i^P + (1 - \theta_i^\pi) \Pr[T_i \notin G_i] \underbrace{\left(- c_i^F + \mathbb{E}[(V_i - \sigma_m)^+ \mid T_i \notin G_i] \right)}_{\leq 0 \text{ since } \sigma_m \geq \sigma_i^{F|T_i} \text{ for } T_i \notin G_i}
 \end{aligned}$$

$$\begin{aligned}
&\geq c_i^P + \Pr[T_i \notin G_i] \left(-c_i^F + \mathbb{E}[(V_i - \sigma_m)^+ | T_i \notin G_i] \right) \\
&\geq 0.
\end{aligned} \tag{EC.11}$$

Equation (EC.10) follows from Lemma EC.1. The last inequality proceeds in the same reasoning as (EC.3) in the single-box setting, by noting that $\{T_i \notin G_i\} \Leftrightarrow \{T_i \notin \Lambda_i(\sigma_m)\}$ and $\sigma_i^{F/P} > \sigma_m$.

To complete the proof of Theorem 2, it remains to establish Claim EC.1.

Proof of Claim EC.1: We decompose the expected cumulative costs E into three types of nodes (inspections or selections) in the decision-tree: (i) F-opening a box at an interior node, then selecting its reward when it is greater than σ_m ; (ii) P-opening a box at an interior node, then F-opening it and selecting the reward when the conditional threshold and reward are greater than σ_m , respectively; and (iii) selecting a box at a terminal node with reward smaller or equal to σ_m . These three types partition all the nodes of the decision tree. Indeed, (i) when σ_m is greater than all remaining thresholds, by Theorem 1, it is optimal to select a box if its prize is found to be greater than σ_m ; (ii) when a conditional threshold greater than σ_m is found, the optimal policy immediately F-opens it by Lemma EC.1; and (iii) in all other cases, the prize value is bounded by σ_m . Consequently, we obtain the following expression:

$$\begin{aligned}
E &= \sum_{\substack{\text{interior node } \eta \\ \text{outside } tree(i) \\ \text{action}(\eta)=\text{F-open} \\ \sigma_{\text{box}(\eta)} \leq \sigma_m}} \Pr[\pi \text{ reaches } \eta | T_i \notin G_i] \left(-c_{\text{box}(\eta)}^F + \Pr[V_{\text{box}(\eta)} > \sigma_m] \mathbb{E}[V_{\text{box}(\eta)} | V_{\text{box}(\eta)} > \sigma_m] \right) \\
&+ \sum_{\substack{\text{interior node } \eta \\ \text{outside } tree(i) \\ \text{action}(\eta)=\text{P-open}}} \Pr[\pi \text{ reaches } \eta | T_i \notin G_i] \left(-c_{\text{box}(\eta)}^P + \Pr[\sigma_{\text{box}(\eta)}^{F|t_{\text{box}(\eta)}} > \sigma_m] \right. \\
&\quad \cdot \left(-c_{\text{box}(\eta)}^F + \Pr[V_{\text{box}(\eta)} > \sigma_m | \sigma_{\text{box}(\eta)}^{F|t_{\text{box}(\eta)}} > \sigma_m] \mathbb{E}[V_{\text{box}(\eta)} | V_{\text{box}(\eta)} > \sigma_m, \sigma_{\text{box}(\eta)}^{F|t_{\text{box}(\eta)}} > \sigma_m] \right) \Big) \\
&+ \sum_{\substack{\text{terminal node } \tau \\ \text{outside } tree(i) \\ V_{\text{box}(\tau)} \leq \sigma_m}} \Pr[\pi \text{ reaches } \tau | T_i \notin G_i] \mathbb{E}[V_{\text{box}(\tau)} | V_{\text{box}(\tau)} \leq \sigma_m] \\
&\leq \sum_{\substack{\text{interior node } \eta \\ \text{outside } tree(i) \\ \text{action}(\eta)=\text{F-open} \\ \sigma_{\text{box}(\eta)} \leq \sigma_m}} \Pr[\pi \text{ reaches } \eta | T_i \notin G_i] \left(\Pr[V_{\text{box}(\eta)} > \sigma_m] \cdot \sigma_m \right)
\end{aligned}$$

$$\begin{aligned}
& + \sum_{\substack{\text{interior node } \eta \\ \text{outside } \text{tree}(i) \\ \text{action}(\eta)=\text{P-open}}} \Pr [\pi \text{ reaches } \eta \mid T_i \notin G_i] \left(\Pr \left[\sigma_{\text{box}(\eta)}^{F|t_{\text{box}(\eta)}} > \sigma_m \right] \Pr \left[V_{\text{box}(\eta)} > \sigma_m \mid \sigma_{\text{box}(\eta)}^{F|t_{\text{box}(\eta)}} > \sigma_m \right] \cdot \sigma_m \right) \\
& + \sum_{\substack{\text{terminal node } \tau \\ \text{outside } \text{tree}(i) \\ V_{\text{box}(\tau)} \leq \sigma_m}} \Pr [\pi \text{ reaches } \tau \mid T_i \notin G_i] \cdot \sigma_m \\
& = \left(\sum_{\substack{\text{terminal node } \tau \\ \text{outside } \text{tree}(i)}} \Pr [\pi \text{ reaches } \tau \mid T_i \notin G_i] \right) \cdot \sigma_m \\
& = (1 - \theta_i^\pi) \cdot \sigma_m.
\end{aligned}$$

The above inequality is due to the following two facts:

- *Fact 1:* For any closed box j , by the definition of the F-threshold and monotonicity of the function $\mathbb{E}[(V_j - x)^+]$ in x , we obtain $c_j^F \geq \widetilde{c}_j^F \triangleq \mathbb{E}[(V_j - \sigma_m)^+]$, which follows because $\sigma_m \geq \sigma_j^F$. Thus,

$$\Pr[V_j > \sigma_m] \sigma_m = -\widetilde{c}_j^F + \Pr[V_j > \sigma_m] \mathbb{E}[V_j \mid V_j > \sigma_m] \geq -c_j^F + \Pr[V_j > \sigma_m] \mathbb{E}[V_j \mid V_j > \sigma_m].$$

- *Fact 2:* For any closed box j , by definition of the P-threshold and the monotonicity of $\mathbb{E}[\max\{0, -c_j^F + \mathbb{E}[(V_j - x)^+]\}]$ in x , we obtain $c_j^P \geq \widetilde{c}_j^P \triangleq \mathbb{E}[\max\{0, -c_j^F + \mathbb{E}[(V_j - \sigma_m)^+]\}]$ since $\sigma_m \geq \sigma_j^P$. Thus,

$$\begin{aligned}
& -c_j^P + \Pr[\sigma_j^{F|T_j} > \sigma_m] \left(-c_j^F + \Pr[V_j > \sigma_m \mid \sigma_j^{F|T_j} > \sigma_m] \mathbb{E}[V_j \mid V_j > \sigma_m, \sigma_j^{F|T_j} > \sigma_m] \right) \\
& \leq -\widetilde{c}_j^P + \Pr[\sigma_j^{F|T_j} > \sigma_m] \left(-c_j^F + \Pr[V_j > \sigma_m \mid \sigma_j^{F|T_j} > \sigma_m] \mathbb{E}[V_j \mid V_j > \sigma_m, \sigma_j^{F|T_j} > \sigma_m] \right) \\
& = \Pr[\sigma_j^{F|T_j} > \sigma_m] \Pr[V_j > \sigma_m \mid \sigma_j^{F|T_j} > \sigma_m] \sigma_m.
\end{aligned}$$

□

□

B.5. Proof of Theorem 3

We start by establishing an auxiliary lemma that states an optimal policy should not F-open a well-classified box. We then proceed to prove Theorem 3.

LEMMA EC.4. *For every state (C, \mathcal{P}, y) , F-opening a well-classified box $i \in C$ is suboptimal.*

Proof of Lemma EC.4: We prove this by contradiction. Suppose that in the current state (C, \mathcal{P}, y) an optimal policy π F-opens box i . In such case, the expected profit can be written as

$$\begin{aligned} J^\pi(C, \mathcal{P}, y) &= -c_i^F + \mathbb{E} [J^\pi(C \setminus \{i\}, \mathcal{P}, \max\{y, V_i\})] \\ &= -c_i^F + \sum_{t_i: \sigma_i^{F|t_i} > \sigma_{-i}} \Pr [T_i = t_i] \cdot \mathbb{E} [J^\pi(C \setminus \{i\}, \mathcal{P}, \max\{y, V_i\}) \mid T_i = t_i] \\ &\quad + \sum_{t_i: \sigma_i^{F|t_i} \leq \sigma_{-i}} \Pr [T_i = t_i] \cdot \mathbb{E} [J^\pi(C \setminus \{i\}, \mathcal{P}, \max\{y, V_i\}) \mid T_i = t_i]. \end{aligned}$$

We define an alternative policy π' and show that it attains a higher expected profit than policy π . At state (C, \mathcal{P}, y) , policy π' P-opens box i . If the observed type t_i satisfies $\sigma_i^{F|t_i} > \sigma_{-i}$, then policy π' F-opens box i and imitates policy π thereafter. Otherwise, if $\sigma_i^{F|t_i} \leq \sigma_{-i}$, then instead of F-opening the partially open box i , policy π' generates an independent sample denoted by \hat{v}_i from the prize distribution of V_i conditional on $T_i = t_i$. Policy π' treats \hat{v}_i as the realized prize of box i and imitates policy π afterward, with the only difference that, whenever policy π selects box i , policy π' selects the box whose value is y . Therefore, the expected profit of policy π' can be expressed as follows:

$$\begin{aligned} J^{\pi'}(C, \mathcal{P}, y) &= -c_i^P + \sum_{t_i: \sigma_i^{F|t_i} > \sigma_{-i}} \Pr [T_i = t_i] \left(-c_i^F + \mathbb{E} [J^\pi(C \setminus \{i\}, \mathcal{P}, \max\{y, V_i\}) \mid T_i = t_i] \right) \\ &\quad + \sum_{t_i: \sigma_i^{F|t_i} \leq \sigma_{-i}} \Pr [T_i = t_i] \cdot J^{\pi'}(C \setminus \{i\}, \mathcal{P} \cup \{(i, t_i)\}, y). \end{aligned}$$

It follows that the difference in the expected profits between the two policies satisfies

$$\begin{aligned} &J^{\pi'}(C, \mathcal{P}, y) - J^\pi(C, \mathcal{P}, y) \\ &= -c_i^P + \sum_{t_i: \sigma_i^{F|t_i} \leq \sigma_{-i}} \Pr [T_i = t_i] \left(c_i^F + J^{\pi'}(C \setminus \{i\}, \mathcal{P} \cup \{(i, t_i)\}, y) - \mathbb{E} [J^\pi(C \setminus \{i\}, \mathcal{P}, \max\{y, V_i\}) \mid T_i = t_i] \right) \\ &= -c_i^P + \sum_{t_i: \sigma_i^{F|t_i} \leq y} \Pr [T_i = t_i] \left(c_i^F + J^{\pi'}(C \setminus \{i\}, \mathcal{P} \cup \{(i, t_i)\}, y) - \mathbb{E} [J^\pi(C \setminus \{i\}, \mathcal{P}, \max\{y, V_i\}) \mid T_i = t_i] \right) \\ &\geq -c_i^P + \sum_{t_i: \sigma_i^{F|t_i} \leq y} \Pr [T_i = t_i] \left(c_i^F - \mathbb{E} [(V_i - y)^+ \mid T_i = t_i] \right). \end{aligned} \tag{EC.12}$$

The second equation follows from the assumption that box i is well-classified. The inequality holds since the difference in the total expected profits garnered by policies π' and π is at most $(V_i - y)^+$ on any sample path. We note that equation (EC.12) is exactly the negative of equation (EC.3). Therefore, by Theorem 1, we have $J^{\pi'}(C, \mathcal{P}, y) - J^\pi(C, \mathcal{P}, y) > 0$, provided that $y > \sigma_i^{F/P}$ (which holds since box i is well-classified box). \square

Proof of Theorem 3: Similarly to the proof of Lemma EC.1, we prove the theorem by induction over the integer $2|C| + |\mathcal{P}|$, which corresponds to the maximal possible number of openings from state (C, \mathcal{P}, y) onwards.

- *Base case:* $2|C| + |\mathcal{P}| = 2$. There is only one closed box. The well-classified condition is clearly satisfied, since the problem reduces to the single box case. The desired claim immediately follows from Theorem 1.

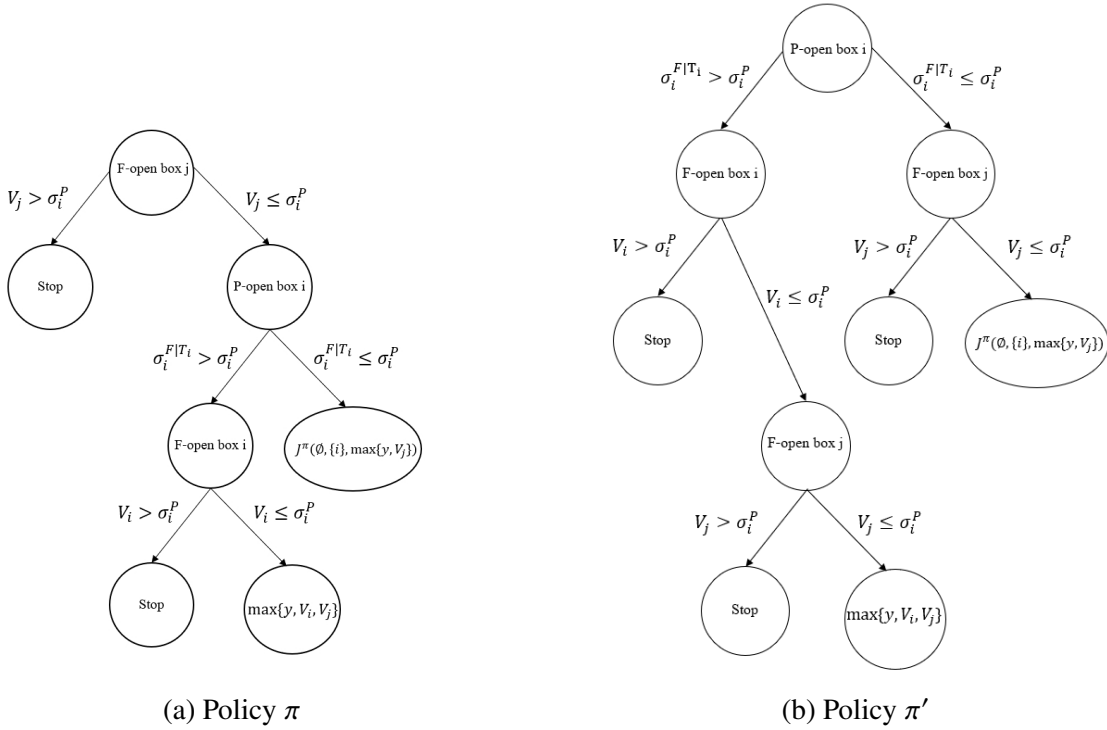


Figure EC.5 A graphical illustration of policy π and policy π' , for the base case $2|C| + |\mathcal{P}| = 3$ in the proof of Theorem 3.

- *Base case:* $2|C| + |\mathcal{P}| = 3$. Let i be the single box contained in C and let j be the single box contained in \mathcal{P} . By Lemma EC.4, F-opening box i is suboptimal. As such, suppose that an optimal policy π does not P-open box i , and instead it F-opens box j , and then the system transitions to a state in which $2|C| + |\mathcal{P}| = 2$. By Theorem 1, policy π selects box j if $V_j > \sigma_i^P$. Otherwise, when $V_j \leq \sigma_i^P$, we observe that box i is the only one remaining and it is still well-classified, therefore Theorem 1 implies that policy π proceeds to P-open this box. Applying Theorem 2, policy π then F-opens box i if the revealed type t_i is such that $\sigma_i^{F|T_i} > \sigma_i^P$, and further selects box i if $V_i > \sigma_i^P$ (this follows from Theorem 1). The policy is illustrated in Figure EC.5a. We let $q_i = \Pr \left[\sigma_i^{F|T_i} > \sigma_i^P \right]$,

$p_i = \Pr \left[V_i > \sigma_i^P \mid \sigma_i^{F|t_i} > \sigma_i^P \right]$, and $p_j = \Pr \left[V_j > \sigma_i^P \right]$. Therefore, the expected profit evaluated in state $(\{i\}, \{j\}, y)$ is

$$\begin{aligned} J^\pi((\{i\}, \{j\}, y)) &= -c_j^F + p_j \cdot \mathbb{E} \left[V_j \mid V_j > \sigma_i^P \right] \\ &+ (1 - p_j) \cdot \left(-c_i^P + q_i \left(-c_i^F + p_i \cdot \mathbb{E} \left[V_i \mid V_i > \sigma_i^P \right] + (1 - p_i) \cdot \mathbb{E} \left[\max\{y, V_i, V_j\} \mid V_i \leq \sigma_i^P, V_j \leq \sigma_i^P \right] \right) \right. \\ &\left. + (1 - q_i) \cdot \mathbb{E} \left[J^\pi((\emptyset, \{(i, t_i)\}, \max\{y, V_j\})) \mid V_j \leq \sigma_i^P, \sigma_i^{F|t_i} \leq \sigma_i^P \right] \right). \end{aligned}$$

We construct an alternative policy π' that P-opens box i in state $(\{i\}, \{j\}, y)$, as specified in Figure EC.5b. The expected profit generated by policy π' is

$$\begin{aligned} J^{\pi'}((\{i\}, \{j\}, y)) &= -c_i^P + q_i \cdot \left(-c_i^F + p_i \cdot \mathbb{E} \left[V_i \mid V_i > \sigma_i^P \right] \right. \\ &\quad \left. + (1 - p_i) \cdot \left(-c_j^F + p_j \cdot \mathbb{E} \left[V_j \mid V_j > \sigma_i^P \right] + (1 - p_j) \cdot \mathbb{E} \left[\max\{y, V_i, V_j\} \mid V_i \leq \sigma_i^P, V_j \leq \sigma_i^P \right] \right) \right) \\ &\quad + (1 - q_i) \cdot \left(-c_j^F + p_j \cdot \mathbb{E} \left[V_j \mid V_j > \sigma_i^P \right] \right. \\ &\quad \left. + (1 - p_j) \cdot \mathbb{E} \left[J^\pi((\emptyset, \{(i, t_i)\}, \max\{y, V_j\})) \mid V_j \leq \sigma_i^P, \sigma_i^{F|t_i} \leq \sigma_i^P \right] \right). \end{aligned}$$

The difference between the expected profit obtained by the two policies is

$$\begin{aligned} J^{\pi'}((\{i\}, \{j\}, y)) - J^\pi((\{i\}, \{j\}, y)) &= (-q_i(1 - p_i) - (1 - q_i) + 1) \cdot c_j^F + (-1 + (1 - p_j)) \cdot c_i^P + (-q_i + q_i(1 - p_j)) \cdot c_i^F \\ &+ (q_i p_i - (1 - p_j) q_i p_i) \cdot \mathbb{E} \left[V_i \mid V_i > \sigma_i^P \right] + (q_i(1 - p_i) p_j + (1 - q_i) p_j - p_j) \cdot \mathbb{E} \left[V_j \mid V_j > \sigma_i^P \right] \\ &= p_j \cdot \left(-c_i^P - q_i c_i^F + q_i p_i \cdot \mathbb{E} \left[V_i \mid V_i > \sigma_i^P \right] \right) - p_i q_i \cdot \left(-c_j^F + p_j \cdot \mathbb{E} \left[V_j \mid V_j > \sigma_i^P \right] \right) \\ &= p_j q_i p_i \cdot \sigma_i^P - p_i q_i \cdot \left(-c_j^F + p_j \cdot \mathbb{E} \left[V_j \mid V_j > \sigma_i^P \right] \right) \\ &\geq p_j q_i p_i \cdot \sigma_i^P - p_i q_i p_j \cdot \sigma_i^P \\ &= 0. \end{aligned}$$

The inequality holds due to the same reasoning as in the derivation of equation (EC.12). Hence, policy π' can achieve an expected profit as high as policy π .

• *Induction step.* Suppose that the induction hypothesis holds when $2 | C | + | \mathcal{P} | = 2, 3, \dots, n$ with $n \geq 3$. We wish to establish a similar claim when $2 | C | + | \mathcal{P} | = n + 1$. Suppose an optimal policy π does not P-open immediately. We show that we can construct alternative policies that perform at least as well as policy π . Depending on whether policy π F-opens or P-opens a box next, we distinguish between the following two cases.

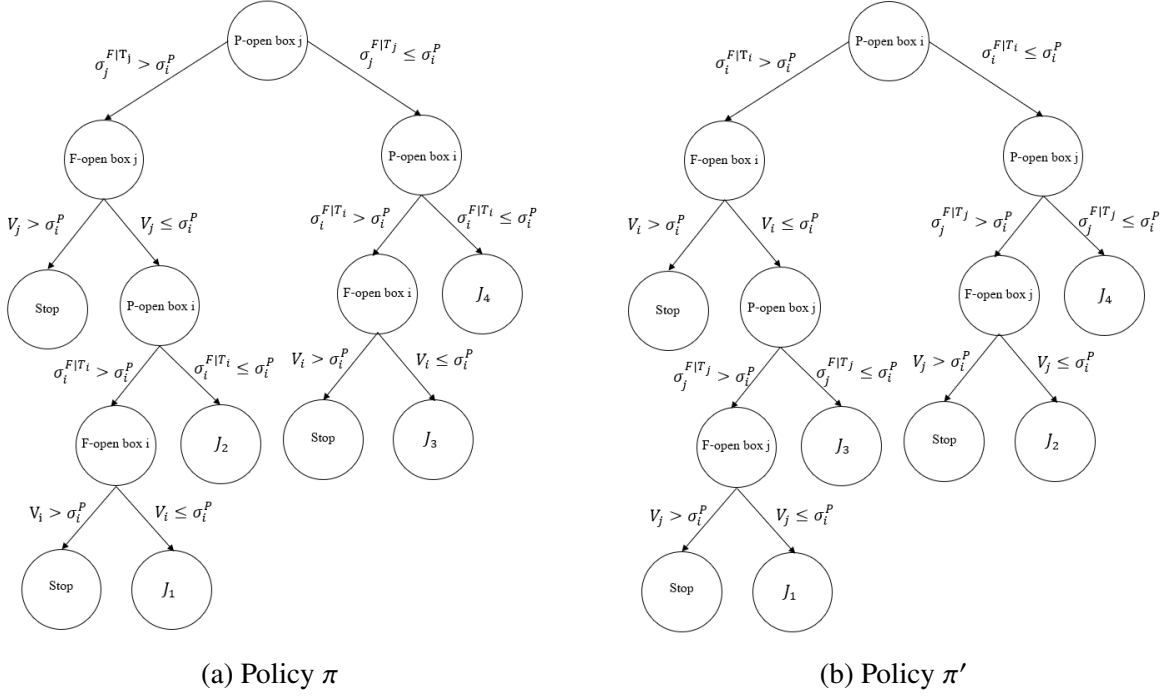


Figure EC.6 Graphical illustration of policy π and policy π' , for Case 1 in the induction step, in the proof of Theorem 3.

Case 1: Policy π starts by P-opening. Suppose that policy π P-opens a box $j \in C$ (without loss of generality, we assume has not expired). Consequently, the system transitions to a state (C', \mathcal{P}', y) for which $2 | C' | + | \mathcal{P}' | = n - 1$. By Theorem 2, if the type T_j is found to be such that $\sigma_j^{F|T_j} > \sigma_i^P$, policy π will further F-open box j , and select box j if $V_j > \sigma_i^P$ according to Theorem 1. If $V_j \leq \sigma_i^P$, policy π will P-open box i by the induction hypothesis, since box i remains well-classified, and σ_i^P is still the largest threshold. Otherwise, if $\sigma_j^{F|T_j} \leq \sigma_i^P$, by the induction hypothesis, policy π will P-open box i . Let $q_k = \Pr \left[\sigma_k^{F|T_k} > \sigma_i^P \right]$ and $p_k = \Pr \left[V_k > \sigma_i^P \mid \sigma_k^{F|T_k} > \sigma_i^P \right]$ for each box $k \in \{i, j\}$. With this definition at hand, the expected profit obtained by policy π can be written as

$$J^\pi(C, \mathcal{P}, y) = -c_j^P + q_j \cdot \left(-c_j^F + p_j \cdot E[V_j \mid V_j > \sigma_i^P] \right)$$

$$\begin{aligned}
& + (1 - p_j) \left(-c_i^P + q_i \left(-c_i^F + p_i \cdot \mathbb{E} [V_i | V_i > \sigma_i^P] + (1 - p_i) \cdot J_1 \right) + (1 - q_i) \cdot J_2 \right) \\
& + (1 - q_j) \left(-c_i^P + q_i \left(-c_i^F + p_i \cdot \mathbb{E} [V_i | V_i > \sigma_i^P] + (1 - p_i) \cdot J_3 \right) + (1 - q_i) \cdot J_4 \right),
\end{aligned}$$

$$\begin{aligned}
\text{where } J_1 &= \mathbb{E} \left[J^\pi(C \setminus \{i, j\}, \mathcal{P}, \max\{y, V_i, V_j\}) \mid \sigma_j^{F|T_j} > \sigma_i^P, V_j \leq \sigma_i^P, \sigma_i^{F|T_i} > \sigma_i^P, V_i \leq \sigma_i^P \right], \\
J_2 &= \mathbb{E} \left[J^\pi(C \setminus \{i, j\}, \mathcal{P} \cup \{(i, t_i)\}, \max\{y, V_j\}) \mid \sigma_j^{F|T_j} > \sigma_i^P, V_j \leq \sigma_i^P, \sigma_i^{F|T_i} \leq \sigma_i^P \right], \\
J_3 &= \mathbb{E} \left[J^\pi(C \setminus \{i, j\}, \mathcal{P} \cup \{(j, T_j)\}, \max\{y, V_i\}) \mid \sigma_j^{F|T_j} \leq \sigma_i^P, \sigma_i^{F|T_i} > \sigma_i^P, V_i \leq \sigma_i^P \right], \\
J_4 &= \mathbb{E} \left[J^\pi(C \setminus \{i, j\}, \mathcal{P} \cup \{(j, T_j), (i, t_i)\}, y) \mid \sigma_j^{F|T_j} \leq \sigma_i^P, \sigma_i^{F|T_i} \leq \sigma_i^P \right].
\end{aligned}$$

We construct an alternative policy π' that P-opens box i in state (C, \mathcal{P}, y) , as specified in Figure EC.6. The expected profit in state (C, \mathcal{P}, y) under policy π' is

$$\begin{aligned}
J^{\pi'}(C, \mathcal{P}, y) &= -c_i^P + q_i \left(-c_i^F + p_i \mathbb{E} [V_i | V_i > \sigma_i^P] \right. \\
&\quad \left. + (1 - p_i) \left(-c_j^P + q_j \left(-c_j^F + p_j \cdot \mathbb{E} [V_j | V_j > \sigma_i^P] + (1 - p_j) \cdot J_1 \right) + (1 - q_j) \cdot J_3 \right) \right. \\
&\quad \left. + (1 - q_i) \left(-c_j^P + q_j \left(-c_j^F + p_j \cdot \mathbb{E} [V_j | V_j > \sigma_i^P] + (1 - p_j) \cdot J_2 \right) + (1 - q_j) \cdot J_4 \right) \right).
\end{aligned}$$

Consequently, the difference between the expected profits under policies π' and π is

$$\begin{aligned}
& J^{\pi'}(C, \mathcal{P}, y) - J^\pi(C, \mathcal{P}, y) \\
&= (-1 + q_j(1 - p_j) + (1 - q_j)) \cdot c_i^P + (-q_i + q_j(1 - p_j)q_i + (1 - q_j)q_i) \cdot c_i^F \\
&\quad + (-q_i(1 - p_i) - (1 - q_i) + 1) \cdot c_j^P + (-q_i(1 - p_i)q_j - (1 - q_i)q_j + q_j) \cdot c_j^F \\
&\quad + (q_i p_i - q_j(1 - p_j)q_i p_i - (1 - q_j)q_i p_i) \cdot \mathbb{E} [V_i | V_i > \sigma_i^P] \\
&\quad + (q_i(1 - p_i)q_j p_j + (1 - q_i)q_j p_j - q_j p_j) \cdot \mathbb{E} [V_j | V_j > \sigma_i^P] \\
&= -q_j p_j c_i^P - q_j p_j q_i c_i^F + q_j p_j q_i p_i \cdot \mathbb{E} [V_i | V_i > \sigma_i^P] \\
&\quad + q_i p_i c_j^P + q_i p_i q_j c_j^F - q_i p_i q_j p_j \cdot \mathbb{E} [V_j | V_j > \sigma_i^P] \\
&= q_j p_j \cdot \left(-c_i^P - q_i c_i^F + q_i p_i \cdot \mathbb{E} [V_i | V_i > \sigma_i^P] \right) \\
&\quad - q_i p_i \cdot \left(-c_j^P - q_j c_j^F + q_j p_j \cdot \mathbb{E} [V_j | V_j > \sigma_i^P] \right) \\
&= q_j p_j q_i p_i \cdot \sigma_i^P - q_i p_i \cdot \left(-c_j^P - q_j c_j^F + q_j p_j \cdot \mathbb{E} [V_j | V_j > \sigma_i^P] \right) \\
&\geq q_j p_j q_i p_i \cdot \sigma_i^P - q_i p_i q_j p_j \cdot \sigma_i^P \\
&= 0.
\end{aligned}$$

The last inequality is due to the same reasoning as in equation (EC.12). Hence, policy π' can achieve an expected profit as high as policy π .

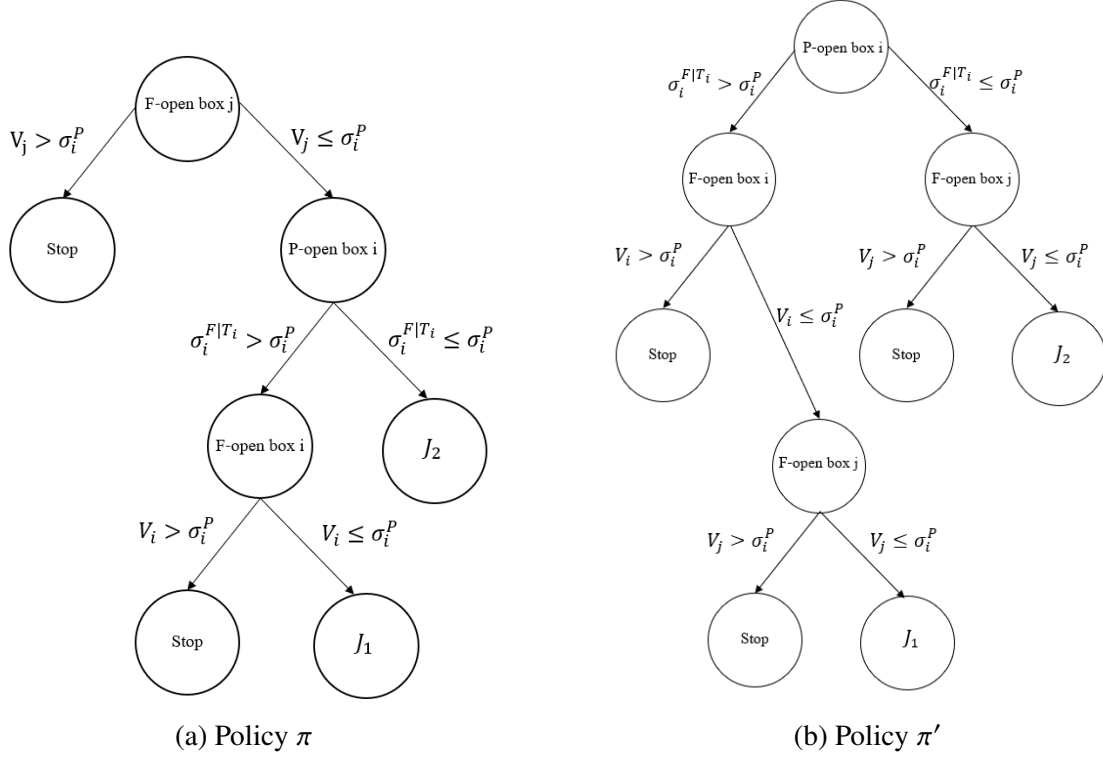


Figure EC.7 Graphical illustration of policy π and policy π' , for Case 2 in the induction step, in the proof of Theorem 3.

Case 2: Policy π starts by F-opening. Suppose that policy π F-opens a box $j \in C$. Consequently, the system transitions to a state (C', \mathcal{P}', y') for which $2 \mid C' \mid + \mid \mathcal{P}' \mid = n - 1$. Alternatively, if we suppose that policy π F-opens a box $j \in \mathcal{P}$, then it transitions to a state (C', \mathcal{P}', y') for which $2 \mid C' \mid + \mid \mathcal{P}' \mid = n$. Observe that, in both settings, box i will remain well-classified. Indeed, after box j is opened, if it is suboptimal to stop, then the P-threshold of box i is still the largest opening threshold. Hence, in view of the induction hypothesis, policy π will P-open box i . Let $p_j = \Pr[V_j > \sigma_i^P]$, $q_i = \Pr[\sigma_i^{F|T_i} > \sigma_i^P]$ and $p_i = \Pr[V_i > \sigma_i^P \mid \sigma_i^{F|T_i} > \sigma_i^P]$. The expected profit in state (C, \mathcal{P}, y) under policy π is

$$J^\pi(C, \mathcal{P}, y) = -c_j^F + p_j \cdot \mathbb{E}[V_j \mid V_j > \sigma_i^P] + (1 - p_j) \cdot \left(-c_i^P + q_i \left[-c_i^F + p_i \cdot \mathbb{E}[V_i \mid V_i > \sigma_i^P] \right] + q_i(1 - p_i) \cdot J_1 + (1 - q_i) \cdot J_2 \right).$$

We construct an alternative policy π' that P-opens box i first, as specified in Figure EC.7. The expected profit in state (C, \mathcal{P}, y) under policy π' is

$$\begin{aligned} J^{\pi'}(C, \mathcal{P}, y) &= -c_i^P + q_i \cdot \left(-c_i^F + p_i \cdot \mathbb{E}[V_i | V_i > \sigma_i^P] + (1 - p_i) \left(-c_j^F + p_j \cdot \mathbb{E}[V_j | V_j > \sigma_i^P] + (1 - p_j) \cdot J_1 \right) \right) \\ &\quad + (1 - q_i) \cdot \left(-c_j^F + p_j \cdot \mathbb{E}[V_j | V_j > \sigma_i^P] + (1 - p_j) \cdot J_2 \right). \end{aligned}$$

Hence, the difference in the expected profits under policies π' and π is

$$\begin{aligned} J^{\pi'}(C, \mathcal{P}, y) - J^{\pi}(C, \mathcal{P}, y) &= (-1 + 1 - p_j) \cdot c_i^P + (-q_i + (1 - p_j)q_i) \cdot c_i^F + (-q_i(1 - p_i) - (1 - q_i) + 1) \cdot c_j^F \\ &\quad + (q_i p_i - (1 - p_j)q_i p_i) \cdot \mathbb{E}[V_i | V_i > \sigma_i^P] + (q_i(1 - p_i)p_j + (1 - q_i)p_j - p_j) \cdot \mathbb{E}[V_j | V_j > \sigma_i^P] \\ &= p_j \cdot \left(-c_i^P - q_i c_i^F + q_i p_i \cdot \mathbb{E}[V_i | V_i > \sigma_i^P] \right) - q_i p_i \cdot \left(-c_j^F + p_j \cdot \mathbb{E}[V_j | V_j > \sigma_i^P] \right) \\ &= p_j q_i p_i \cdot \sigma_i^P - q_i p_i \cdot \left(-c_j^F + p_j \cdot \mathbb{E}[V_j | V_j > \sigma_i^P] \right) \\ &\geq p_j q_i p_i \cdot \sigma_i^P - p_j q_i p_i \cdot \sigma_i^P = 0. \end{aligned}$$

The inequality follows from the same reasoning as the derivation of equation (EC.12). Hence, policy π' can achieve an as high expected profit as policy π . This concludes the proof of the induction step. □

B.6. Proof of Theorem 4

We first introduce notation prior to proceeding to prove each of the five properties. Since boxes are identical, for brevity, we drop the subscript of box indices throughout this proof. Recall that each box can be of either a good type (G) or a bad type (B), and the rewards are binary with values 0 or 1. Let p^G and p^B denote the probabilities of a box being of good type and bad type, respectively. Conditional on the box being of good (bad) type, the probability that it contains a reward is q^G (q^B).

To encode the state space, it suffices to record the number of closed boxes n_C , the number of good-type boxes n_G , the number of bad-type boxes n_B and the current best reward: (n_C, n_G, n_B, y) . The initial state is then $(n, 0, 0, 0)$. Since a good type will be F-opened immediately upon observation and it is optimal to stop once a prize is found, we can “skip” states in which good-type boxes were discovered and “collapse” states where the reward was found, to obtain a more compact

representation using the tuple (n_C, n_B) . Finally, we denote by J^F (J^P) the expected payoff of F-opening (P-opening) a closed box and then following the optimal policy at any state (n_C, n_B) with at least one closed box ($n_C \geq 1$).

We proceed to prove each stated property.

Property: Policy π assigns the highest priority to F-opening good-type boxes. Applying Observation 1 to the two-type case, we observe that

1. $\sigma^{F/P} < \sigma^F \leq \sigma^P$ (the theorem assumes that $\sigma^F < \sigma^P$), and
2. $\sigma^{F|B} \leq \sigma^{F/P}$ and $\sigma^{F|G} \geq \sigma^P$.

Theorem 2 then implies that an optimal policy directly F-opens a good type upon its discovery.

Property: Policy π either discards all bad-type boxes, or assigns the lowest priority to F-opening them. We consider two cases, depending on whether $\sigma^{F|B}$ is positive. For the case when $\sigma^{F|B} \leq 0$, a bad type is not worth opening. For the case when $\sigma^{F|B} > 0$, the property can be proven by an interchange argument. For brevity, the proof is omitted.

Property: Policy π is a committing policy. To show that there exists an optimal policy that is a committing policy (see Section 4.3 for a formal definition and discussion), we first prove that the presence of bad-type boxes does *not* affect the decision to F-open versus P-open the closed boxes. We formalize this in the following lemma, which for ease of reading we defer proving to the end of this section.

LEMMA EC.5. *For a PSI-B2I instance at state (n_C, n_B) where $n_C \geq 1$, the difference $J^F(n_C, n_B) - J^P(n_C, n_B)$ is a function of n_C and is independent of n_B .*

Recall from Theorem 1 that at state $(1, 0)$, it is optimal to F-open the closed box (see for illustration Figure 2). Lemma EC.5 then implies that at any state $(1, n_B)$, it is optimal to F-open the closed box. The latter serves as a basis for an inductive argument for the optimal policy being a committing policy. At a state (n_C, n_B) , the difference in the expected costs between F-opening versus P-opening depends on actions taken at future states of the form $(n_C - 1, n_B)$ and $(n_C - 1, n_B + 1)$, that are reached depending on the opening mode and whether the opened job is of a bad type or not. Regardless, the policy takes actions that are independent of the number of bad-type boxes, meaning that whether or not P-opening or F-opening is better at state (n_C, n_B) , one can already commit at that state to the opening mode of each of the remaining boxes.

Property: There exists a finite threshold N_C such that policy π prioritizes P-opening closed boxes over F-opening them iff the number of closed boxes $n_C > N_C$. From Lemma 2, we know

that once an optimal policy commits to an opening mode for each closed box, the optimal opening sequence is prescribed by the opening thresholds (Algorithm 1). Therefore, P-openings will take place prior to F-openings.

We proceed by induction on n_C . At any state $(1, n_B)$, F-opening is optimal. At the state (n_C, n_B) either P-opening or F-opening is optimal. If we add a closed box to create state $(n_C + 1, n_B)$, F-opening could be optimal only if P-opening is not optimal at state (n_C, n_B) (because this would imply that a committing policy F-opens prior to P-opening).

The finiteness of N_C follows from Lemma EC.6 and the asymptotic optimality of the policy that exclusively P-opens closed boxes.

Proof of Lemma EC.5: We prove the claim by induction.

- We start from the base case where $n_C = 1$. Let the closed box be box i . Following the reasoning of Lemma 2, the expected payoff of P-opening box i is given by

$$J^P(1, n_B) = -c^P + \mathbb{E} \left[\max \left\{ \min \{V_i, \sigma^{F|T_i}\}, \max_{j \in \mathcal{P}} \{ \min \{V_j, \sigma^{F|B}\} \} \right\} \right]$$

The expected payoff of F-opening box i is given by

$$J^F(1, n_B) = \mathbb{E} \left[\max \left\{ \min \{V_i, \sigma^F\}, \max_{j \in \mathcal{P}} \{ \min \{V_j, \sigma^{F|B}\} \} \right\} \right]$$

For brevity let $\mathcal{P}(n_B)$ denote the set of n_B bad boxes. We let $X(n_B) \stackrel{\text{def}}{=} \max_{j \in \mathcal{P}(n_B)} \{ \min \{V_j, \sigma^{F|B}\} \}$. We note the difference

$$\begin{aligned} & J^P(1, n_B) - J^F(1, n_B) \\ &= -c^P + \mathbb{E} [\max \{ \min \{V_i, \sigma^{F|T_i}\}, X \}] - \mathbb{E} [\max \{ \min \{V_i, \sigma^F\}, X \}] \\ &\stackrel{(a)}{=} -c^P + \mathbf{Pr} [\min \{V_i, \sigma^{F|T_i}\} = 0] \mathbb{E} [X] + \mathbf{Pr} [\min \{V_i, \sigma^{F|T_i}\} \neq 0] \mathbb{E} [\min \{V_i, \sigma^{F|T_i}\} \mid \min \{V_i, \sigma^{F|T_i}\} \neq 0] \\ &\quad - \mathbf{Pr} [\min \{V_i, \sigma^F\} = 0] \mathbb{E} [X] - \mathbf{Pr} [\min \{V_i, \sigma^F\} \neq 0] \mathbb{E} [\min \{V_i, \sigma^F\} \mid \min \{V_i, \sigma^F\} \neq 0] \\ &\stackrel{(b)}{=} -c^P + \mathbf{Pr} [\min \{V_i, \sigma^{F|T_i}\} \neq 0] \mathbb{E} [\min \{V_i, \sigma^{F|T_i}\} \mid \min \{V_i, \sigma^{F|T_i}\} \neq 0] \\ &\quad - \mathbf{Pr} [\min \{V_i, \sigma^F\} \neq 0] \mathbb{E} [\min \{V_i, \sigma^F\} \mid \min \{V_i, \sigma^F\} \neq 0], \end{aligned}$$

which is independent of X . Step (a) follows by noting that X can only take values 0 or $\sigma^{F|B}$, and $\sigma^{F|T_i} \geq \sigma^{F|B}$ almost surely. Hence, conditioned on the event $\{\min \{V_i, \sigma^{F|T_i}\} \neq 0\}$, we have $X \leq \min \{V_i, \sigma^{F|T_i}\}$ almost surely. Step (b) follows from the simple observation that $\mathbf{Pr} [\min \{V_i, \sigma^{F|T_i}\} = 0] = \mathbf{Pr} [\min \{V_i, \sigma^F\} = 0]$.

• Suppose the induction hypothesis is true for $n_C \geq 1$. At state $(n_C + 1, n_B)$ where $n_C \geq 1$, the expected payoff of F-opening a closed box and following the optimal policy afterwards is

$$J^F(n_C + 1, n_B) = -c^F + (p^G q^G + p^B q^B) \cdot 1 + \left(p^G (1 - q^G) + p^B (1 - q^B) \right) \cdot J(n_C, n_B) .$$

The expected payoff of P-opening a closed box and following the optimal policy afterwards is

$$J^P(n_C + 1, n_B) = -c^P + p^G \cdot \left(-c^F + q_G \cdot 1 + (1 - q_G) \cdot J(n_C, n_B) \right) + p^B \cdot (J(n_C, n_B + 1)) ,$$

where we note that an optimal policy F-opens a good box immediately. The difference is

$$J^F(n_C + 1, n_B) - J^P(n_C + 1, n_B) = c^P + p^B \cdot \left(q^B - c^F + (1 - q^B) J(n_C, n_B) - J(n_C, n_B + 1) \right) . \quad (\text{EC.13})$$

To prove the lemma, it suffices to show the right-hand-side of above expression is a constant with respect to n_B .

Next, we proceed to show that the term $J(n_C, n_B + 1) - (1 - q^B)J(n_C, n_B)$ in the above display is indeed independent of n_B . To this end, we consider how the optimal strategies at state (n_C, n_B) and $(n_C, n_B + 1)$ differ. By the induction hypothesis, the decision to whether F-open or P-open a closed box is only a function of n_C and is *independent* of the number of bad type. Therefore, the optimal courses of actions starting from state (n_C, n_B) and $(n_C, n_B + 1)$ coincide, up to the moment (if occurs) when the latter has one last bad type and the former has nothing left.

Let $\mathcal{C}(n_C)$ be the set of indices of n_C closed boxes. For brevity, we denote $J(n_C, 0) = \max_{\pi} \mathbb{E} \left[\sum_{i \in \mathcal{C}(n_C)} (\mathbb{S}_i^{\pi} V_i - \mathbb{F}_i^{\pi} c_i^F - \mathbb{P}_i^{\pi} c_i^P) \right]$. Let $\theta(n) = (p^G (1 - q^G) + p^B (1 - q^B))^n$ be the probability that the optimal policy ends up with no rewards at state $(n, 0)$. Further, we recall that the optimal policy will postpone opening the bad boxes to the very end. Therefore, we can decompose $J(n_C, n_B)$ and $J(n_C, n_B + 1)$ in the following way.

$$J(n_C, n_B) = J(n_C, 0) + \theta(n_C) \mathbb{E} [X(n_B)] \quad \text{and} \quad J(n_C, n_B + 1) = J(n_C, 0) + \theta(n_C) \mathbb{E} [X(n_B + 1)]$$

Hence, we have

$$\begin{aligned} & J(n_C, n_B + 1) - (1 - q^B)J(n_C, n_B) \\ &= \theta(n_C) \cdot \left(\mathbb{E} [X(n_B + 1)] - (1 - q^B) \mathbb{E} [X(n_B)] \right) + q_B J(n_C, 0) \\ &= \theta(n_C) \cdot \left(\mathbb{E} \left[\max_{j \in \mathcal{P}(n_B + 1)} \{ \min \{ V_j, \sigma^{F|B} \} \} \right] - (1 - q^B) \mathbb{E} \left[\max_{j \in \mathcal{P}(n_B)} \{ \min \{ V_j, \sigma^{F|B} \} \} \right] \right) + q_B J(n_C, 0) \\ &= \theta(n_C) \cdot \left(q^B \sigma^{F|B} + (1 - q^B) \mathbb{E} \left[\max_{j \in \mathcal{P}(n_B)} \{ \min \{ V_j, \sigma^{F|B} \} \} \right] - (1 - q^B) \mathbb{E} \left[\max_{j \in \mathcal{P}(n_B)} \{ \min \{ V_j, \sigma^{F|B} \} \} \right] \right) \\ &\quad + q_B J(n_C, 0) \\ &= q^B \cdot \left(\theta(n_C) \sigma^{F|B} + J(n_C, 0) \right) \end{aligned}$$

which is independent of n_B , as desired.

□

B.7. Optimal policy for PSI-B2I

Algorithm 4 specifies a class of policies that satisfy the structural properties required by Theorem 4 for PSI-B2I instances. An optimal policy can then be obtained by enumerating this finite class and selecting the best-performing configuration, as described in Algorithm 3.

Algorithm 3 Computing an optimal PSI-B2I policy by enumerating the threshold class (Theorem 4).

```

1: Initialize  $V^* \leftarrow -\infty$ ;  $(N_C^*, \text{OPENBAD}^*) \leftarrow (0, \text{False})$ ;
2: for  $N_C \in \{0, 1, \dots, N\}$  do
3:   for  $\text{OPENBAD} \in \{\text{False}, \text{True}\}$  do
4:     Compute  $V(N_C, \text{OPENBAD}) = \text{expected payoff of policy } \pi^{N_C, \text{OPENBAD}}$  (Algorithm 4);
5:     if  $V(N_C, \text{OPENBAD}) > V^*$  then
6:        $V^* \leftarrow V(N_C, \text{OPENBAD})$ ;
7:        $(N_C^*, \text{OPENBAD}^*) \leftarrow (N_C, \text{OPENBAD})$ ;
8:     end if
9:   end for
10: end for
11: return the policy  $\pi^{N_C^*, \text{OPENBAD}^*}$ ;
```

Algorithm 4 Policy template $\pi^{N_C, \text{OPENBAD}}$ for PSI-B2I.

```

1: Initialize counts  $(n_C, n_G, n_B, y) \leftarrow (N, 0, 0, 0)$ ;
2: while  $y = 0$  and  $(n_C + n_G + n_B) > 0$  do
3:   if  $n_G > 0$  then                                     ▶ Highest priority: F-open good-type boxes
4:     F-open a good-type box; update  $y \leftarrow \max\{y, v\}$ ;  $n_G \leftarrow n_G - 1$ ;
5:   else if  $n_C > 0$  and  $n_C > N_C$  then                     ▶ P-open closed boxes
6:     P-open a closed box;  $n_C \leftarrow n_C - 1$ ;
7:     if revealed type is good then
8:        $n_G \leftarrow n_G + 1$ ;
9:     else
10:       $n_B \leftarrow n_B + 1$ ;
11:    end if
12:   else if  $n_C > 0$  and  $n_C \leq N_C$  then
13:     F-open a closed box; update  $y \leftarrow \max\{y, v\}$ ;  $n_C \leftarrow n_C - 1$ ;
14:   else                                                   ▶ Bad-type boxes are lowest priority
15:     if  $\text{OPENBAD} = \text{True}$  and  $n_B > 0$  then
16:       F-open a bad-type box; update  $y \leftarrow \max\{y, v\}$ ;  $n_B \leftarrow n_B - 1$ ;
17:     else
18:       break;
19:     end if
20:   end if
21: end while
22: Select  $y$  and stop;

```

Appendix C: Proofs of Section 4

In this section, we present all the proofs for Section 4. We present the proof of Theorem 5 to the end due to the dependence on other results of which we present the proofs first.

C.1. Proof of Lemma 1

The proof of the first inequality mostly follows similar lines of reasoning to that of Brown and Smith (2013)[Proposition 4.2], and for brevity is thus omitted. The remainder of the proof is dedicated to showing that $J^W(C, \mathcal{P}, y) \leq J^K(C, \mathcal{P}, y)$.

Define X^K as follows

$$X^K = \max \left\{ y, \max_{i \in C} \left\{ \max \{K_i, \tilde{K}_i\} \right\}, \max_{i \in \mathcal{P}} \left\{ \tilde{K}_i^{t_i} \right\} \right\}.$$

Recall from equation (9) that $\sigma_M = \max \left\{ \max_{j \in C} \{ \sigma_j^F, \sigma_j^P \}, \max_{j \in \mathcal{P}} \{ \sigma_j^{F|t_j} \} \right\}$. Since we assume that it is suboptimal to stop in current state (C, \mathcal{P}, y) , we infer that $y < \sigma_M$, and thus, $\sigma_M \geq X^K \geq y \geq 0$.

Consequently, we have

$$\begin{aligned}
 J^K(C, \mathcal{P}, y) &\triangleq \mathbb{E}[X^K] \\
 &= \int_0^{\sigma_M} \Pr[X^K > s] ds \\
 &= \sigma_M - \int_y^{\sigma_M} \Pr[X^K \leq s] ds \\
 &= \sigma_M - \int_y^{\sigma_M} \left(\prod_{i \in C} \Pr[\max\{K_i, \tilde{K}_i\} \leq s] \right) \cdot \left(\prod_{i \in \mathcal{P}} \Pr[\tilde{K}_i^{t_i} \leq s \mid T_i = t_i] \right) dy. \quad (\text{EC.14}) \\
 &\geq \sigma_M - \int_y^{\sigma_M} \prod_{i=1}^N \omega_i(x_i, y) dy \\
 &= J^W(C, \mathcal{P}, y).
 \end{aligned}$$

Equation EC.14 follows from the independence between boxes. The inequality is justified in the remainder of the proof. Recall that, if $i \in \mathcal{P}$, we have

$$\omega_i(x_i, y) = \Pr \left[\min\{V_i, \sigma_i^{F|t_i}\} \leq y \mid T_i = t_i \right] = \Pr \left[\tilde{K}_i^{t_i} \leq y \mid T_i = t_i \right].$$

Now, if $i \in C$, we have

$$\begin{aligned}
 \omega_i(x_i, y) &= \mathbf{I}\{y > \sigma_i^{F/P}\} \Pr \left[\min \left\{ V_i, \sigma_i^{F|T_i}, \sigma_i^P \right\} \leq y \right] + \mathbf{I}\{y \leq \sigma_i^{F/P}\} \Pr \left[\min \{V_i, \sigma_i^F\} \leq y \right] \\
 &= \mathbf{I}\{y > \sigma_i^{F/P}\} \Pr \left[\tilde{K}_i \leq y \right] + \mathbf{I}\{y \leq \sigma_i^{F/P}\} \Pr \left[K_i \leq y \right] \\
 &= \Pr \left[\max \left\{ \mathbf{I}\{y > \sigma_i^{F/P}\} \tilde{K}_i, \mathbf{I}\{y \leq \sigma_i^{F/P}\} K_i \right\} \leq y \right] \\
 &\geq \Pr \left[\max \{ \tilde{K}_i, K_i \} \leq y \right].
 \end{aligned}$$

C.2. Asymptotic optimality

LEMMA EC.6 (Asymptotic optimality). *Suppose that there exists $\theta > 0$ such that $N_k > \theta N$ for all $k \in [K]$. Let $u = \Pr[V_i > \sigma_i^F]$, $q = \Pr[\sigma_i^{F|T_i} > \sigma_i^P]$ and $p = \Pr[V_i > \sigma_i^P \mid \sigma_i^{F|T_i} > \sigma_i^P]$, where $i \in [K]$ is the class of boxes that initially attains the largest opening threshold $\sigma_M([N], \emptyset)$. We distinguish between two cases:*

- *Case 1: $\sigma_i^F = \sigma_M([N], \emptyset)$. The policy that always F-opens closed boxes in class i and implements the stopping rule of Theorem 1 is a factor $(1 - (1 - u)^{\theta N})$ -close to the optimum.*

• *Case 2:* $\sigma_i^P = \sigma_M([N], \emptyset)$. The policy that always P -opens closed boxes in class i , F -opens partially open boxes when $\sigma_i^{F|t_i} > \sigma_i^P$, and implements the stopping rule of Theorem 1 is a factor $(1 - (1 - pq)^{\theta N})$ -close to the optimum.

The proof is based on the simple observation that the expected profit under the policy described in Cases 1 and 2 asymptotically matches Whittle's integral (11). Intuitively, when we have a sufficiently large number of “good” boxes (with associated threshold $\sigma_M(C, \mathcal{P})$), the actual probability that the effort of opening them is in vain approaches zero, i.e., the gap between Whittle's integral and $\sigma_M(C, \mathcal{P})$ decreases exponentially in the number of good boxes. Conversely, the policy of Lemma EC.6 has an expected profit approaching $\sigma_M(C, \mathcal{P})$ from below.

We note in passing, that the Pandora's box model with infinite number of boxes has been commonly used in the literature as a model of consumer search (e.g., Wolinsky 1986). The fact that our problem admits an asymptotically optimal index policy offers researchers the ability to use the problem as a building block for constructing richer models that capture more realistic behavior.

Proof: We let $u = \Pr[V_i > \sigma_i^F]$, $q = \Pr[\sigma_i^{F|t_i} > \sigma_i^P]$ and $p = \Pr[V_i > \sigma_i^P \mid \sigma_i^{F|t_i} > \sigma_i^P]$, where $i \in [K]$ is the class of boxes that initially attains the largest opening threshold $\sigma_M([N], \emptyset)$.

We first examine Case 1, where $\sigma_i^F = \sigma_M([N], \emptyset)$. In this setting, our policy achieves an expected profit of at least

$$\begin{aligned} & -c_i^F + u\mathbb{E}[V_i \mid V_i > \sigma_i^F] + (1-u) \left(-c_i^F + u\mathbb{E}[V_i \mid V_i > \sigma_i^F] + (1-u)(\dots) \right) \\ & \geq u\sigma_i^F + (1-u)u\sigma_i^F + (1-u)^2u\sigma_i^F + \dots + (1-u)^{\lfloor \theta \cdot N \rfloor - 1} u\sigma_i^F \\ & = u\sigma_i^F \frac{1 - (1-u)^{\lfloor \theta \cdot N \rfloor}}{1 - (1-u)} \xrightarrow{N \rightarrow \infty} \sigma_i^F, \end{aligned}$$

where we use the fact

$$u \cdot \sigma_i^F = -c_i^F + u \cdot \mathbb{E}[V_i \mid V_i > \sigma_i^F].$$

By Lemma 1, we know that σ_i^F is an upper bound on the optimal expected profit. Hence, we derive the approximation ratio stated in Lemma EC.6.

Next, we examine Case 2, where $\sigma_i^P = \sigma_M([N], \emptyset)$. Here, our policy achieves an expected profit of at least

$$\begin{aligned} & -c_i^P + q \left(-c_i^F + p\mathbb{E}[V_i \mid V_i > \sigma_i^P, \sigma_i^{F|t_i} > \sigma_i^P] \right) \\ & + (1-q + q(1-p)) \left(-c_i^P + q \left(-c_i^F + p\mathbb{E}[V_i \mid V_i > \sigma_i^P, \sigma_i^{F|t_i} > \sigma_i^P] + (1-q + q(1-p))(\dots) \right) \right) \\ & \geq pq\sigma_i^P + (1-pq)pq\sigma_i^P + (1-pq)^2pq\sigma_i^P + \dots + (1-pq)^{\lfloor \theta \cdot N \rfloor - 1} pq\sigma_i^P \\ & = pq\sigma_i^P \frac{1 - (1-pq)^{\lfloor \theta \cdot N \rfloor}}{1 - (1-pq)} \xrightarrow{N \rightarrow \infty} \sigma_i^P, \end{aligned}$$

where we use the fact

$$pq\sigma_i^P = -c_i^P + q \left(-c_i^F + p\mathbb{E} \left[V_i \mid V_i > \sigma_i^P, \sigma_i^{F|T_i} > \sigma_i^P \right] \right).$$

By Lemma 1, we know that σ_i^P is an upper bound on the optimal expected profit. Hence, the approximation ratio of Lemma EC.6 immediately follows. \square

C.3. Proof of Lemma 2

First, we identify necessary and sufficient conditions such that there is no gap between the free-info upper bound and the value function.

LEMMA EC.7. *For every policy π , we have $J^{\pi,K}(s_0) = J^\pi(s_0)$ if and only if policy π satisfies the following conditions:*

1. π selects any F -opened box i (which was not P -opened) whose value satisfies $V_i > \sigma_i^F$;
2. π fully opens any P -opened box i whose type satisfies $\sigma_i^{F|T_i} > \sigma_i^P$;
3. π selects any F -opened box i (which was already P -opened) that satisfies $V_i > \min\{\sigma_i^P, \sigma_i^{F|T_i}\}$.

Consistently with the terminology used in previous literature, we call these policies *non-exposed*. Note that non-exposed policies were introduced by Beyhaghi and Kleinberg (2019). Our proof of Lemma EC.7 in Appendix C.4 requires a slightly more general argument for our sequential inspection processes and the corresponding two-stage (rather than single-stage) call options.

Now, using equation (14), for any committing policy $\pi^{F,P}$, the expected profit in the free-info problem is

$$\begin{aligned} J^{\pi,K}(s_0) &= \mathbb{E} \left[\sum_{i=1}^N \left(\mathbb{S}_i^\pi K_i (1 - \mathbb{P}_i) + \tilde{\mathbb{S}}_i^\pi \tilde{K}_i \mathbb{P}_i \right) \right] \\ &= \mathbb{E} \left[\sum_{i \in F} \mathbb{S}_i^\pi K_i (1 - \mathbb{P}_i) + \sum_{i \in P} \tilde{\mathbb{S}}_i^\pi \tilde{K}_i \mathbb{P}_i \right] \\ &\leq \mathbb{E} \left[\max \left\{ \max_{i \in F} \{K_i\}, \max_{i \in P} \{\tilde{K}_i\} \right\} \right], \end{aligned}$$

where the first equality holds due to the construction of the committing policy $\pi^{F,P}$, and the inequality follows from the constraints $\sum_{i=1}^N \mathbb{S}_i^\pi + \sum_{i=1}^N \tilde{\mathbb{S}}_i^\pi \leq 1$, and $0 \leq \mathbb{P}_i^\pi \leq 1$.

To conclude the proof, it suffices to show that this upper bound is matched by $\pi^{F,P}$ in the real-world problem. To see this, we first remark that the committing policy $\pi^{F,P}$ is non-exposed, and thus, by Lemma EC.7, $J^{\pi^{F,P}}(s_0) = J^{\pi^{F,P},M}(s_0)$. Therefore, to prove the lemma, it suffices to show that in the free-info problem, policy $\pi^{F,P}$ achieves the maximal capped realization. Letting $\hat{K}_i = K_i$

if $i \in F$ and $\widehat{K}_i = \widetilde{K}_i$ if $i \in P$ be the capped value of each box i , we would like to show that that policy $\pi^{F,P}$ selects $\max_{i \in [N]} \{\widehat{K}_i\}$ under any realization.

Let j denote the box selected by the policy. We note that, at termination, except for \widehat{K}_j , there are two groups of boxes: (1) boxes that were F-opened (perhaps after P-opening) ; and (2) boxes that were not F-opened (but possibly were P-opened). We argue that the capped value of each box in both groups is smaller than or equal to \widehat{K}_j .

Specifically, for any box i in group 1, observe that

- Fact 1: $V_i \leq V_j$. This is due to the fact that policy $\pi^{F,P}$ always selects the best prize so far, and thus $V_i \leq V_j$;

- Fact 2: $V_i < \sigma_j$. To see this, note that if $V_j \geq \sigma_j$ the policy immediately stops after F-opening box j . In this case, $V_i \leq \sigma_j$ (otherwise box j would not have been opened). Alternatively, if $V_j < \sigma_j$, then it follows from fact 1 that $V_i < \sigma_j$.

Recalling that $\widehat{K}_i \leq V_i$, we have $\widehat{K}_j \geq \widehat{K}_i$ for any box i in group (1).

Next, for any box i in group (2), we observe that

- Fact 3: $\sigma_j \geq \sigma_i \geq \widehat{K}_i$, since policy $\pi^{F,P}$ always opens the box with largest threshold;
- Fact 4: $V_j \geq \sigma_i \geq \widehat{K}_i$, since policy $\pi^{F,P}$ obeys the optimal stopping rule.

Hence, \widehat{K}_j is indeed the largest capped value. It then follows that policy $\pi^{F,P}$ attains the largest capped realization.

C.4. Proof of Lemma EC.7

Recall the expressions of $J^\pi(s_0)$ and $J^{\pi,K}(s_0)$

$$J^\pi(s_0) = \mathbb{E} \left[\sum_{i=1}^N \left(\mathbb{S}_i^\pi V_i - \mathbb{F}_i^\pi c_i^F - \mathbb{P}_i^\pi c_i^P \right) + \sum_{i=1}^N \left(\widetilde{\mathbb{S}}_i^\pi V_i - \widetilde{\mathbb{F}}_i^\pi c_i^F \right) \right],$$

and

$$J^{\pi,K}(s_0) = \mathbb{E} \left[\sum_{i=1}^N \left(\mathbb{S}_i^\pi K_i (1 - \mathbb{P}_i) + \widetilde{\mathbb{S}}_i^\pi \widetilde{K}_i \mathbb{P}_i \right) \right].$$

To prove the lemma¹, it suffices to show that for all i ,

$$\mathbb{E} \left[\mathbb{S}_i V_i + \widetilde{\mathbb{S}}_i V_i - \mathbb{F}_i c_i^F - \widetilde{\mathbb{F}}_i c_i^F - \mathbb{P}_i c_i^P \right] \leq \mathbb{E} \left[\mathbb{S}_i K_i (1 - \mathbb{P}_i) + \widetilde{\mathbb{S}}_i \widetilde{K}_i \mathbb{P}_i \right],$$

and that the inequalities hold with equalities if and only if the policy satisfies the three criteria for all boxes i . We proceed to show that this is indeed true.

¹ For brevity, we mute the superscripts of $\mathbb{S}_i^\pi, \widetilde{\mathbb{S}}_i^\pi, \mathbb{F}_i^\pi, \widetilde{\mathbb{F}}_i^\pi, \mathbb{P}_i^\pi$ below in this proof.

For any box $i \in [N]$, its contribution to the total expected profit is

$$\underbrace{\mathbb{E}[\mathbb{S}_i V_i - \mathbb{F}_i c_i^F]}_{\textcircled{1}} + \underbrace{\mathbb{E}[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{F}}_i c_i^F - \mathbb{P}_i c_i^P]}_{\textcircled{2}},$$

where the term $\textcircled{1}$ corresponds to the expected profit from box i by directly F-opening it, while the term $\textcircled{2}$ corresponds to a P-opening followed by a F-opening. We derive upper bounds on each of two terms, and show these bounds are exactly tight for a non-exposed policy.

The first condition of non-exposed policies. For the first term, we note that

$$\textcircled{1} = \mathbb{E}[\mathbb{S}_i V_i - \mathbb{F}_i c_i^F \mid \mathbb{P}_i = 0] \Pr[\mathbb{P}_i = 0] + \underbrace{\mathbb{E}[\mathbb{S}_i V_i - \mathbb{F}_i c_i^F \mid \mathbb{P}_i = 1]}_{=0} \Pr[\mathbb{P}_i = 1] \quad (\text{EC.15})$$

$$= \mathbb{E}[\mathbb{S}_i V_i - \mathbb{F}_i \mathbb{E}[(V_i - \sigma_i^F)^+ \mid \mathbb{P}_i = 0] \mid \mathbb{P}_i = 0] \Pr[\mathbb{P}_i = 0] \quad (\text{EC.16})$$

$$= \mathbb{E}[\mathbb{S}_i V_i - \mathbb{F}_i (V_i - \sigma_i^F)^+ \mid \mathbb{P}_i = 0] \Pr[\mathbb{P}_i = 0] \quad (\text{EC.17})$$

$$\leq \mathbb{E}[\mathbb{S}_i V_i - \mathbb{S}_i (V_i - \sigma_i^F)^+ \mid \mathbb{P}_i = 0] \Pr[\mathbb{P}_i = 0] \quad (\text{EC.18})$$

$$= \mathbb{E}[\mathbb{S}_i K_i \mid \mathbb{P}_i = 0] \Pr[\mathbb{P}_i = 0], \quad (\text{EC.19})$$

where equation (EC.15) is obtained by conditioning on the decision to P-open, and by showing that the underbraced expression is equal to zero. The latter fact follows by noting that $\mathbb{S}_i \leq \mathbb{F}_i = 1 - \mathbb{P}_i = 0$ when $\mathbb{P}_i = 1$. Equation (EC.16) follows from the definition of the F-threshold σ_i^F , and equation (EC.17) is due to the independence of \mathbb{F}_i and V_i , since the decision of directly F-opening or not does not depend on the realization of V_i . The inequality again proceeds from constraint $\mathbb{S}_i \leq \mathbb{F}_i$. We observe that this inequality is tight if and only if $\mathbb{E}[(\mathbb{F}_i - \mathbb{S}_i)(V_i - \sigma_i^F)^+ \mid \mathbb{P}_i = 0] = 0$. In particular, when $V_i \leq \sigma_i^F$, this inequality is satisfied with equality. Moreover, when V_i satisfies $V_i > \sigma_i^F$, the first criterion in the definition of non-exposed policies is a sufficient and necessary condition for the box to be selected, in which case $\mathbb{S}_i = \mathbb{F}_i = 1$ and the above inequality is satisfied with equality. Finally, equation (EC.19) immediately follows from the definition of K_i . We mention in passing that our analysis of term $\textcircled{1}$ is nearly identical to that of Kleinberg et al. (2016)[Lemma 1].

The second condition of non-exposed policies. Now, let us turn to the second term:

$$\textcircled{2} = \mathbb{E}[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{F}}_i c_i^F - \mathbb{P}_i c_i^P \mid \mathbb{P}_i = 1] \Pr[\mathbb{P}_i = 1] + \underbrace{\mathbb{E}[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{F}}_i c_i^F - \mathbb{P}_i c_i^P \mid \mathbb{P}_i = 0]}_{=0} \Pr[\mathbb{P}_i = 0] \quad (\text{EC.20})$$

$$= \mathbb{E}[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{F}}_i c_i^F - \mathbb{P}_i \mathbb{E}[\max\{0, -c_i^F + \mathbb{E}[(V_i - \sigma_i^P)^+ \mid T_i]\} \mid \mathbb{P}_i = 1] \mid \mathbb{P}_i = 1] \Pr[\mathbb{P}_i = 1] \quad (\text{EC.21})$$

$$= \mathbb{E}[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{F}}_i c_i^F - \mathbb{P}_i \max\{0, -c_i^F + \mathbb{E}[(V_i - \sigma_i^P)^+ \mid T_i]\} \mid \mathbb{P}_i = 1] \Pr[\mathbb{P}_i = 1] \quad (\text{EC.22})$$

$$\leq \mathbb{E}[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{F}}_i c_i^F - \tilde{\mathbb{F}}_i \max\{0, -c_i^F + \mathbb{E}[(V_i - \sigma_i^P)^+ \mid T_i]\} \mid \mathbb{P}_i = 1] \Pr[\mathbb{P}_i = 1], \quad (\text{EC.23})$$

where equation (EC.20) is obtained by conditioning on the decision to P-open, and the underbraced term is equal to zero when $\mathbb{P}_i = 0$ (this fact can be verified from the definition of $\widetilde{\mathbb{S}}_i$ and $\widetilde{\mathbb{F}}_i$). Equation (EC.21) is due to definition of the P-threshold σ_i^P . Equation (EC.22) is due to the independence of \mathbb{P}_i and T_i , since the decision to partially open does not depend on the realization of the type T_i . The inequality (EC.23) holds since $\widetilde{\mathbb{F}}_i \leq \mathbb{P}_i$.

We next examine when the inequality EC.23 is tight. To this end, let $\varphi_i(t_i) = \max \{0, -c_i^F + \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i]\}$. We can write the difference between equations (EC.23) and (EC.22) as follows:

$$\begin{aligned} & (\text{EC.23}) - (\text{EC.22}) \\ &= \mathbb{E} \left[\left(\mathbb{P}_i - \widetilde{\mathbb{F}}_i \right) \varphi(T_i) \mid \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \end{aligned} \quad (\text{EC.24})$$

$$= \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbf{I}\{\varphi(t_i) = 0\} \Pr[\mathbb{P}_i = 1] \cdot 0 \quad (\text{EC.25})$$

$$+ \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbf{I}\{\varphi(t_i) > 0\} \Pr[\mathbb{P}_i = 1] \Pr[\widetilde{\mathbb{F}}_i = 1 \mid \mathbb{P}_i = 1, T_i = t_i] \cdot 0 \quad (\text{EC.26})$$

$$+ \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbf{I}\{\varphi(t_i) > 0\} \Pr[\mathbb{P}_i = 1] \Pr[\widetilde{\mathbb{F}}_i = 0 \mid \mathbb{P}_i = 1, T_i = t_i] \cdot \underbrace{\varphi_i(t_i)}_{>0}, \quad (\text{EC.27})$$

In equation (EC.24), we substitute the function $\varphi_i(t_i)$ into the difference between (EC.23) and (EC.22). We then decompose the expectation into three summations based on the type t_i . Specifically, expression (EC.25) corresponds to the “bad” types t_i , for which F-opening is not worthwhile. Expression (EC.26) corresponds to the case where box i is F-opened ($\widetilde{\mathbb{F}}_i = 1$) after being P-opened ($\mathbb{P}_i = 1$). Finally, expression (EC.27) corresponds to the case where box i is P-opened ($\mathbb{P}_i = 1$) but not F-opened ($\widetilde{\mathbb{F}}_i = 0$). Therefore, equations (EC.23) and (EC.22) are equal if and only if $\varphi_i(t_i) > 0$ implies that $\Pr[\widetilde{\mathbb{F}}_i = 0 \mid \mathbb{P}_i = 1, T_i = t_i] = 0$. Recall that $c_i^F = \mathbb{E}_{V_i|T_i=t_i}[(V_i - \sigma_i^{F|t_i})^+]$ by definition of $\sigma_i^{F|t_i}$, and that $\varphi_i(t_i) = \max \{0, -c_i^F + \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i]\}$. Therefore, we have

$$\begin{aligned} \varphi_i(t_i) > 0 &\Leftrightarrow -c_i^F + \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i] > 0 \\ &\Leftrightarrow \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i] > c_i^F \\ &\Leftrightarrow \mathbb{E}_{V_i|T_i=t_i}[(V_i - \sigma_i^P)^+] > \mathbb{E}_{V_i|T_i=t_i}[(V_i - \sigma_i^{F|t_i})^+] \\ &\Leftrightarrow \sigma_i^{F|t_i} > \sigma_i^P. \end{aligned}$$

This is exactly the second criterion in the definition of non-exposed policies, which requires that if $\sigma_i^{F|t_i} > \sigma_i^P$ then F-opening will succeed a P-opening.

The third condition of non-exposed policies. Now we get back to characterizing an upper bound on term ②:

$$\begin{aligned}
& \mathbb{E} \left[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{F}}_i \left(c_i^F + \max\{0, -c_i^F + \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i] \} \right) \mid \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \\
&= \mathbb{E} \left[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{F}}_i \left(\max\{c_i^F, \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i] \} \right) \mid \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \\
&= \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbb{I}\{c_i^F \geq \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i]\} \mathbb{E} \left[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{F}}_i c_i^F \mid T_i = t_i, \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \\
&+ \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbb{I}\{c_i^F < \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i]\} \mathbb{E} \left[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{F}}_i \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i] \mid T_i = t_i, \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \quad (\text{EC.28})
\end{aligned}$$

$$\begin{aligned}
&= \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbb{I}\{c_i^F \geq \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i]\} \mathbb{E} \left[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{F}}_i \mathbb{E}[(V_i - \sigma_i^{F|t_i})^+ | T_i = t_i] \mid T_i = t_i, \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \\
&+ \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbb{I}\{c_i^F < \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i]\} \mathbb{E} \left[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{F}}_i \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i] \mid T_i = t_i, \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \quad (\text{EC.29})
\end{aligned}$$

$$\begin{aligned}
&= \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbb{I}\{c_i^F \geq \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i]\} \mathbb{E} \left[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{F}}_i (V_i - \sigma_i^{F|t_i})^+ \mid T_i = t_i, \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \\
&+ \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbb{I}\{c_i^F < \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i]\} \mathbb{E} \left[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{F}}_i (V_i - \sigma_i^P)^+ \mid T_i = t_i, \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \quad (\text{EC.30})
\end{aligned}$$

The first equation is identical to equation (EC.23). Equation (EC.28) is obtained by conditioning on the type t_i revealed after P-opening box i . Equation (EC.29) is derived by substituting the definition of $\sigma_i^{F|t_i}$ (cf. Definition 1). We obtain equation (EC.30) by consolidating the expectation operator around $(V_i - \sigma_i^P)^+$ and $(V_i - \sigma_i^{F|t_i})^+$ and by observing the following sequence of equalities:

$$\begin{aligned}
& \mathbb{E} \left[\tilde{\mathbb{F}}_i \mathbb{E} \left[(V_i - \sigma_i^{F|t_i})^+ \mid T_i = t_i \right] \mid T_i = t_i, \mathbb{P}_i = 1 \right] \\
&= \mathbb{E} \left[\tilde{\mathbb{F}}_i \mid T_i = t_i, \mathbb{P}_i = 1 \right] \mathbb{E} \left[(V_i - \sigma_i^{F|t_i})^+ \mid T_i = t_i, \mathbb{P}_i = 1 \right] \\
&= \mathbb{E} \left[\tilde{\mathbb{F}}_i (V_i - \sigma_i^{F|t_i})^+ \mid T_i = t_i, \mathbb{P}_i = 1 \right].
\end{aligned}$$

The first equality follows from the fact that $\mathbb{E} \left[(V_i - \sigma_i^{F|t_i})^+ \mid T_i = t_i \right]$ is a constant, independent of $\tilde{\mathbb{F}}_i$. The second equality follows from the independence between $\tilde{\mathbb{F}}_i$ and V_i conditional on $T_i = t_i$. Moving forward, we have:

$$\begin{aligned}
(\text{EC.30}) &= \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbb{I}\{c_i^F \geq \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i]\} \mathbb{E} \left[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{F}}_i (V_i - \sigma_i^{F|t_i})^+ \mid T_i = t_i, \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \\
&+ \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbb{I}\{c_i^F < \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i]\} \mathbb{E} \left[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{F}}_i (V_i - \sigma_i^P)^+ \mid T_i = t_i, \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \quad (\text{EC.31})
\end{aligned}$$

$$\begin{aligned}
&\leq \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbb{I}\{c_i^F \geq \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i]\} \mathbb{E} \left[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{S}}_i (V_i - \sigma_i^{F|t_i})^+ \mid T_i = t_i, \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \\
&+ \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbb{I}\{c_i^F < \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i]\} \mathbb{E} \left[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{S}}_i (V_i - \sigma_i^P)^+ \mid T_i = t_i, \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \quad (\text{EC.32})
\end{aligned}$$

$$\begin{aligned}
&= \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbb{I}\{\sigma_i^{F|t_i} \leq \sigma_i^P\} \mathbb{E} \left[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{S}}_i (V_i - \sigma_i^{F|t_i})^+ \mid T_i = t_i, \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \\
&+ \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbb{I}\{\sigma_i^{F|t_i} > \sigma_i^P\} \mathbb{E} \left[\tilde{\mathbb{S}}_i V_i - \tilde{\mathbb{S}}_i (V_i - \sigma_i^P)^+ \mid T_i = t_i, \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \quad (\text{EC.33})
\end{aligned}$$

$$\begin{aligned}
&= \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbb{I}\{\sigma_i^{F|t_i} \leq \sigma_i^P\} \mathbb{E} \left[\tilde{\mathbb{S}}_i \min\{V_i, \sigma_i^{F|t_i}\} \mid T_i = t_i, \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \\
&+ \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbb{I}\{\sigma_i^{F|t_i} > \sigma_i^P\} \mathbb{E} \left[\tilde{\mathbb{S}}_i \min\{V_i, \sigma_i^P\} \mid T_i = t_i, \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \quad (\text{EC.34})
\end{aligned}$$

$$= \mathbb{E} \left[\tilde{\mathbb{S}}_i \min\{V_i, \sigma_i^P, \sigma_i^{F|T_i}\} \mid \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1] \quad (\text{EC.35})$$

$$= \mathbb{E} \left[\tilde{\mathbb{S}}_i \tilde{K}_i \mid \mathbb{P}_i = 1 \right] \Pr[\mathbb{P}_i = 1]. \quad (\text{EC.36})$$

The first equation is identical to equation EC.30. The inequality EC.32 follows from the inequality $\tilde{\mathbb{S}}_i \leq \tilde{\mathbb{F}}_i$. For equation(EC.33), note that by definition of $\sigma_i^{F|t_i}$,

$$\begin{aligned} c_i^F \geq \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i] &\Leftrightarrow \mathbb{E}[(V_i - \sigma_i^{F|t_i})^+ | T_i = t_i] \geq \mathbb{E}[(V_i - \sigma_i^P)^+ | T_i = t_i] \\ &\Leftrightarrow \sigma_i^{F|t_i} \leq \sigma_i^P, \end{aligned}$$

where the last equivalence follows from the monotonicity of the mapping $x \rightarrow \mathbb{E}[(V_i - x)^+ | T_i = t_i]$. Equation (EC.34) follows from the identity $x - (x - y)^+ = \min\{x, y\}$. Equation (EC.35) can be justified by conditioning on T_i , and finally, equation(EC.36) follows from the definition of \tilde{K}_i .

To identify under which condition we have (EC.31) = (EC.32), we consider the difference between these terms:

$$\begin{aligned} (\text{EC.32}) - (\text{EC.31}) &= \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbf{I}\{\sigma_i^{F|t_i} \leq \sigma_i^P\} \mathbb{E}[(\tilde{\mathbb{F}}_i - \tilde{\mathbb{S}}_i)(V_i - \sigma_i^{F|t_i})^+ | T_i = t_i] \Pr[\mathbb{P}_i = 1] \\ &\quad + \sum_{t_i \in \Gamma_i} \Pr[T_i = t_i] \mathbf{I}\{\sigma_i^{F|t_i} > \sigma_i^P\} \mathbb{E}[(\tilde{\mathbb{F}}_i - \tilde{\mathbb{S}}_i)(V_i - \sigma_i^P)^+ | T_i = t_i] \Pr[\mathbb{P}_i = 1]. \end{aligned}$$

We note that the latter sum-expressions are equal to zero if and only if the policy π is designed such that

- $\{V_i > \sigma_i^{F|t_i}\} \cap \{\sigma_i^{F|t_i} \leq \sigma_i^P\} \cap \{\tilde{\mathbb{F}}_i = 1\} \Rightarrow \{\tilde{\mathbb{S}}_i = 1\}$, and
- $\{V_i > \sigma_i^P\} \cap \{\sigma_i^{F|t_i} > \sigma_i^P\} \cap \{\tilde{\mathbb{F}}_i = 1\} \Rightarrow \{\tilde{\mathbb{S}}_i = 1\}$.

The latter conditions can be reformulated as $\{V_i > \min\{\sigma_i^P, \sigma_i^{F|t_i}\}\} \cap \{\tilde{\mathbb{F}}_i = 1\} \Rightarrow \{\tilde{\mathbb{S}}_i = 1\}$. This is precisely the third criterion in the definition of non-exposed policies.

Finally, combining inequalities (EC.19) and (EC.36) yields that the expected profit from box i is bounded above by:

$$\begin{aligned} \mathbb{E}[\mathbb{S}_i V_i + \tilde{\mathbb{S}}_i V_i - \mathbb{F}_i c_i^F - \tilde{\mathbb{F}}_i c_i^F - \mathbb{P}_i c_i^P] &\leq \mathbb{E}[\mathbb{S}_i K_i | \mathbb{P}_i = 0] \Pr[\mathbb{P}_i = 0] + \mathbb{E}[\tilde{\mathbb{S}}_i \tilde{K}_i | \mathbb{P}_i = 1] \Pr[\mathbb{P}_i = 1] \\ &= \mathbb{E}[\mathbb{S}_i K_i (1 - \mathbb{P}_i) + \tilde{\mathbb{S}}_i \tilde{K}_i \mathbb{P}_i]. \end{aligned}$$

Following from the discussion above, we know that the inequalities hold as equalities for all i if and only if the policy is non-exposed.

C.5. Proof of Theorem 6

In the spirit of Beyhaghi and Kleinberg (2019), we cast our free-info problem into a respective stochastic submodular optimization problem. The adaptivity gap result by Asadpour and Nazerzadeh

(2015) guarantees that the performance of some committing policy is comparable to a fully adaptive policy, up to a constant-factor approximation.

The proof is organized as follows. We begin by summarizing results from Asadpour and Nazerzadeh (2015) on stochastic monotone submodular optimization, in section C.5.1. In section C.5.2, we then formulate specific submodular optimization problems related to our Pandora's box with sequential inspection. In section C.5.3, we continue to define the transformation of policies between the free-info problem and the submodular problem and prove relations between these transformations. Finally, in section C.5.4, we combine earlier results to prove the main theorem.

C.5.1. Submodular monotone nonnegative maximization subject to a matroid constraint.

We consider the class of stochastic optimization problems, represented by a three-tuple $(f, \mathcal{A}, \mathcal{F})$, where the goal is to maximize the expectation of a stochastic monotone submodular function f , subject to a matroid constraint $\mathcal{M} = (\mathcal{A}, \mathcal{F})$. That is, $f : \mathbb{R}_+^n \rightarrow \mathbb{R}_+$ satisfies

$$\forall x, y \in \mathbb{R}_+^n : f(x \vee y) + f(x \wedge y) \leq f(x) + f(y),$$

where we denote by $x \vee y$ the element-wise maximum and $x \wedge y$ the element-wise minimum of vectors x and y . The matroid \mathcal{M} captures the feasibility constraint, where $\mathcal{A} = \{X_1, X_2, \dots, X_n\}$ is a ground set of n independent random variables, and \mathcal{F} is a collection of independent sets.²

A policy π generates a sequence of *selections* of elements in the ground set. More precisely, it is a mapping from the set of all possible states to a distribution over elements in the ground set. After each selection, the realization of the corresponding element is revealed and the policy may adapt to this new information accordingly. The state of this problem Θ can be described by an n -tuple $(\theta_1, \theta_2, \dots, \theta_n)$, where θ_i denotes the realization of X_i , if $X_i \in \mathcal{A}$ is selected by the policy, and is equal to \circ otherwise, which means this element has not been selected yet. We say a policy is adaptive if the selection decisions are based on the outcomes of the previous actions; otherwise, we say that a policy is nonadaptive. Denote by Θ^π the random n -tuple corresponding to the last state reached by the policy π . The value of this policy is defined to be $\mathbb{E}[f(\Theta^\pi)]$. We denote by $\mathcal{F}(\mathcal{A}, \mathcal{F})$ the set of feasible policies, in which the set of elements chosen by π always belong to \mathcal{F} .

The goal of this optimization problem is to find a feasible policy with respect to \mathcal{M} that obtains the maximum expected value of its final state, i.e.,

$$\max_{\pi \in \mathcal{F}(\mathcal{A}, \mathcal{F})} \mathbb{E}[f(\Theta^\pi)].$$

² Definition of a matroid $\mathcal{M} = (\mathcal{A}, \mathcal{F})$: all elements of \mathcal{F} are subsets of \mathcal{A} satisfying the following properties: (i) $\emptyset \in \mathcal{F}$; (ii) If $A \in \mathcal{F}$ and $B \subseteq A$, then $B \in \mathcal{F}$; (iii) If $A \in \mathcal{F}$, $B \in \mathcal{F}$ and $|B| > |A|$, then there exists $b \in B \setminus A$ with $A \cup \{b\} \in \mathcal{F}$.

In what follows, we restate the main result of Asadpour and Nazerzadeh (2015) as a theorem, which we will use as the main machinery for our analysis.

THEOREM EC.1. (*Asadpour and Nazerzadeh (2015), Theorem 1*). *There exists a nonadaptive policy that achieves $1 - \frac{1}{e} \approx 0.63$ fraction of the optimal policy in maximizing a stochastic monotone submodular function with respect to a matroid constraint $(f, \mathcal{A}, \mathcal{F})$.*

C.5.2. Modeling Pandora's box problem with sequential inspection as a submodular optimization problem. The Pandora's box problem can be essentially viewed as a selection problem. Following Lemma 2 and the discussion above, we define two associated problems that capture exactly the tradeoff of inspection modes.

DEFINITION EC.1. (The associated stochastic submodular optimization problems). Given an instance of a real-world problem \mathcal{I} , which has N boxes with costs and distributions $\{c_i^F, c_i^P, \mathcal{D}_i\}_{i=1}^N$, we define two stochastic submodular optimization problems $\mathcal{J} = (f, \mathcal{A}, \mathcal{F})$ and $\mathcal{J}' = (f, \mathcal{A}', \mathcal{F})$ associated with the free-info problem:

- *Ground sets.* We define the ground set \mathcal{A} is

$$\mathcal{A} = \{K_1, \tilde{K}_1, K_2, \tilde{K}_2, \dots, K_N, \tilde{K}_N\},$$

where K_i and \tilde{K}_i are correlated and induced from \mathcal{D}_i . Thus, \mathcal{I} and \mathcal{J} are defined over the same probability space. However, the ground set \mathcal{A}' is

$$\mathcal{A}' = \{K'_1, \tilde{K}'_1, K'_2, \tilde{K}'_2, \dots, K'_N, \tilde{K}'_N\},$$

where K'_i and \tilde{K}'_i are drawn independently from the same distribution as K_i and \tilde{K}_i .

- *Matroid constraint.* The matroid constraint \mathcal{M} is defined by the partition matroid whose independent sets are all the subsets of \mathcal{A} (resp., \mathcal{A}') that contain at most one element of each pair $\{K_i, \tilde{K}_i\}_{i=1}^N$ (resp., $\{K'_i, \tilde{K}'_i\}_{i=1}^N$).

- *Objective function.* The submodular objective function $f : \mathbb{R}_+^{2N} \rightarrow \mathbb{R}_+$ is defined as

$$f(\Theta) = \max\{\Theta_i : 1 \leq i \leq 2N, \theta_i \neq \text{NULL}\}.$$

We denote the expected profit of policy π in \mathcal{J} by $J^{\pi, M}(s_0)$ and in \mathcal{J}' by $J^{\pi, M'}(s_0)$.

Remark. Clearly, any feasible policy π for problem \mathcal{J} is also feasible for problem \mathcal{J}' (that is, $\pi \in \mathfrak{F}(\mathcal{A}, \mathcal{F}) \Leftrightarrow \pi \in \mathfrak{F}(\mathcal{A}', \mathcal{F})$). A perhaps simple, yet crucial observation is that for any feasible policy π , we have $J^{\pi, M}(s_0) = J^{\pi, M'}(s_0)$. The equality holds by noting that only one single variable from each pair $\{K_i, \tilde{K}_i\}$ or $\{K'_i, \tilde{K}'_i\}$, is picked by any feasible policy.

Summary of Notations for Theorem 6	
\mathcal{I}	A real-world instance
\mathcal{J}	The corresponding stochastic submodular problem
ρ	A nonadaptive policy in \mathcal{J}
$\Psi(\rho)$	A committing policy in \mathcal{I} induced by ρ
π	Any policy in \mathcal{I}
$\Phi(\pi)$	The transformed policy in \mathcal{J} from policy π

C.5.3. Transformation of policies. Let ρ be a *nonadaptive* policy in \mathcal{J} and $A(\rho) \subseteq \mathcal{A}$ be the set of random variables that ρ selects. Let $F(\rho)$ denote the (deterministic) set of boxes $\{i \in [N] \mid K_i \in A(\rho)\}$ and let $P(\rho)$ denote the (deterministic) set of boxes $\{i \in [N] \mid \tilde{K}_i \in A(\rho)\}$. Clearly, the expected profit of policy ρ in \mathcal{J} is

$$J^{\rho, M}(s_0) = \mathbb{E} \left[\max \left\{ \max_{i \in F(\rho)} \{K_i\}, \max_{i \in P(\rho)} \{\tilde{K}_i\} \right\} \right].$$

Moreover, ρ naturally induces a corresponding committing policy $\Psi(\rho)$ in \mathcal{I} by specifying $F = F(\rho)$ and $P = P(\rho)$. We mention in passing that $F \cup P$ might not be equal to $[N]$, but as long as $F \cap P = \emptyset$, Lemma 2 still holds. Thus, we have

$$J^{\Psi(\rho)}(s_0) = \mathbb{E} \left[\max \left\{ \max_{i \in F(\rho)} \{K_i\}, \max_{i \in P(\rho)} \{\tilde{K}_i\} \right\} \right].$$

Since \mathcal{I} and \mathcal{J} share the same probability space, we have the following lemma:

LEMMA EC.8. *Let ρ be a nonadaptive policy in \mathcal{J} , and $\Psi(\rho)$ be the corresponding committing policy in \mathcal{I} . We have $J^{\Psi(\rho)}(s_0) = J^{\rho, M}(s_0) = J^{\rho, M'}(s_0)$.*

Conversely, let π be any policy in \mathcal{I} . Since problem \mathcal{I} and problem \mathcal{J} share the same probability space, we transform π into policy $\Phi(\pi)$ in \mathcal{J} in the following way: (i) if policy π directly F-opens box i without a P-opening in \mathcal{I} , then policy $\Phi(\pi)$ selects K_i in \mathcal{J} ; (ii) if policy π P-opens box i in \mathcal{I} , then policy $\Phi(\pi)$ selects \tilde{K}_i in \mathcal{J} ; (iii) if policy π stops in \mathcal{I} , then policy $\Phi(\pi)$ does not select any other box as well.

LEMMA EC.9. *Let π be a policy in \mathcal{I} , and $\Phi(\pi)$ be the corresponding transformed policy in \mathcal{J} . We have $J^{\Phi(\pi), M'}(s_0) = J^{\Phi(\pi), M}(s_0) \geq J^\pi(s_0)$.*

Proof: By construction, given any policy π in \mathcal{I} , the *random* set of random variables selected by the transformed policy $\Phi(\pi)$ in \mathcal{J} is $\{K_i \mid \mathbb{P}_i^\pi = 1\} \cup \{\tilde{K}_i \mid \mathbb{P}_i^\pi = 1\}$. Let $F(\pi) = \{i \in [N] \mid \mathbb{P}_i^\pi = 1\}$ be the random set of boxes that is directly F-opened by policy π , and similarly, let $P(\pi) = \{i \in [N] \mid$

$\mathbb{P}_i^\pi = 1\}$ be the set of boxes that will be P-opened first by policy π . Thus, the expected profit of $\Phi(\pi)$ in \mathcal{J} is

$$J^{\Phi(\pi),M}(s_0) = \mathbb{E} \left[\max \left\{ \max_{i \in F(\pi)} \{K_i\}, \max_{i \in P(\pi)} \{\tilde{K}_i\} \right\} \right]. \quad (\text{EC.37})$$

We emphasize that the expectation is taken over $F(\pi)$ and $P(\pi)$ in addition to K_i and \tilde{K}_i , since the policy can adaptively adjust which inspection mode to apply on boxes based on the observed information. On the other hand, we note that

$$\begin{aligned} J^\pi(s_0) &\leq \mathbb{E} \left[\sum_{i=1}^N \mathbb{S}_i^\pi K_i (1 - \mathbb{P}_i^\pi) + \tilde{\mathbb{S}}_i^\pi \tilde{K}_i \mathbb{P}_i^\pi \right] \\ &= \mathbb{E} \left[\sum_{i \in F(\pi)} \mathbb{S}_i^\pi K_i + \sum_{i \in P(\pi)} \tilde{\mathbb{S}}_i^\pi \tilde{K}_i \right] \\ &\leq \mathbb{E} \left[\max \left\{ \max_{i \in F(\pi)} \{K_i\}, \max_{i \in P(\pi)} \{\tilde{K}_i\} \right\} \right], \end{aligned} \quad (\text{EC.38})$$

$$= J^{\Phi(\pi),M}(s_0), \quad (\text{EC.39})$$

where inequality EC.38 proceeds from the constraint that only one box can be selected. Equation EC.39 follows by noting equation EC.37, and that \mathcal{I} and \mathcal{J} share the same probability space.

□

Armed with the preceding results, we are ready to establish our main theorem.

C.5.4. Proof of Theorem 6. Let π be the optimal policy in \mathcal{I} . By Lemma EC.9, we have

$$J^{\Phi(\pi),M'}(s_0) = J^{\Phi(\pi),M}(s_0) \geq J^\pi(s_0) \triangleq J(s_0). \quad (\text{EC.40})$$

By Theorem EC.1, there exists a nonadaptive policy ρ in \mathcal{J}' such that

$$J^{\rho,M}(s_0) \geq \left(1 - \frac{1}{e}\right) J^{\Phi(\pi),M'}(s_0). \quad (\text{EC.41})$$

By Lemma EC.8, we know that the transformed policy $\Psi(\rho)$ in \mathcal{I} satisfies

$$J^{\Psi(\rho)}(s_0) = J^{\rho,M}(s_0) = J^{\rho,M'}(s_0). \quad (\text{EC.42})$$

Combining EC.40, EC.41 and EC.42 together yields

$$J^{\Psi(\rho)}(s_0) \geq \left(1 - \frac{1}{e}\right) J(s_0).$$

Finally, we remark here that a good committing policy can be found efficiently, by invoking another technical result established in Asadpour and Nazerzadeh (2015), which we restate as follows.

THEOREM EC.2. (Asadpour and Nazerzadeh (2015), Theorem 2). For any $\epsilon > 0$ and any instance $(f, \mathcal{A}, \mathcal{F})$, a nonadaptive policy that obtains a $(1 - 1/e - \epsilon)$ -fraction of the value of the optimal adaptive policy can be found in polynomial time with respect to $|\mathcal{A}|$ and $1/\epsilon$.

C.6. Proof of Theorem 5

In this section, we show that computing optimal policies for *Pandora box problem with sequential inspection* (PSI) is NP-hard. We devise the reduction from a family of NP-hard subset sum problems.

The hard instances. We devise a reduction from the subset sum problem (SSP). We are given positive integers a_1, \dots, a_n and the goal is to find $W \subseteq [n]$ such that $\sum_{i \in W} a_i = T$. Without loss of generality, we assume that $T \leq \frac{1}{2} \sum_{i=1}^n a_i$ as otherwise, we consider the complementary target $\sum_{i=1}^n a_i - T$, corresponding to selecting the complement in $[n]$. For brevity, we use the notation $L \stackrel{\text{def}}{=} \sum_{i=1}^n a_i$ throughout. Moreover, we focus on a family of SSP which we call the ε -SSP, where we have $\frac{a_i}{\sum_{i=1}^n a_i} \leq \varepsilon$ for every $i \in [n]$ for some $\varepsilon \in (0, 1)$. This family is still NP-hard. To see this, suppose there exists a polynomial-time algorithm that solves the ε_0 -SSP for some fixed ε_0 . Now given a general SSP instance $\{\{a_i\}_{i=1}^n, T\}$, we can transform it to a ε_0 -SSP instance by adding “fake items” $a_0 > L$. Since $a_0 > L$, we know for sure we will not include them in a feasible solution. In addition, we can ensure for every $i \in [n]$, we have $\frac{a_i}{N \cdot a_0 + \sum_{i=1}^n a_i} < \frac{a_0}{N \cdot a_0 + \sum_{i=1}^n a_i} < \varepsilon_0$ for large enough $N = O(\frac{1}{\varepsilon_0})$ (independent of n). The following lemma ensues.

LEMMA EC.10. The ε -subset sum problem is NP-hard for any fixed $\varepsilon \in (0, 1)$.

Our reduction makes use of the following class of PSI instances.

DEFINITION EC.2 (THE 2T2L-PSI INSTANCE). Given positive integers $\{a_i\}_{i=1}^n$, an instance of PSI with n boxes is a two-type two-support low-cost-low-return (2T2L) instance with parameter (ν, α) where $\nu \in (0, 1), \alpha > 0$ if the following conditions hold:

1. all rewards are binary random variables $V_i \in \{0, 1\}$. Specifically, for box i , we set its success probability as

$$p_i \stackrel{\text{def}}{=} \Pr[V_i = 1] = 1 - \exp(-\alpha a_i),$$

for some constant $\alpha > 0$. We define p'_i as

$$p'_i \stackrel{\text{def}}{=} 1 - \exp\left(-\frac{1}{2}\alpha a_i\right);$$

2. each box i has 2 types: good and bad, $\Gamma_i = \{G, B\}$, and moreover $\Pr[T_i = G] = (1 - \nu)p_i$, $\Pr[T_i = B] = 1 - (1 - \nu)p_i$, $\Pr[V_i = 1 | T_i = G] = \frac{p'_i}{p_i(1-\nu)} + p_i - p'_i$, $\Pr[V_i = 1 | T_i = B] = p_i - p'_i$;

3. box i has F-opening cost $c_i^F = \frac{3}{5}p_i$ and P-opening cost $c_i^P = (1 - \nu)p_i \left(\frac{1}{3} \left(\frac{p'_i}{p_i(1-\nu)} + p_i - p'_i \right) - \frac{3}{5}p_i \right)$.

The reduction. Given an instance of the ε -Subset Sum Problem (ε -SSP) $\{\{a_i\}_{i=1}^n, T\}$, we construct an associated 2T2L-PSI instance by setting the parameters $\alpha = \frac{\ln(\frac{3}{2})}{L-T}$ and $\nu = \frac{1}{2}$.

We will show that there exists a universal constant $c \in (0, 1)$ such that, whenever $\varepsilon < c$, there exists an optimal policy of the associated PSI instance that admits the form of a committing policy (see Section 4.3 for a formal definition and discussion). Specifically, in this associated PSI instance, the good type and the bad type are highly distinguishable. Learning the type of a box directly informs the decision to either F-open it or discard it. If a box is discarded, it has no further impact on any future decisions. Additionally, as the rewards are binary, it is optimal to stop and select a box immediately upon finding one with a reward of 1. Consequently, there exists an optimal policy for this instance that is a committing policy. If a PSI instance admits the above-mentioned properties, we call it a *well-separated* instance. Suppose that there is an oracle which returns an optimal policy for such a well-separated instance. Then one can read off a partition of boxes, of which the policy is committed to the opening mode.

Moreover, the parameter $\alpha = \frac{\ln(\frac{3}{2})}{L-T}$ of the associated 2T2L-PSI instance is carefully set such that the optimal partition (F^*, P^*) is exactly achieved at the condition $\sum_{i \in P^*} a_i = T$, if such a subset P^* exists.

One caveat, however, is that in the associated PSI instance, the inputs p_i and p'_i are irrational, potentially requiring exponential bit complexity. We note that for the purpose of a polynomial-sized reduction,³ it might be infeasible to feed the *true* associated low-cost-low-return PSI instance to the Pandora oracle. Instead, we approximate the inputs and feed the Pandora oracle an *approximated version*, using \hat{p}_i and \hat{p}'_i as substitutes for p_i and p'_i , respectively.

We use

$$\delta \stackrel{\text{def}}{=} \max_{i \in [n]} \max \left\{ \frac{|p_i - \hat{p}_i|}{p_i}, \frac{|p'_i - \hat{p}'_i|}{p'_i} \right\}$$

to measure the encoding precision. In the following lemma, we make it precise the conditions needed for the *true* associated 2T2L-PSI instance to be well-separated.

Moreover, we characterize the degree of precision needed in the approximated instance to ensure the same optimal policy structure, in terms of a requirement on δ .

LEMMA EC.11. *Consider a 2T2L-PSI instance, with parameter (ν, α) set to $\alpha = \frac{\ln(\frac{3}{2})}{L-T}$ where $0 < T \leq \frac{1}{2}L$. There exists a universal constant $c \in (0, 1)$ such that if $\frac{a_i}{L} \leq c$ for every i , then an optimal policy F-opens a good box immediately upon observation and discards a bad box (never opens it*

³ The reduction must run in polynomial time in $n \log(\max_{i \in [n]} a_i)$, which is the input size of the ε -SSP instance.

afterwards) upon discovery. Consequently, there exists an optimal policy which is a committing policy.

Moreover, $\delta = \Theta(\frac{1}{L})$ ensures that the approximated PSI instance admits an optimal committing policy as well.

In addition, an optimal policy must be a committing policy in both cases.

Proof: **First half: the true instance.**

We will show that by construction, there exists a universal constant $c \in (0, 1)$ such that if $\frac{a_i}{L} \leq c$ for every $i \in [n]$, then (1) $\sigma_i^F = \frac{2}{5}$, (2) $\sigma_i^{F|G} > \frac{2}{3}$, (3) $\sigma_i^{F|B} < 0$, (4) $\sigma_i^P = \frac{2}{3}$, (5) $\sigma_i^{F/P} > 0$. Therefore, by virtue of Theorem 2, an optimal policy shall immediately F-open a good box upon finding it. By Theorem 1, an optimal policy shall discard any bad boxes. Not only will one never open such a bad box, but also it does not affect any of the subsequent decisions. Consequently, there exists an optimal policy that is a committing policy. It is worth noting that our constructed instance does *not* satisfy the conditions of Corollary 1, where we have identified the optimal committing policy.

In what follows, we compute each of the above-mentioned thresholds. Observe that in order to ensure that $p_i < \theta$ for some $\theta \in (0, 1)$, it suffices to require

$$\frac{a_i}{L} < \frac{\ln(\frac{3}{2})}{2 \ln(\frac{1}{1-\theta})}. \quad (\text{EC.43})$$

1. **The F-threshold** $\sigma_i^F = \frac{2}{5}$. We recall that $c_i^F = \frac{3}{5}p_i$, $\Pr[V_i = 1] = p_i$ and

$$\mathbb{E} \left[\left(V_i - \sigma_i^F \right)^+ \right] = \Pr[V_i = 1] \left(1 - \sigma_i^F \right).$$

Hence, $\sigma_i^F = \frac{2}{5}$ exactly.

2. **The conditional F-threshold** $\sigma_i^{F|G} > \frac{2}{3}$. To see this, we note that by definition

$$c_i^F = \mathbb{E} \left[\left(V_i - \sigma_i^{F|G} \right)^+ \mid T_i = G \right] = \Pr[V_i = 1 \mid T_i = G] \left(1 - \sigma_i^{F|G} \right).$$

Plugging $\Pr[V_i = 1 \mid T_i = G] = \frac{p'_i}{p_i(1-\nu)} + p_i - p'_i$ into the above equation, we have that

$$\sigma_i^{F|G} = 1 - \frac{3}{5} \frac{p_i^2}{\frac{1}{1-\nu} p'_i + p_i(p_i - p'_i)}. \quad (\text{EC.44})$$

In order for $\sigma_i^{F|G} > \frac{2}{3}$, it suffices to set $\frac{p_i^2}{\frac{1}{1-\nu} p'_i + p_i(p_i - p'_i)} < \frac{5}{9}$. By Jensen's inequality, we know $\exp\left(-\frac{1}{2}\alpha a_i\right) < \frac{1}{2}(\exp(-\alpha a_i) + 1)$ which implies that

$$p_i > p'_i > \frac{1}{2}p_i. \quad (\text{EC.45})$$

Hence, $\frac{p_i^2}{\frac{1}{1-\nu}p'_i + p_i(p_i - p'_i)} < \frac{p_i^2}{\frac{1}{1-\nu}\frac{1}{2}p_i + p_i(p_i - p'_i)} < 2p_i$. It suffices to impose $2p_i < \frac{5}{9}$, for which to happen we can require

$$\frac{a_i}{L} \leq \frac{1}{2} \frac{\ln(18/13)}{\ln(3/2)} = O(1). \quad (\text{EC.46})$$

3. **The conditional F-threshold** $\sigma_i^{F|B} < 0$. By straightforward calculation, we see that $\sigma_i^{F|B} = 1 - \frac{c_i^F}{\Pr[V_i=1|T_i=B]} = 1 - \frac{\frac{3}{5}p_i}{p_i - p'_i} < 1 - \frac{\frac{3}{5}p_i}{p_i - \frac{1}{2}p_i} = -\frac{1}{5}$.

4. **The P-threshold** $\sigma_i^P = \frac{2}{3}$. Recall the definition

$$c_i^P = \Pr[T_i = G] \underbrace{\left(-c_i^F + \Pr[V_i = 1 | T_i = G] (1 - \sigma_i^P) \right)}_{(a)} + \Pr[T_i = B] \underbrace{\left(-c_i^F + \Pr[V_i = 1 | T_i = B] (1 - \sigma_i^P) \right)}_{(b)}.$$

If (a) > 0 and (b) < 0, it is easy to verify that $\sigma_i^P = \frac{2}{3}$ is indeed implied from the above equation. In what follows, we show that term (a) > 0 and (b) < 0 when setting $\sigma_i^P = \frac{2}{3}$. To see this, for (b), we observe that $-c_i^F + \Pr[V_i = 1 | T_i = B] \left(1 - \frac{2}{3} \right) = -\frac{3}{5}p_i + \frac{1}{3}(p_i - p'_i) < 0$. Next, for term (a): we have

$$-c_i^F + \Pr[V_i = 1 | T_i = G] \left(1 - \frac{2}{3} \right) = -\frac{3}{5}p_i + \frac{1}{3} \left(\frac{p'_i}{p_i(1-\nu)} + p_i - p'_i \right) = \frac{1}{3} \frac{p'_i}{p_i(1-\nu)} - \frac{4}{15}p'_i - \frac{1}{3}p_i > \frac{1}{3} \frac{1}{2} - \frac{4}{15}p'_i - \frac{1}{3}p_i,$$

as $p'_i > \frac{1}{2}p_i$ and $\frac{1}{1-\nu} > 1$. For the right-hand-side to be positive, it suffices to have

$$\frac{4}{15}p_i + \frac{1}{3}p'_i < \frac{4}{15}p_i + \frac{1}{3}p_i < \frac{1}{6} \quad (\text{EC.47})$$

which translates to $\frac{a_i}{L} = O(1)$, by Equation (EC.43).

5. **The F/P threshold** $\sigma_i^{F/P} > 0$. To prove $\sigma_i^{F/P} > 0$, in view of Theorem 1, it suffices to show that the expected payoff of F-opening is greater than the expected payoff of P-opening in the single-box case. The difference of the expected payoff of P-opening and the expected payoff of F-opening is

$$\begin{aligned} & -c_i^P + \Pr[T_i = G] \left(-c_i^F + \Pr[V_i = 1 | T_i = G] \cdot 1 \right) - \left(-c_i^F + \Pr[V_i = 1] \cdot 1 \right) \\ &= \frac{2}{3}p'_i + \frac{2}{3}(1-\nu)p_i(p_i - p'_i) - \frac{2}{5}p_i \\ &= \frac{2}{15}e^{-2\alpha a_i} \left(e^{\frac{1}{2}\alpha a_i} - 1 \right) \underbrace{\left(2e^{\frac{3}{2}\alpha a_i} + e^{\alpha a_i}(2-5\nu) - 5(1-\nu) \right)}_{(*)}. \end{aligned} \quad (\text{EC.48})$$

Hence, the sign is determined by the (*) term in the last parenthesis. By using $e^x \leq 1 + x + x^2$ for $\forall x \leq 1.79$, we know that whenever

$$\frac{a_i}{L} < \frac{1.79}{3} \frac{1}{\ln(\frac{3}{2})} \quad (\text{EC.49})$$

for every i , we have $(*) = 2e^{\frac{3}{2}\alpha a_i} + e^{\alpha a_i}(2 - 5\nu) - 5(1 - \nu) \leq -1 + 5(1 - \nu)\alpha a_i + (\frac{13}{2} - 5\nu)(\alpha a_i)^2$.

For the latter to be less than 0, it suffices to impose

$$\alpha a_i < \frac{1}{2} \min \left\{ \frac{1}{5(1 - \nu)}, \sqrt{\frac{1}{\frac{13}{2} - 5\nu}} \right\}, \quad (\text{EC.50})$$

which again translates to $\frac{a_i}{L} = O(1)$ for every i .

Second half: the approximated instance. In the approximated instance, we use the notation $\hat{\cdot}$ to denote the approximated version of the threshold, obtained by replacing p_i, p'_i with \hat{p}_i, \hat{p}'_i in the calculation. We proceed to show that by construction, for the same constant $c \in (0, 1)$ prescribed in the first half of the proof, as long as $\delta = \Theta(\frac{1}{L})$, we have (1) $\hat{\sigma}_i^F = \frac{2}{5}$, (2) $\hat{\sigma}_i^{F|G} > \frac{2}{3}$, (3) $\hat{\sigma}_i^{F|B} < 0$, (4) $\hat{\sigma}_i^P = \frac{2}{3}$, (5) $\hat{\sigma}_i^{F/P} > 0$. As a result, there exists an optimal policy that is a committing policy as well in the approximated instance.

Before calculating the thresholds exhaustively, we make some algebraic observations. We recall $\delta \stackrel{\text{def}}{=} \max_{i \in [n]} \max \left\{ \frac{|p_i - \hat{p}_i|}{p_i}, \frac{|p'_i - \hat{p}'_i|}{p'_i} \right\}$. A direct consequence is that

$$|p_i - \hat{p}_i| \leq p_i \delta \leq \delta. \quad (\text{EC.51})$$

In view of Equation EC.45, we have

$$\frac{\hat{p}'_i}{\hat{p}_i} > \frac{(1 - \delta)p'_i}{(1 + \delta)p_i} > \frac{1}{2} \frac{1 - \delta}{1 + \delta} \quad (\text{EC.52})$$

as well.

Moreover, we have $\hat{p}_i > \hat{p}'_i$ as long as $\delta = \Theta(\frac{1}{L})$. To see this, we observe $\hat{p}_i - \hat{p}'_i \geq p_i - p'_i - 2\delta$, and by the Mean Value Theorem, we have $p_i - p'_i = \exp\left(-\frac{1}{2}\alpha a_i\right) - \exp(-\alpha a_i) = \exp(\xi) \frac{1}{2}\alpha a_i$ for some ξ between $-\frac{1}{2}\alpha a_i$ and $-\alpha a_i$. Hence, $\exp(\xi) \frac{1}{2}\alpha a_i \geq \exp(-\alpha a_i) \frac{1}{2}\alpha a_i = \exp\left(-\ln\left(\frac{3}{2}\right) \frac{a_i}{L-T}\right) \frac{1}{2} \ln\left(\frac{3}{2}\right) \frac{a_i}{L-T} \geq \exp\left(-2 \ln\left(\frac{3}{2}\right)\right) \frac{1}{2} \ln\left(\frac{3}{2}\right) \frac{1}{L}$, as $\frac{a_i}{L-T} \leq \frac{2a_i}{L} \leq 2$. Therefore, in order for

$$p_i - p'_i = \exp(\xi) \frac{1}{2}\alpha a_i \geq \exp\left(-2 \ln\left(\frac{3}{2}\right)\right) \frac{1}{2} \ln\left(\frac{3}{2}\right) \frac{1}{L} > 2\delta,$$

it suffices to require

$$\delta = \exp\left(-2 \ln\left(\frac{3}{2}\right)\right) \frac{1}{4} \ln\left(\frac{3}{2}\right) \frac{1}{L} = \Theta\left(\frac{1}{L}\right).$$

Now, we turn to the computation of thresholds.

1. **The F-threshold** $\hat{\sigma}_i^F = \frac{2}{5}$. Recall the definition $c_i^F = \mathbb{E}[(V_i - \sigma_i^F)^+] = \Pr[V_i = 1](1 - \sigma_i^F)$. In the approximated version, $\hat{\sigma}_i^F$ is the solution to $\frac{3}{5}\hat{p}_i = \hat{p}_i \cdot (1 - \hat{\sigma}_i^F)$, which can be calculated exactly.

2. **The F-threshold** $\hat{\sigma}_i^{F|G} > \frac{2}{3}$. Following the same reasoning as Equation EC.44, we have

$$\hat{\sigma}_i^F = 1 - \frac{3}{5} \frac{\hat{p}_i^2}{\frac{1}{1-\nu}\hat{p}'_i + \hat{p}_i(\hat{p}_i - \hat{p}'_i)}.$$

Since we ensure $\hat{p}_i > \hat{p}'_i$ and Inequality EC.52, we can upper bound

$$\frac{\hat{p}_i^2}{\frac{1}{1-\nu}\hat{p}'_i + \hat{p}_i(\hat{p}_i - \hat{p}'_i)} < \frac{\hat{p}_i^2}{\frac{1}{1-\nu}\hat{p}'_i} < (1-\nu)2\frac{1+\delta}{1-\delta}\hat{p}_i \leq 2\frac{1+\delta}{1-\delta}(p_i + \delta).$$

Moreover, $\frac{1+\delta}{1-\delta}(p_i + \delta) < 2(p_i + \delta)$ if $\delta < \frac{1}{3}$. Hence, it suffices to impose $p_i < \frac{5}{36} - 0.0001$ and $\delta < 0.0001$ to make sure that $\hat{\sigma}_i^{F|G} > \frac{2}{3}$.

3. **The F-threshold** $\hat{\sigma}_i^{F|B} < 0$. For the approximated threshold $\hat{\sigma}_i^{F|B} = 1 - \frac{\frac{3}{5}\hat{p}_i}{\hat{p}_i - \hat{p}'_i} < 1 - \frac{\frac{3}{5}\hat{p}_i}{\hat{p}_i - \frac{1}{2}\frac{1-\delta}{1+\delta}\hat{p}_i} = 1 - \frac{3}{5}\frac{2+2\delta}{1+\delta}$. It is easy to verify that $\hat{\sigma}_i^{F|B} < 0$ as long as $\delta < \frac{1}{9}$.

4. **The P-threshold** $\hat{\sigma}_i^P = \frac{2}{3}$. Recall the definition

$$c_i^P = \Pr[T_i = G] \left(\underbrace{-c_i^F + \Pr[V_i = 1 | T_i = G] (1 - \sigma_i^P)}_{(a)} \right)^+ + \Pr[T_i = B] \left(\underbrace{-c_i^F + \Pr[V_i = 1 | T_i = B] (1 - \sigma_i^P)}_{(b)} \right)^+.$$

In the approximated instance, when setting $\sigma_i^P = \frac{2}{3}$, it is easy to see $(b) = -\frac{3}{5}\hat{p}_i + \frac{1}{3}(\hat{p}_i - \hat{p}'_i) < 0$ as long as $\delta = O(1)$. Likewise, for term (a), we observe that $\frac{1}{3}\frac{\hat{p}'_i}{\hat{p}_i(1-\nu)} - \frac{4}{15}\hat{p}'_i - \frac{1}{3}\hat{p}_i > \frac{1}{3}\frac{1}{2}\frac{1-\delta}{1+\delta}\frac{1}{1-\nu} - \frac{4}{15}\hat{p}'_i - \frac{1}{3}\hat{p}_i > \frac{1}{3}\frac{1}{2}\frac{1-\delta}{1+\delta} - \frac{4}{15}(p'_i + \delta) - \frac{1}{3}(p_i + \delta)$. One can verify that the latter is strictly positive as long as $\delta = O(1)$.

5. **The F/P-threshold** $\hat{\sigma}_i^{F/P} > 0$. In the approximated version, the difference of the expected payoff of P-opening and the expected payoff of F-opening is

$$\begin{aligned} & -c_i^P + \Pr[T_i = G] \left(-c_i^F + \Pr[V_i = 1 | T_i = G] \cdot 1 \right) - \left(-c_i^F + \Pr[V_i = 1] \cdot 1 \right) \\ &= \frac{2}{3}\hat{p}'_i + \frac{2}{3}(1-\nu)\hat{p}_i(\hat{p}_i - \hat{p}'_i) - \frac{2}{5}\hat{p}_i \\ &\leq \frac{2}{3}(1+\delta)p'_i + \frac{2}{3}(1-\nu)(1+\delta)p_i((1+\delta)p_i - (1-\delta)p'_i) - \frac{2}{5}(1-\delta)p_i \\ &\stackrel{\text{def}}{=} E(\delta, \nu, \alpha). \end{aligned} \tag{EC.53}$$

Recall the fact that $\frac{x}{1+x} < 1 - e^{-x} < \frac{x+x^2}{1+x+x^2}$ for $-1 < x < 1.79$. Moreover, $\frac{x+x^2}{1+x+x^2} = x - x^3 + O(x^4)$ and $\frac{x}{1+x} = x + O(x^2)$. We hence have

$$\frac{\alpha a_i}{1 + \alpha a_i} < p_i < \frac{\alpha a_i + \alpha^2 a_i^2}{1 + \alpha a_i + \alpha^2 a_i^2} \text{ and } \frac{\frac{1}{2}\alpha a_i}{1 + \frac{1}{2}\alpha a_i} < p'_i < \frac{\frac{1}{2}\alpha a_i + \frac{1}{4}\alpha^2 a_i^2}{1 + \alpha a_i + \frac{1}{4}\alpha^2 a_i^2}$$

as long as $\alpha a_i < 1.79$, which translates to the condition $\frac{a_i}{L} < \frac{1.79}{\ln(3/2)} = O(1)$. Therefore,

$$\begin{aligned}
& E(\delta, \nu, \alpha) \\
& < \frac{2}{3}(1+\delta)p'_i + \frac{2}{3}(1-\nu)(1+\delta)p_i((1+\delta)p_i - (1-\delta)p'_i) - \frac{2}{5}(1-\delta)p_i \\
& < \frac{2}{3}(1+\delta) \frac{\frac{1}{2}\alpha a_i + \frac{1}{4}\alpha^2 a_i^2}{1 + \alpha a_i + \frac{1}{4}\alpha^2 a_i^2} \\
& \quad + \frac{2}{3}(1-\nu)(1+\delta) \frac{\alpha a_i + \alpha^2 a_i^2}{1 + \alpha a_i + \alpha^2 a_i^2} \left((1+\delta) \frac{\alpha a_i + \alpha^2 a_i^2}{1 + \alpha a_i + \alpha^2 a_i^2} - (1-\delta) \frac{\frac{1}{2}\alpha a_i}{1 + \frac{1}{2}\alpha a_i} \right) \\
& \quad - \frac{2}{5}(1-\delta) \frac{\alpha a_i}{1 + \alpha a_i} \\
& = \left(-\frac{1}{15} + \frac{11}{15}\delta \right) \alpha a_i + O(\alpha^2 a_i^2).
\end{aligned}$$

In order for the latter to be negative, it suffices to impose $\delta < \frac{1}{11}$ and $\alpha a_i = O(1)$.

An optimal policy must be a committing policy. We proceed by induction on the number of closed boxes.

- Base case: For $n = 1$, the claim is immediate.
- Assume the statement holds for all well-separated instances of $n \geq 1$ closed boxes. Consider an instance with $n + 1$ closed boxes and let π^* be an optimal policy. π^* must either F-open or P-open a closed box:

— *F-open*: If π^* finds a reward, it stops immediately. Otherwise, we transition to a state with n boxes, where by the inductive hypothesis, a committing policy is optimal.

— *P-open*: If the box is good, it would be suboptimal not to F-open it immediately. If it is bad, the box is discarded and does not affect subsequent decisions. Either way, we again reach a state with n boxes, which by the inductive hypothesis admits a committing policy.

In both cases, the policy transitions to a scenario where a committing policy is optimal, so π^* itself must be committing.

□

The next Lemma shows the parameter α is carefully set so that we can use a Pandora oracle to solve a ε -SSP instance.

LEMMA EC.12. *There exists a universal constant $c \in (0, 1)$, such that given any instance of ε -SSP where $\varepsilon < c$, the answer to it is YES if and only if the optimal policy for the associated well-separated PSI P-opens a subset W^* that satisfies $T = \sum_{i \in W^*} a_i$.*

Proof: The “if” part is obvious, as such a subset W^* is an solution to the SSP.

For the “only if” part, we first note that by construction of the PSI instance and in light of Lemma EC.11, finding an optimal policy for this PSI instance reduces to identifying an optimal committing policy. Recall $K_i = \min \{V_i, \sigma^F\}$, $\tilde{K}_i = \min \{V_i, \sigma^P, \sigma_i^{F|T_i}\}$. For a committing policy (F, P) , we define $x(P) \stackrel{\text{def}}{=} \frac{\sum_{j \in P} a_j}{\sum_{i=1}^n a_i}$. Recall $\alpha = \frac{\ln(\frac{3}{2})}{L-T}$. Lemma 2 shows that the expected payoff of a committing policy (F, P) is exactly

$$\begin{aligned}
 \Psi(x(P)) &= \mathbb{E} \left[\max \left\{ \max_{i \in F} \{K_i\}, \max_{j \in P} \{\tilde{K}_j\} \right\} \right] \\
 &= \sigma^P \cdot \Pr \left[\max_{j \in P} \{\tilde{K}_j\} = \sigma^P \right] \\
 &\quad + \sigma^F \cdot \Pr \left[\max_{i \in F} \{K_i\} = \sigma^F \right] \left(1 - \Pr \left[\max_{j \in P} \{\tilde{K}_j\} = \sigma^P \right] \right) \\
 &\quad + 0 \cdot \Pr \left[\max_{i \in F} \{K_i\} = 0 \right] \Pr \left[\max_{j \in P} \{\tilde{K}_j\} \leq 0 \right] \\
 &= \sigma^P \cdot \left(1 - \prod_{j \in P} (1 - p'_j) \right) + \sigma^F \cdot \left(1 - \prod_{i \in F} (1 - p_i) \right) \left(\prod_{j \in P} (1 - p'_j) \right) \\
 &= \frac{2}{3} \left(1 - \exp \left(-\alpha \sum_{i \in P} \frac{1}{2} a_i \right) \right) + \frac{2}{5} \left(1 - \exp \left(-\alpha \sum_{i \in F} a_i \right) \right) \exp \left(-\alpha \sum_{i \in P} \frac{1}{2} a_i \right) \\
 &= \frac{2}{3} \left(1 - \exp \left(-\frac{1}{2} \frac{\ln(\frac{3}{2})}{1 - \frac{T}{\sum_{i=1}^n a_i}} x \right) \right) + \frac{2}{5} \left(1 - \exp \left(-\frac{\ln(\frac{3}{2})}{1 - \frac{T}{\sum_{i=1}^n a_i}} (1-x) \right) \right) \exp \left(-\frac{1}{2} \frac{\ln(\frac{3}{2})}{1 - \frac{T}{\sum_{i=1}^n a_i}} x \right). \tag{EC.54}
 \end{aligned}$$

When viewing Ψ as a function defined on $[0, 1]$, its derivative is

$$\frac{d\Psi}{dx}(x) = \frac{2}{15} \left(5 - e^{-\frac{Lx \ln(3)}{2L-2T}} \left(3e^{\frac{L(x-1) \ln(3)}{L-T}} + 2 \right) \right).$$

It is easy to verify that this function is unimodal on $[0, 1]$, and hence the first-order-condition $\frac{d\Psi}{dx} = 0$ leads to the optimal solution $x^* = \frac{T}{L}$.

Suppose the answer to the SSP instance is YES. Namely, there exists a partition W^* such that $\sum_{i \in W^*} a_i = T$. We can use the PSI oracle to solve the associated PSI instance. As there is an optimal policy that is a committing policy, we can read off the partition (F^*, P^*) . From the discussion above, we must have $\sum_{i \in P^*} a_i = T = \sum_{i \in W^*} a_i$, as desired. \square

Polynomial-sized encoding. At this point, the remaining challenge is to prove that the computational effort required to compute the approximated associated PSI instance, such that it retains the same optimal committing policy as the true associated instance, is still polynomial in $n \log(\max_{i \in [n]} a_i)$, which is number of bits required to encode the original SSP instance.

Given $P \subseteq [n]$, the true expected payoff of a committing policy (F, P) is

$$\Psi(x(P)) = \mathbb{E} \left[\max \left\{ \max_{i \in F} \{K_i\}, \max_{j \in P} \{\tilde{K}_j\} \right\} \right] = \sigma^P \cdot \left(1 - \prod_{j \in P} (1 - p'_j) \right) + \sigma^F \cdot \left(1 - \prod_{i \in F} (1 - p_i) \right) \left(\prod_{j \in P} (1 - p'_j) \right).$$

In the reduction, instead of computing exactly p_i, p'_i 's, we can only afford to compute the approximated inputs \hat{p}_i, \hat{p}'_i 's and feed them to the PSI oracle. Hence, the approximated expected payoff function is

$$\widehat{\Psi}(P) = \sigma^P \cdot \left(1 - \prod_{j \in P} (1 - \hat{p}'_j) \right) + \sigma^F \cdot \left(1 - \prod_{i \in F} (1 - \hat{p}_i) \right) \left(\prod_{j \in P} (1 - \hat{p}'_j) \right). \quad (\text{EC.55})$$

Hence, we are able to characterize the approximation error $|\Psi(x(P)) - \widehat{\Psi}(P)|$ in terms of the encoding precision δ .

CLAIM EC.2. *For all $P \subseteq [n]$, $|\Psi(x(P)) - \widehat{\Psi}(P)| \leq (\sigma^P + \sigma^F) \cdot (n\delta)$.*

Proof: Given fixed set $P \in 2^{[n]}$, consider the following function $f : [0, 1]^n \rightarrow \mathbb{R}$

$$f(\mathbf{x}) = \sigma^P \cdot \left(1 - \prod_{j \in P} (1 - x_j) \right) + \sigma^F \cdot \left(1 - \prod_{j \in [n] \setminus P} (1 - x_j) \right) \left(\prod_{j \in P} (1 - x_j) \right),$$

where \mathbf{x} denotes a n -dimensional vector and x_j is the j -th element of it. Its partial derivative with respect the i th element is

$$\begin{aligned} \frac{\partial f}{\partial x_i} &= \mathbf{I}[i \in P] \left[\sigma^P \cdot \prod_{j \in P \setminus \{i\}} (1 - x_j) + (-\sigma^F) \cdot \left(1 - \prod_{j \in [n] \setminus P} (1 - x_j) \right) \left(\prod_{j \in P \setminus \{i\}} (1 - x_j) \right) \right] \\ &\quad + \mathbf{I}[i \in [n] \setminus P] \left[\sigma^F \cdot \left(\prod_{j \in P} (1 - x_j) \right) \left(\prod_{j \in [n] \setminus (P \cup \{i\})} (1 - x_j) \right) \right]. \end{aligned}$$

It is easy to see that $\left| \frac{\partial f}{\partial x_i}(\mathbf{x}) \right| \leq \sigma^P + \sigma^F$ for $\mathbf{x} \in [0, 1]^n$. When we calculate \hat{p}_i, \hat{p}'_i 's, we can easily keep them in the range of $[0, 1]$. Therefore, applying Mean Value Theorem yields that

$$\left| \Psi(x(P)) - \widehat{\Psi}(P) \right| \leq (\sigma^P + \sigma^F) \cdot \left| \sum_{j \in P} (p'_j - \hat{p}'_j) + \sum_{j \in [n] \setminus P} (p_j - \hat{p}_j) \right| \leq (\sigma^P + \sigma^F) (n\delta).$$

We note that the above argument works for any fixed P . The proof is hence complete. \square

Moreover, the function Ψ has enough curvature at the point $\frac{T}{L}$.

CLAIM EC.3. *For any $0 \leq T \leq \frac{L}{2}$, where $L \geq 2$, $\Psi(\frac{T}{L}) - \Psi(\frac{T}{L} + \frac{1}{L}) \geq (\frac{\ln(3)}{2})^2 3^{-3} \frac{1}{L^2}$ and $\Psi(\frac{T}{L}) - \Psi(\frac{T}{L} - \frac{1}{L}) \geq (\frac{\ln(3)}{2})^2 3^{-3} \frac{1}{L^2}$.*

Proof: We only prove the first inequality. The second can be handled similarly. When viewing Ψ as a function defined on $[0, 1]$,

$$\Psi(x) = \frac{2}{3} \left(1 - \exp \left(-\frac{1}{2} \ln(3) \frac{L}{L-T} x \right) \right) + \frac{1}{2} \left(1 - \exp \left(-\ln(3) \frac{L}{L-T} (1-x) \right) \right) \exp \left(-\frac{1}{2} \ln(3) \frac{L}{L-T} x \right).$$

As it is smooth enough, by Taylor's expansion, we have

$$\Psi \left(\frac{T}{L} + \frac{1}{L} \right) = \Psi \left(\frac{T}{L} \right) + \underbrace{\Psi' \left(\frac{T}{L} \right)}_{=0} \frac{1}{L} + \frac{1}{2} \Psi''(\xi) \left(\frac{1}{L} \right)^2,$$

for some ξ on the line segment of $\frac{T \pm 1}{L}$ and $\frac{T}{L}$. In what follows, we proceed to lower bound $|\Psi''(\xi)|$.

To straightforward calculation, we have

$$\Psi''(x) = -\frac{\ln(3)^2}{8} \cdot \left(\frac{L}{L-T} \right)^2 \cdot 3^{\frac{L(x-4)+2T}{2(L-T)}} \cdot \left(3^{\frac{L-Lx}{L-T}} + 3 \right).$$

By noting that for any $x \in [0, 1]$ and $0 \leq T \leq \frac{L}{2}$:

- $3^{\frac{L-Lx}{L-T}} \geq 3^0 = 1$,
- $3^{\frac{L(x-4)+2T}{2(L-T)}} \geq 3^{\frac{-4L+2T}{2(L-T)}} = 3^{\frac{-2L+T}{L-T}} \geq 3^{-3}$,
- $\frac{L}{L-T} \geq 1$,

we conclude that $|\Psi''(x)| \geq \left(\frac{L}{L-T} \right)^2 \frac{4\ln(3)^2}{8} 3^{-3} \geq \frac{\ln(3)^2}{2} 3^{-3}$ for any $x \in [0, 1]$ and $0 \leq T \leq \frac{L}{2}$. □

Combining Lemma EC.2 and Lemma EC.3, we have a chain of inequalities

$$\hat{\Psi} \left(\frac{T}{L} \right) \geq \Psi \left(\frac{T}{L} \right) - O(n\delta) \geq \Psi \left(\frac{T \pm 1}{L} \right) + \Omega \left(\frac{1}{L^2} \right) - O(n\delta) \geq \hat{\Psi} \left(\frac{T \pm 1}{L} \right) + \Omega \left(\frac{1}{L^2} \right) - O(n\delta).$$

Therefore, it suffices to make δ small enough so that $\Omega \left(\frac{1}{L^2} \right) \geq O(n\delta)$ to make $\hat{\Psi} \left(\frac{T}{L} \right) \geq \hat{\Psi} \left(\frac{T \pm 1}{L} \right)$. In addition, we note that it is easy to show $\hat{\Psi}$ is unimodal as a function defined on $[0, 1]$. Hence, when viewed as functions on subsets, both $\hat{\Psi}$ and Ψ will attain the global maximum at the same subset.

Lastly, we conclude that the polynomial-sized encoding is feasible, by virtue of the following classic result (stated in our context).

LEMMA EC.13 (Brent (1976)). *Assuming access to multitape Turing machine, let $M(d)$ be the number of bit-operations required to multiply d -bit numbers with d -bit accuracy. It takes $O(\log(d)M(d))$ bit operations to ensure that $\max \left\{ \frac{|p_i - \hat{p}_i|}{p_i}, \frac{|p'_i - \hat{p}'_i|}{p'_i} \right\} \leq 2^{-d}$ for each i , for the parameter $\alpha = \frac{\ln(\frac{3}{2})}{L-T}$.*

Appendix D: Additional Tables

Table EC.1 Runtime (in seconds) of different policies.

N	π^{OPT}		π^{index}		π^{W}		π^{F^*, P^*}		π^{STP}	
	mean	std	mean	std	mean	std	mean	std	mean	std
2	0.002	0.002	0.001	0.001	0.005	0.004	0.001	0.001	0.000	0.000
3	0.014	0.013	0.002	0.002	0.021	0.024	0.004	0.004	0.001	0.001
4	0.101	0.078	0.004	0.005	0.084	0.121	0.013	0.014	0.002	0.002
5	0.738	0.577	0.007	0.010	0.276	0.466	0.032	0.049	0.004	0.003
6	5.698	5.128	0.012	0.017	0.884	1.729	0.081	0.144	0.007	0.008
7	41.049	39.646	0.025	0.050	2.933	6.081	0.248	0.563	0.015	0.020
8	316.648	311.921	0.047	0.114	10.831	30.604	0.707	1.902	0.028	0.031
9	2472.509	2585.495	0.088	0.323	29.127	72.949	1.662	3.831	0.052	0.060
10	-	-	0.128	0.316	75.226	228.431	3.292	7.010	0.072	0.104
11	-	-	0.261	1.018	152.678	1817.640	9.082	25.463	0.141	0.220
12	-	-	0.372	0.778	718.722	3458.725	20.250	40.365	0.227	0.382
13	-	-	0.823	1.851	1613.774	4178.379	79.878	266.576	0.438	0.594
14	-	-	1.556	6.139	3668.967	10860.112	245.827	1494.645	0.793	1.365
15	-	-	2.977	3.895	10325.838	13620.728	810.565	1658.033	1.885	2.392
16	-	-	5.475	11.173	10168.432	15753.618	869.394	5183.688	2.025	3.310

Appendix E: When is P-opening worthwhile?

We conduct an additional numerical study with the goal of examining when P-opening is worthwhile.

We apply a similar process for generating instances (Section 5.1) with the exception choosing 100 as the number prototypical boxes, which are illustrated in Figure EC.8.

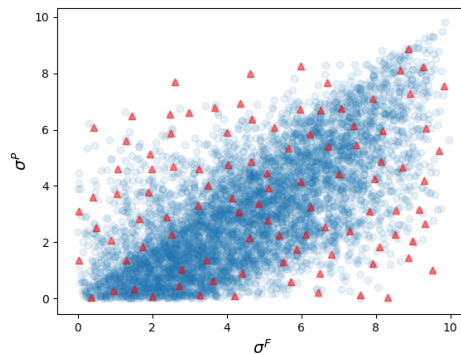


Figure EC.8 An example of how prototypical boxes are generated. Each point is a box sampled from the distribution. Red triangles are those which are selected, while blue circles are those which are not.

To get a sense for when P-opening is particularly useful under an optimal policy, we define the *P-ratio*. This is the ratio between the expected number of P-openings and the total expected number of P-openings and F-openings (of closed-boxes), both computed for an optimal policy at the initial state of a problem instance.

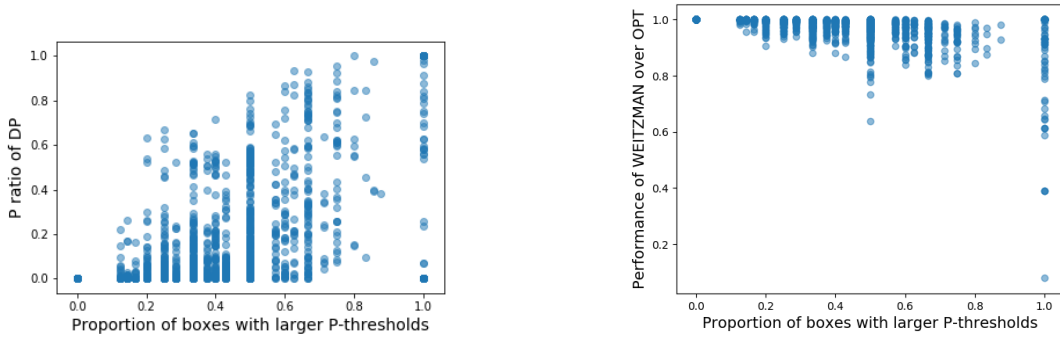


Figure EC.9 When is P-opening important?

Left figure: P-ratio of π^{OPT} versus proportion of boxes with $\sigma_i^P > \sigma_i^F$.

Right figure: Suboptimality (ratio of the value of π^{Weitzman} to that of the optimal policy)) of $\pi^{[N],\emptyset}$ versus proportion of boxes with $\sigma_i^P > \sigma_i^F$.

Figure EC.9 (left) plots the P-ratio of π^{OPT} versus the proportion of boxes with $\sigma_i^P > \sigma_i^F$. We observe a positive correlation between these two quantities, indicating that when there are more boxes with $\sigma_i^P > \sigma_i^F$, the optimal policy tends to take advantage of P-opening more frequently. We also calculate the performance of the policy π^{Weitzman} , which simply ignores all the P-opening options, i.e., the committing policy $\pi^{[N],\emptyset}$. Figure EC.9 (right) shows its suboptimality versus the proportion of boxes with $\sigma_i^P > \sigma_i^F$. We similarly observe that it tends to perform worse when there are more boxes with larger P-thresholds. These observations further suggest that the interaction between the P-thresholds and F-thresholds impacts optimal decisions, and the relative magnitude of P-thresholds compared to F-thresholds determine the extent to which P-opening should be utilized. This may also explain why the simple index policy π^{index} performs so well.

Next, we consider the heterogeneity of boxes. For each problem instance, we define *dispersion* as the sum of standard deviations of the P-thresholds and F-thresholds that are associated with closed boxes. Intuitively, when the dispersion is higher, boxes are more heterogeneous. The upper panels of Figure EC.10 illustrate dispersion. Each panel shows a problem instance, where every point on the scatter plot represents a box (e.g., the left panel shows an instance with 9 boxes, while the right panel shows a problem instance with 10 boxes). The coordinates of each box correspond to its F-opening and P-opening thresholds. Intuitively, the instance on the left has boxes that are more homogeneous while the instance on the right has boxes that are more heterogeneous, which are reflected by the dispersion metric (the dispersion of the left instance is smaller than 5, while the dispersion of the right instance is greater than 20).

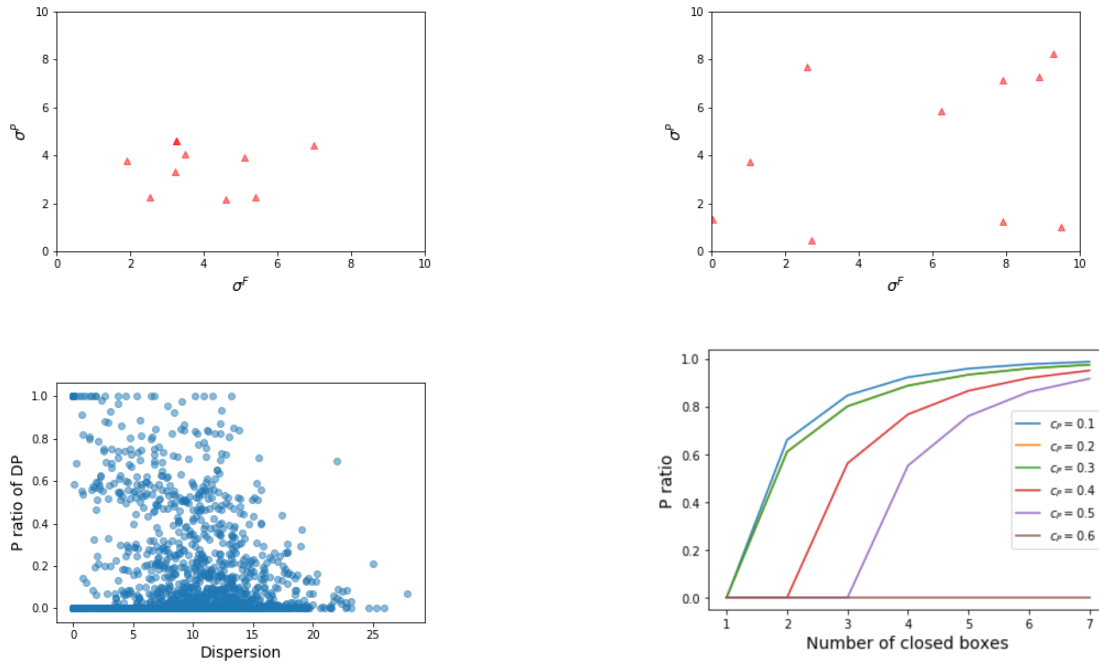


Figure EC.10 Heterogeneity of boxes. Figures from left to right, top to bottom correspond to: example instance where dispersion is relatively small (< 5), example instance where dispersion is relatively large (> 20), P-ratio of the optimal policy versus dispersion, and P-ratio of the optimal policy versus the number of boxes.

The lower-left panel of Figure EC.10 plots the P-ratio versus dispersion. We observe that, when the dispersion is large (i.e., above 15), the P-ratio tends to be small, meaning that there are relatively fewer P-openings in the optimal policy. Intuitively, when boxes are very different in terms of the potential information gain which is proxied by the opening thresholds, boxes are more likely to dominate each other. In such settings, an optimal policy tends to directly F-open the boxes according to a predefined order.

Finally, the lower-right panel of Figure EC.10 shows the P-ratio versus the number of boxes of the problem instance. To generate this figure, we randomly sample several box instances, vary their P-opening costs, and investigate how an optimal policy behaves when there are N i.i.d. such boxes. We observe that the optimal policy exploits P-opening more often when there are sufficiently many boxes. The panel complements panel (c), showing that when the dispersion is low (here instances are homogeneous), there is interaction affect with the number of boxes.