Dynamics of Multi-Stage Screening*

David Lagziel† and Ehud Lehrer‡

April 13, 2026

Abstract:

This study explores dynamic screening problems in which elements undergo noisy evaluations. Some elements are discarded at each stage, while the remainder are re-evaluated independently in subsequent stages. We demonstrate that, ceteris paribus, the quality of a screening process may not improve as the number of stages increases. Specifically, we examine the resulting elements' values and demonstrate that adding a single stage to a screening process may produce inferior results in terms of stochastic dominance, while increasing the number of stages substantially leads to a first-best outcome.

Journal of Economic Literature classification numbers: C70, D49, D81.

Keywords: dynamic screening; perfect screening; threshold strategies

---

*For their valuable comments, the authors wish to thank the participants of the One World Mathematical Game Theory Seminar, the Aix-Marseille School of Economics Interaction seminar, the Adam Smith Business School Microtheory Seminar, the Manchester Economic Theory Seminar, ORSIS 2019, the Tel Aviv University Game Theory and Mathematical Economics Research Seminar, the Technion Game Theory Seminar, and the Bar Ilan University Game Theory Seminar. This work was supported by the Israel Science Foundation, Grant #2074/23.

†Department of Economics, Ben-Gurion University of the Negev, Israel. E-mail: Davidlag@bgu.ac.il.

‡Department of Economics, Durham University, UK. E-mail: ehud.m.lehrer@durham.ac.uk.

1 Introduction

Imagine a bacterial infection that spreads through a population. Preventive treatment, such as pre-exposure prophylaxis, is available but the supply is limited. Therefore, the government seeks to identify and select high-risk individuals for this treatment. Fortunately, a simple test that is easy to administer across the population, provides a reasonable (though noisy) identification of individual risk. Accordingly, the existing policy grants treatment to individuals who score high on this test. Recently, a more accurate but costly test became available and was added as a second step to improve the screening. Specifically, the first test is used to identify a subset of high-risk individuals, who are then re-examined using the second test, and only those with the highest risk-score in the second test receive the treatment. The final number of people receiving treatment remains as before. After some experience with the new screening procedure, it appears that the two-stage process does not improve the identification. In fact, it produced worse results overall. Specifically, the two-stage procedure is more likely to misidentify low-risk individuals as high-risk individuals compared with the previously used one-stage screening. How is this possible?

This study answers this question by analyzing dynamic screening processes such as that described above. We focus on a decision maker (DM) who screens elements from a general set based on noisy evaluations. The screening process could vary from a single to multiple stages while maintaining an overall capacity constraint on accepted elements. More accurately, once a dynamic screening process is in place, a subset of the elements are rejected in every stage based on their single-stage score, whereas the remaining elements are independently re-evaluated in subsequent stages.[1]

We intend to identify the key advantages and disadvantages of dynamic screening within this framework, pursuing this goal through two complementary paths. The first path compares one- and two-stage screening. The second path follows an asymptotic approach to examine multistage screening in which the number of stages increases. The combined results of these two approaches provide insights into why people employ dynamic screening and the conditions under which additional screening stages result in suboptimal outcomes.

Specifically, the first research path introduces a basic concept that we refer to as *the hidden cost of dynamic screening.* At face value, additional screening stages impose a direct cost in time and effort. Intuitively, one would expect the added costs to be fully justified and compensated for by the anticipated superior nature of the obtained outcomes. However, we demonstrate that even when the accuracy of the additional stages is superior, they may actually generate inferior results in terms of stochastic dominance.

[1]For clarity, although we use the term "screening," this study focuses on a single DM rather than dynamic interactions between multiple strategic agents as in, e.g., Courty and Hao (2000) and Pavan et al. (2014).

The root of this phenomenon is the *self-induced slack* inherent to dynamic screening. Introducing more screening stages requires easing the acceptance criteria (i.e., setting a lower threshold) in earlier stages. Screening stages cannot simply be added without either admitting potentially subpar elements by relaxing initial criteria or overstepping the capacity constraint. Once these less-than-optimal elements are factored in, the noise in later stages can have a more detrimental effect, even if it provides more detailed information (e.g., as suggested by Lehmann (1988)) than earlier stages. Essentially, an inherent trade-off exists in which, as the sample size shrinks, the noise also diminishes, rendering the expected outcome unpredictable.

The dynamic screening process need not be dynamic in time. Specifically, similar results can be obtained by using several tests concurrently while following an unanimity rule. Although technically not dynamic in time, the unanimity rule assures that the dual screening is essentially a dynamic process. For example, in academia, students are required to pass all courses to be eligible to graduate and earn their degrees. This use of a "unanimity rule" for completion is, de facto, a dynamic screening process with distinct, independent stages.

What conditions lead to suboptimal outcomes? First, it is crucial to note that our results hold for general nonatomic distributions and additive noise. Second, as in previous studies (see Section 1.1 for a broader review), we focus on threshold strategies for a few obvious reasons. Threshold strategies are simple, commonly used, and are indeed optimal in the absence of noise. Third, as has already been proven,[2] a one-stage screening process based solely on a more-accurate second test yields better results than a one-stage screening given the first, basic test. Therefore, the fact that the majority of the screening is performed in less accurate stages has a significant influence on generating this phenomenon. In fact, the effect is completely reversed when shifting from "elite" screening (where only a minor fraction is accepted) to "low-level" screening. Essentially, the hidden cost of additional stages becomes pronounced when the overall acceptance constraint is minimal, rendering later stages more susceptible to pronounced noise.

This distinction brings us to the second part of our analysis, i.e., the asymptotic approach. We next consider an alternate example of screening job applicants. Because the stakes are high, many cutting-edge institutions conduct prolonged applicant screening processes with multiple stages. It seems quite implausible that all these institutions are under-performing. To be clear, we do not make this claim. Once multiple stages are introduced, then the overall constraint could be maintained by consistently screening in small portions. That is, every stage can support a high acceptance rate, effectively making it a low-level screening, whereas the final outcome matches the overall constraint.

In light of this understanding, we reach our second main result which establishes a *convergence to perfect screening*. We prove that a multistage asymptotic screening process yields a first-best posterior

[2] See the notion of a *contraction mapping* in Lagziel and Lehrer (2022).

distribution as if the screening were performed with no noise whatsoever, even if all stages are subject to the same noise. We refer to this outcome as *perfect screening*, demonstrating that standard stationary strategies such as a fixed-threshold or a fixed-capacity strategies, result in a perfect screening outcome. Note that we obtain this result even when the same noise is used repeatedly; that is, even if accuracy does not improve along the stages.

A combination of the results that originate from these two research paths leads to a (somewhat striking) conclusion that the quality of a screening process, as a function of the number of stages, is not necessarily monotone. In other words, introducing one additional referee or one additional stage may be detrimental, whereas adding multiple stages would likely improve the screening. In what follows, we provide the specific details and conditions to support this statement.

1.1 Relation to relevant literature

Accurately positioning this work in the existing literature is anything but trivial. By contrast, this study is associated with the extensive theory of statistical decision making that concerns dynamic decision problems [going back to the works of Wald (1939, 1947) and Arrow et al. (1949)] and information structure comparisons.[3] Conversely, this study also concerns the theory of information aggregation, which varies from social learning and information cascade to group decisions and committees of experts.[4]

Although these two branches of literature are extensive and thorough, our work does not precisely fit either branch. There are three basic elements that distinguish this study from previous research. First, our formulation of screening problems, in the context of statistical decision theory, naturally combines a capacity constraint that is typically missing from the previous research [with the exception of Lagziel and Lehrer (2019, 2022)]. This constraint is necessary for our analysis and outcomes. Second, many of the above-mentioned studies extended the seminal work of Condorcet (1785) by introducing costly observations and strategic information accumulation. This study is more basic in this sense because it raises the question of whether another signal is beneficial altogether, irrespective of its price, and independent of the evaluators' preferences, which renders information cascades and herding less relevant in our framework. Third, our framework builds on general, nonatomic distributions in which signal classification as either "true" or "false" is irrelevant. In our model, every signal provides more information about the actual value; however, an additional signal (given another screening stage) might still be detrimental. These stark differences are best exemplified by Bikhchandani et al. (2021), who states that "In purely individual decision making, an extra signal always makes an agent weakly

[3] See Milgrom (1981), Quah and Strulovici (2009), Ganuza and Penalva (2010), Chambers and Healy (2011), and more recently Athey and Levin (2018) and Lagziel and Lehrer (2022).

[4] See Bikhchandani et al. (2021) for a recent survey on social learning and information cascade.

better off" (Section 2.6 herein). Although this statement is completely true for relevant models of information aggregation and social learning, one of our main results proves the opposite.

Nevertheless, some basic similarities remain between our study and previous research. The first, rather basic, similarity is the fact that our asymptotic analysis yields the first-best screening outcome. This is also the key insight of Condorcet (1785), with the obvious distinctions from our model. Condorcet (1785) and subsequent works built on a binary state of the world, majority rule, and a necessary condition concerning signals' informativeness, which are irrelevant for our framework. Another similarity, which is primarily related to the research approach, arises from Ben-Yashar and Paroush (2000) and Berend and Sapir (2005), who proved that random committees of at least three experts outperform a single expert. However, the key differences between the models and assumptions completely change the outcome under our analysis.

The current study is also related to two recent articles of Lagziel and Lehrer (2019, 2022). Lagziel and Lehrer (2019) examines one-stage screening problems under threshold strategies, and proves that the expected value of accepted elements is not a monotone function of the capacity constraint. Specifically, it shows that the expected value of accepted elements may *decrease* when the screening threshold is raised. The current study complements this result by demonstrating that a "more restrictive" screening, either through additional stages or higher thresholds, may yield inferior outcomes overall.

Furthermore, Lagziel and Lehrer (2022) compares screening problems under different noise levels, providing a characterization in which one noise variable dominates another in terms of screening under threshold strategies. The study shows that a screening process can sometimes improve by adding unbiased independent noise. Our results complement those of Lagziel and Lehrer (2022) by showing that a "noisier" screening process may prove superior, through either fewer stages or additional noise.

However, we emphasize that the current study provides significantly more general and robust results compared with the two aforementioned works, particularly regarding screening strategies and noise conditions. For example, Theorem 3 in this study accommodates distinct noise distributions at different stages and permits varying threshold strategies across stages. These features were not feasible in the frameworks of Lagziel and Lehrer (2019) and Lagziel and Lehrer (2022).[5] Additionally, the current study explores the possibility of perfect screening, a topic that has not been addressed in previous research.

[5] For example, the screening biases in Lagziel and Lehrer (2019) rely on specific noise distributions, and may not hold under general distributions, as considered in the current study. Furthermore, the main analysis in Lagziel and Lehrer (2022) does not account for varying thresholds that are uniquely determined by the capacity constraints.

1.2 Structure of the paper

The remainder of this paper proceeds as follows. In Section 2 we describe the basic model and key definitions. In Section 3 we present the main results, which are divided into two subsections. In Subsection 3.1 we compare one- and two-stage screening processes, and in Subsection 3.2 we conduct an asymptotic analysis of screening problems. Concluding remarks, including a discussion about cost functions, are presented in Section 4.

2 Preliminaries

Consider a set of elements with intrinsic values that are distributed according to a non-constant random variable V, which is referred to as an *impact* variable. For every $i \in \mathbb{N}$, let N_i be a random variable which defines the additive evaluation errors in stage i, and is referred to as the *stage-i noise*. We generally assume that all noise variables are unbiased — symmetrically distributed around zero and jointly independent of V and of each other.[6] A *capacity* $p \in (0,1)$ dictates the proportion of accepted elements. That is, the screening is constrained by the requirement to accept a fraction p of the proposed elements.

A k-stage screening problem $\text{SP} = \big(V, \{N_i\}_{i=1}^k, p\big)$ consists of an impact variable V, noise variables $N_1, \ldots, N_k$, and a capacity p. The screening problem evolves as follows. Denote $V_1 = V$. In each stage $i \geqslant 1$, the DM observes $V_i + N_i$ and sets a threshold $t_i \in \mathbb{R}$, so that the distribution of every V_{i+1} is as follows:

$$V_{i+1} \sim V_i | \{V_i + N_i \geqslant t_i\}, \tag{1}$$

where

$$\Pr(V_i + N_i \geqslant t_i) = p_i \quad \text{and} \quad \prod_{i=1}^{k} p_i = p. \tag{2}$$

Specifically, in every stage i, the DM observes the noisy valuation $V_i + N_i$ and sets a screening threshold t_i, so that only elements with noisy valuations of at least t_i proceed to the next stage. Formally, $\Pr(V_{i+1} \leqslant t) = \Pr(V_i \leqslant t | V_i + N_i \geqslant t_i)$ such that $\Pr(V_i + N_i \geqslant t_i) = p_i$. At each stage i, the DM maintains a capacity of $p_i \in [0,1]$ to support an overall capacity of $p = \prod_{i=1}^k p_i$.

Given $\text{SP} = \big(V, \{N_i\}_{i=1}^k, p\big)$, a *strategy* τ is a sequence of threshold values $\tau = (t_1, \ldots, t_k) \in \mathbb{R}^k$.[7] Let $V_{\text{SP}}(\tau)$ denote the post-screening valuations of accepted elements. That is, $V_{\text{SP}}(\tau) = V_{k+1}$ where V_{k+1} is defined according to Eq. (1), Eq. (2), and τ. The main goal of the DM is to maximize the expected value $\mathbb{E}[V_{\text{SP}}(\cdot)]$ of accepted elements.

[6] We stress that the symmetry assumption is simply a matter of exposition, and all the statements and proofs in this study also hold for asymmetric distributions.

[7] Strategies are stationary by definition, because the threshold in every stage relates solely to the realized outcome in that stage, rather than to the entire history.

Note that we also allow for *infinite* screening problems, denoted by $\text{SP} = (V, \{N_i\}_{i=1}^{\infty}, p)$, by taking an unbounded number of stages. In this case, a strategy $\tau = (t_1, t_2, \dots)$ is an infinite sequence of threshold values, and $V_{\text{SP}}(\tau)$ is a random variable such that $\Pr(V_{\text{SP}}(\tau) \leqslant t) = \lim_{k\to\infty} \Pr(V_k \leqslant t)$ for every t (namely, convergence in distribution).

For tractability, we make the following assumptions: (i) every variable X (of the above) has a continuous density function f_X and a CDF F_X that are fully supported (namely, f_X is strictly positive) on some bounded interval, denoted as $[\underline{X}, \overline{X}]$, and (ii) unless stated otherwise, all noises are i.i.d. random variables. As will later become evident, these assumptions can be altered (specifically, by varying the noise across stages to be more informative in the sense of Lehmann (1988) or by using a contracting mapping as in Lagziel and Lehrer (2022); see Subsection 3.2 therein) and still maintain our key insights. Specifically, we relax the second assumption (of identically distributed noises) in the first part of our analysis when examining a two-stage screening with a more informative second stage, relative to the first stage.[8]

2.1 Fixed stationary strategies

Our analysis is based on threshold strategies, and in the context of dynamic screening, one can also consider two types of fixed strategies: fixed-threshold strategies and fixed-capacity strategies. Formally, fix a k-stage screening problem SP, and consider the following two fixed (stationary) strategies. The first strategy, referred to as the *fixed-threshold strategy*, dictates that $\tau = (t, \dots, t)$, which means that all threshold values, throughout the k stages, are identical. By continuity, a (unique) threshold value t can be fixed to maintain the capacity constraint p; thus, the strategy is well defined. The second strategy, referred to as the *fixed-capacity strategy*, dictates that $p_i = p^{1/k}$ for every stage $1 \leqslant i \leqslant k$. That is, all threshold values are fixed so that, in every stage, a fraction $p^{1/k}$ of the elements being evaluated proceed to the subsequent stage. We employ these strategies in both parts of our analysis.

The first part of our analysis also requires a precise definition concerning stochastic dominance of one screening process over another. The following definition captures this notion.

Definition 1. *Consider a k-stage screening problem* SP_k *with a specific strategy* τ_k*, where* $k = k_1, k_2$*. We assert that* k_1-stage screening under SP_{k_1} given τ_{k_1} stochastically dominates k_2-stage screening under SP_{k_2} given τ_{k_2} *if* $V_{\text{SP}_{k_1}}(\tau_{k_1})$ *first-order stochastically dominates* $V_{\text{SP}_{k_2}}(\tau_{k_2})$*.*

We also want to clarify that a one-stage screening problem requires no specification regarding the strategy, since the relevant threshold is well defined and unique. Therefore, we do not specify the type of the one-stage strategies going forward.

[8] Although the extension to correlated noises is beyond the scope of this study, note that adding stages with positively correlated noises would typically generate significant distortions to the screening procedure, supporting our key insights.

3 Main results

Our analysis consists of two parts. In Subsection 3.1, we compare one- and two-stage screening processes, and in Subsection 3.2 we apply an asymptotic approach by focusing on a dynamic screening while substantially increasing the number of stages.

3.1 A comparison of one-stage and two-stage screening

Our comparison of one- and two-stage screening processes yields four relevant results. First, in Theorem 1, we consider two-stage screening performed under a fixed-threshold strategy in which all noises are identically distributed. Under these conditions, we demonstrate that one-stage screening stochastically dominates the two-stage process for a sufficiently low acceptance rate (i.e., elite screening). Next, we extend this result in Theorem 2 and Theorem 3 by omitting the fixed-threshold condition and allowing the noise distributions to vary between the two stages. Finally, in Theorem 4, we show how these results can be reverted by adopting a sufficiently high capacity constraint, i.e., through low-level screening.

We begin with Theorem 1, which relates to i.i.d. noises along with an implementation of a fixed-threshold strategy in a two-stage process. In other words, Theorem 1 captures the alternative interpretation (for dynamic screening) of using two evaluators instead of one, when both are subjected to the same noise and use the same screening strategy. The theorem states that it is better to use one stage instead of two for a sufficiently low capacity with very restrictive screening. (All proofs are presented in the Appendix.)

Theorem 1. *For every impact variable V and i.i.d. noise variables N_1 and N_2, there exists $p_0 > 0$ such that for every capacity $p < p_0$, one-stage screening under (V, N_1, p) stochastically dominates two-stage screening under $(V, \{N_1, N_2\}, p)$ given a fixed-threshold strategy.*

Although all proofs are given in the Appendix, we wish to provide some intuition on its structure and technique to better explain the result. In the proof, we consider the probability density functions $((f_{V_1}, f_{V_2}))$ of one- and the two-stage screening processes, respectively. To preserve the capacity constraint, the two-stage process must follow a lower threshold value so that f_{V_2} is supported on a larger interval and some lower values are generated with positive probabilities, whereas similar values are eliminated under the one-stage screening. Therefore, to establish (first-order) stochastic dominance, we demonstrate that the graphs of the two densities intersect only once. Graph (a) in Figure 1 provides some intuition for this.

Note that the two-stage screening in Figure 1 is represented by a parabolic graph, rather than a straight line, due to the fact that the probability of passing two independent tests is the product of two probabilities; one for each stage. Specifically, the area above the parabolic line describes the probability of each value passing the two-stage screening process (and not just the first stage of the two).[9]

Note the critical role of the capacity in our analysis. Generally, the parabolic and linear graphs exhibit two intersections as the two-stage screening process relies on lower (fixed) threshold values. When the screening becomes stricter (i.e., a smaller capacity p), all thresholds increase, causing the two curves to shift to the right, as shown in Figure 1(a). This shift moves the left-hand side (LHS) intersection of the two curves beyond the upper bound of V, thereby preserving the single-crossing property across the entire support of V and ensuring the dominance of one-stage screening over the two-stage process. Conversely, a more lenient screening (i.e., a higher capacity p) reduces the threshold criteria and shifts both curves to the left, as shown in Figure 1(b). This adjustment allows the right-hand side (RHS) intersection of the two curves to appear while eliminating the LHS crossing from the support of V, resulting in the two-stage screening dominating the one-stage process. (This is the main result of Theorem 4 below.)

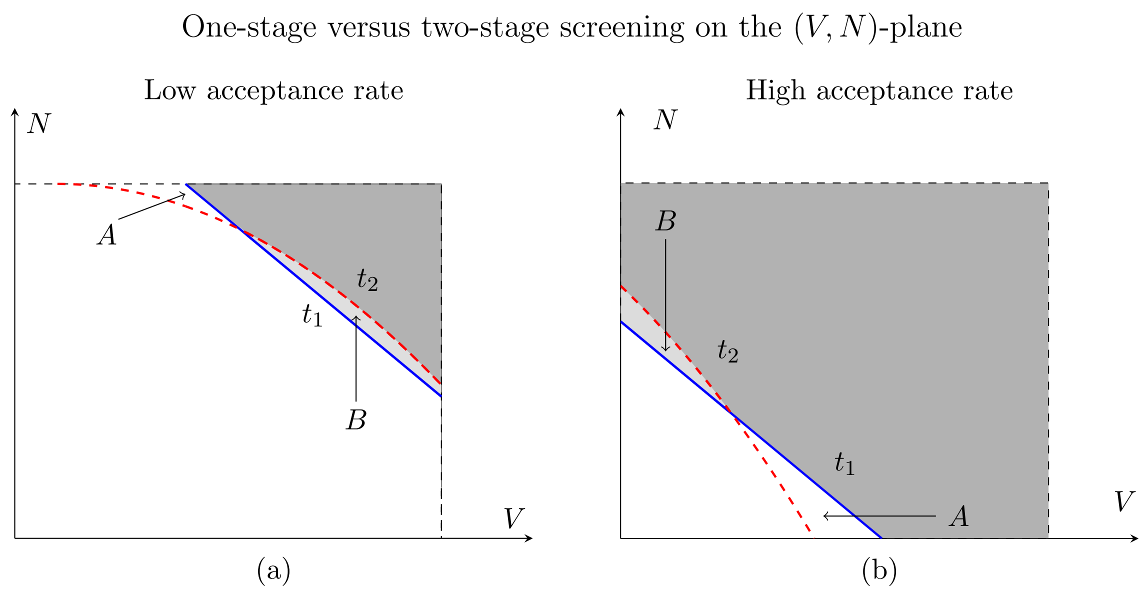

Figure 1: Both figures illustrate a comparison between one- and two-stage screening. Graph (a) provides a comparison under a low acceptance rate (a small p) and Graph (b) provides a similar comparison under a high acceptance rate (i.e., a large p). For every v, the interval above each curve, at $V = v$, describes the conditional probability that a value of v passes the i-stage screening, which is $[\Pr(N \geqslant t_i - v)]^i$, for $i = 1, 2$. Under the one-stage screening, which is depicted by the straight and solid (blue) lines t_1, only gray areas pass the screening. Under the two-stage screening, depicted by the parabolic and dashed (red) lines t_2, the light gray area (B) is eliminated, while all elements in the white area (A) pass the screening.

[9] We exploit the facts that all noises are distributed according to N and the same threshold is used twice to translate the cut-off rule from the two-stage process into the (V, N)-plane.

Beyond the technicalities, Theorem 1 is based on the idea that an additional stage must be accompanied by lower thresholds, relative to the one-stage screening under an overall capacity constraint. In the second stage, this reduction introduces a challenge in that the smaller sample size amplifies the effect of noise in distorting the underlying distribution of values. Specifically, once lower values pass through the first stage in the two-stage process, they have a higher likelihood of passing through the second stage as well, due to the heightened impact of noise given the reduced sample size. This phenomenon is illustrated by the LHS intersection in Figure 1(a). In contrast, in the one-stage process, these lower values are entirely eliminated because of the higher threshold, once again giving rise to the LHS crossing of the two curves.

Remark 1. *The result of* Theorem 1 *can be extended beyond the assumptions of bounded supports and positive densities. For example,* Theorem 1 *also holds for* (i) *noises with probability density functions that monotonically diminish from a certain point,* (ii) *noises with asymmetric distributions, and even* (iii) *noises with unbounded supports and probability density functions that decrease with sufficient rapidity. Given its technical nature, a full characterization is left for future research, while all statements and proofs in* Section 3.1 *hold for asymmetric noises.*

The next two theorems extend Theorem 1 in two ways. Theorem 2 considers non-fixed strategies while maintaining i.i.d. noises, and Theorem 3 incorporates different noise distributions throughout the stages. These extensions require certain limitations on the two-stage screening strategy. Specifically, allowing *any* two-stage strategy $\tau_2 = (t_1, t_2)$ can effectively reduce the process to a one-stage strategy. For example, if t_2 is sufficiently low and t_1 is relatively high, then p_2 approaches 1, and the product $p_1 \cdot p_2 = p$ converges to p_1. If p_1 is indeed arbitrarily close to p, the two-stage process will become nearly equivalent to the one-stage process, rendering the second stage redundant with a negligible effect on the expected outcome.

Moreover, if the noise in the second stage is a contraction of the noise in the first stage,[10] performing the majority of the screening in the second stage guarantees a better expected outcome compared with a one-stage process (see Lagziel and Lehrer, 2022). Based on this, we introduce the following definitions of *distinct* two-stage strategies.

Definition 2. *Fix* $\varepsilon \in (0, 1/2)$ *and a two-stage screening problem* SP_2 *given a strategy* $\tau = (t_1, t_2)$*. Note that* τ *defines two specific capacities* (p_1, p_2)*. We assert that* τ *is* ε-distinct (*from a one-stage strategy*) *if* $\max\{p_1, p_2\} < 1 - \varepsilon$*. In addition,* τ *is* fully ε-distinct *if* $\varepsilon < p_2 < 1 - \varepsilon$*.*

In simple terms, a two-stage screening strategy is distinct from a one-stage strategy if no stage is redundant (i.e., if no stage eliminates less than an ε proportion of the inspected elements). Moreover,

[10] A noise variable N_2 is a contraction of N_1 if the two are independent, $N_1 \sim N$, and $N_2 \sim cN$, where $c \in (0, 1)$. That is, N_1 and N_2 are independent and distributed according to N, with N_2 scaled by a contraction factor c.

a fully distinct strategy sustains a constraint on the second stage such that the second stage is not redundant and not all the screening occurs in the second stage. As noted previously, these constraints are necessary, otherwise one stage could be effectively eliminated from the two-stage process and generate results that are at least as good as those in the one-stage screening.

These limitations may seem restrictive, but there is a clear and natural logic behind them. Generally speaking, the advanced stages in many screening processes are more accurate; therefore, costlier. For example, interviews with chief executives or seminars in academic institutions consume a lot of time and effort from busy, time-constrained individuals. It would be extremely costly to conduct most of the screening in these stages rather than eliminate most of the applicants in preliminary stages (e.g., when screening CVs). Therefore, the capacity of elements that reach these advanced stages must be limited, and Definition 2 provides such a limitation. We examine this aspect more broadly in Section 4.

The following Theorem 2 states that even when diverting from fixed-threshold strategies, the dominance of one-stage screening over the two-stage process remains valid, provided the strategy is ϵ-distinct and the capacity is sufficiently small. The intuition behind this result begins with the flexibility to choose any ϵ-distinct strategy in the two-stage screening. To prevent the two-stage process from effectively collapsing into a one-stage process (e.g., when p_2 approaches 1 while p_1 converges to p), we restrict our consideration to ϵ-distinct strategies. These strategies ensure sufficient variation among the valuations passing through the first screening stage. Consequently, the noise variable becomes more disruptive in the second stage, allowing suboptimal elements to pass the second-stage screening as well.

Theorem 2. *Fix $\varepsilon \in (0, 1/2)$. For every impact variable V and i.i.d. noise variables N_1 and N_2, there exists $p_\varepsilon > 0$ such that for every capacity $p < p_\varepsilon$, one-stage screening under (V, N_1, p) stochastically dominates two-stage screening under $(V, \{N_1, N_2\}, p)$ given an ε-distinct strategy.*

Recall that the restriction to an ε-distinct strategy is essential, otherwise one of the two stages can be effectively eliminated, trivially transforming the two-stage screening process into a one-stage process. Subsequently, one could ask whether Theorem 2 holds as long as the capacities p_1 and p_2 remain strictly positive. In that case, the dominance of the one-stage process could vary with the underlying distributions of V and N.

The next theorem extends Theorem 2 by introducing different noise distributions into the two-stage process. For this extension to hold, we focus exclusively on fully distinct strategies, ensuring that neither stage dominates the process. The intuition aligns with that of Theorem 2: Fully distinct strategies guarantee sufficient variation in the valuations reaching the second screening stage while preventing all the screening from occurring in either stage. This condition is essential because no restrictions are imposed on the noise distributions. Consequently, the influence of both noise variables becomes critical, enabling suboptimal elements to pass through the screening process.

Theorem 3. *Fix $\varepsilon \in (0, 1/2)$, an impact variable V, and two independent noise variables N_1 and N_2, with possibly distinct distributions. There exists $p_\varepsilon > 0$ such that for every capacity $p < p_\varepsilon$, one-stage screening under (V, N_1, p) stochastically dominates two-stage screening under $(V, \{N_1, N_2\}, p)$ given any fully ε-distinct strategy.*

Note that we do not limit the distribution of N_2 relative to N_1 in Theorem 3, other than the general condition of a strictly positive density on some interval. This is more than a mere technicality as it demonstrates that the additional stage can still distort the screening, even if the additional stage is extremely accurate (i.e., even if the noise is relatively mild). In other words, esoteric distributions are not needed to exemplify our results as we simply extend the screening to the top of the distribution.

In contrast, if the DM wishes to remove only a small portion from the bottom of the distribution, then two-stage screening becomes superior relative to the one-stage process. The intuition is that once the sample size is significantly large in both stages, the noises' influence becomes limited, and this makes the additional stage worthwhile. The following theorem is a mirror image of Theorem 1 for cases in which the DM focuses on low-level screening (i.e., when the capacity is relatively high), indicating that two-stage screening dominates the one-stage process.

Theorem 4. *For every impact variable V and i.i.d. noise variables N_1 and N_2, there exists $p_0 > 0$ such that for every capacity $p > p_0$, two-stage screening under $(V, \{N_1, N_2\}, p)$ given a fixed-threshold strategy stochastically dominates one-stage screening under (V, N_1, p).*

Combined, Theorem 1 and Theorem 4 imply that the superiority of one screening method over another, varying in the number of stages, is highly dependent on the capacity constraint. The transition from a sufficiently low to a sufficiently high capacity, given a fixed-threshold strategy, exemplifies how an additional screening stage transforms from a burden to an advantage. A question that remains for future research is whether this transition occurs at a single point such that one-stage screening is superior below a given capacity and inferior above it, or whether this transition occurs at multiple points.

Next, we proceed with the second part of our analysis to prove that adding a considerable number of stages strictly improves the screening process.

3.2 Convergence to perfect screening

The results in Section 3.1 may impart the false impression that dynamic screening is inefficient. In this section we prove that this conclusion is false by demonstrating that a multi-stage process eventually yields the first-best outcome. Theorem 4 provides some intuition for this, since a sufficiently high capacity ensures that additional stages only improve the screening, which is indeed the case when using multiple stages. We establish this conclusion through two supporting results. The first, Theorem 5, concerns infinite screening problems and shows that any increasing strategy (i.e., a strategy under which threshold values can only increase along the stages) generates the first-best outcome. The second, Theorem 6, shows that the two previously introduced fixed strategies generate posterior distributions that converge, in distribution, to the first best result.

3.2.1 A Perfect screening strategy

In every screening problem, the best the DM can strive for is a screening procedure that yields a result as if there is no noise whatsoever — a result that we refer to as *a perfect screening.* Formally, given a screening problem SP, a strategy τ yields a *perfect screening* if $V_{\text{SP}}(\tau) \sim V|\{V \geqslant v_p\}$, where v_p denotes the $(1-p)$-quantile of V (i.e., $\Pr(V \geqslant v_p) = p$). In other words, a *perfect screening* strategy induces a distribution that is equivalent to a screening process without noise, while maintaining the same capacity constraint.

Starting with infinite screening problems $\text{SP} = (V, \{N_i\}_{i=1}^{\infty}, p)$, such that all noise variables are i.i.d., we assert that an infinite strategy $\tau = (t_1, t_2, \dots) \in \mathbb{R}^{\infty}$ is *increasing* if $t_{k+1} \geqslant t_k$ for every $k \geqslant 1$. That is, an increasing strategy entails that the screening along the stages becomes stricter, in the sense that the cut-off values increase. The following theorem shows that every increasing strategy produces a perfect screening in every infinite screening problem.

Theorem 5. *In an infinite screening problem, every increasing strategy induces a perfect screening.*

The motivation behind the statement and proof of Theorem 5 originates from the fact that sub-optimal elements are discarded with some "patience" from the DM. Therefore, even if the noise is rather disruptive for screening; for example, an almost-binary noise with a large variance,[11] a perfect screening remains feasible since suboptimal elements are slowly screened throughout the stages.

3.2.2 Fixed stationary strategies converge to perfect screening

After establishing that every infinite and increasing strategy yields a perfect screening, it may be desirable to consider a more practical finite set-up. In practice, whether we consider screening job applicants or "cleansing" datasets, DMs cannot feasibly commit to infinite screening stages, making the finite set-up the only practical choice. In such scenarios, a basic question is whether simple finite strategies converge to a perfect screening. In this section we address this question and provide two fixed strategies that converge to a perfect screening.

[11]These types of noises (and others) are prone to screening biases; see Lagziel and Lehrer (2019) for more details.

As previously defined in Subsection 2.1, we consider the *fixed-threshold strategy*, which maintains the same threshold value throughout the stages, and the *fixed-capacity strategy*, which dictates the same capacity in all stages. For each of these strategies we demonstrate that the induced distribution converges, in distribution, to a perfect screening outcome. More formally, we assert that a strategy τ *converges to a perfect screening* if, for every k-stage screening problem $\text{SP} = (V, \{N\}_{i=1}^{k}, p)$ where all noise variables are i.i.d., we have that $V_{\text{SP}}(\tau) \xrightarrow{d} V|\{V \geqslant v_p\}$ as $k \to \infty$. In light of Theorem 5, note that the fixed-capacity strategy does not converge to an infinite increasing strategy. In fact, the limit of the fixed-capacity strategy is not well defined for an infinite screening problem with a capacity constraint of $p \in (0,1)$. Therefore, establishing its convergence to a perfect screening outcome requires a separate approach and proof, which is given in the following Theorem 6:

Theorem 6. *The fixed-threshold strategy and the fixed-capacity strategy converge to a perfect screening.*

The use of fixed strategies accommodates a gradual screening process such that only considerably low valuations are eliminated, while most elements proceed to subsequent stages. Although we can devise different non-fixed strategies that converge to a perfect screening, it is clear that not all strategies will do so. For example, consider a screening strategy under which the first-stage threshold is too high such that the capacity constraint is not violated, yet some values above v_p are partially eliminated in the first stage. This poses a challenge for perfect screening, which requires that the posterior distribution (after the screening takes place) converges to $V|\{V \geqslant v_p\}$. Once values above v_p are partially eliminated in preliminary screening stages, the resulting posterior distribution will not converge to $V|\{V \geqslant v_p\}$.

4 Discussion

4.1 Cost of screening

Although our comparison of screening processes does not explicitly incorporate a cost function, we do not remain naive to this consideration. The basic intuition is that additional stages and improved accuracy are costlier; therefore, the two criteria with the superior expected results should be balanced. For example, consider a DM who sets a screening process given that advanced stages are more accurate (under some metric) than previous ones. A cost-minimization analysis would typically dictate that most of the screening, in terms of capacities, is performed in preliminary stages rather than in advanced ones. In other words, because the screening becomes costlier, the DM limits the mass of elements that reach advanced stages. Therefore, we can focus on the capacities as an implicit proxy/measure for the needed stages and accuracy. Notably, our results in Subsection 3.1, namely, Theorem 1 and Theorems 2 and 3, implicitly contain this feature even without an explicit cost function.

More formally, consider a k-stage screening process with the illustrative cost function $\sum_i \alpha_i \log\left(\frac{\prod_{j \leqslant i-1} p_j}{p}\right)$, where $p_0 = 1$ and $\alpha_{i+1} \geqslant \alpha_i > 0$ are accuracy indices for stages $i+1$ and i, respectively (i.e., a higher α

indicates more informative screening under some metric, as in Lehmann (1988) or Lagziel and Lehrer (2022), among others). The intuition behind this function is that the cost of every stage i increases with respect to (i) the informativeness α_i, and (ii) the mass of the inspected elements $\prod_{j=0}^{i-1} p_j$. For simplicity, consider $k = 2$ and assume the DM is bound by a binding budget constraint $C > 0$. To meet the condition $\alpha_1 \log\left(\frac{1}{p}\right) + \alpha_2 \log\left(\frac{p_1}{p}\right) \leqslant C$, the DM must ensure p_1 is sufficiently close to p, which subsequently pushes p_2 toward 1. It is straightforward to verify that this constraint becomes even more binding as α_2 increases.

We can also take a different perspective altogether, considering a set-up in which the accuracy of every stage/test, and therefore the cost, depend directly on the induced capacities. For example, consider the extreme case in which an examination is performed in stage i under a capacity constraint of $p_i = 1$. What would the cost be in this case? Evidently, the cost should be zero because no screening occurs. Therefore, we one can consider another illustrative cost function of the form $-\sum_i \log(p_i)$, so that the cost of not performing a screening is zero (that is, $\log(p_i) = 0$ if and only if $p_i = 1$), and it increases as the capacity decreases. In this example, it is clear that $-\sum_i \log(p_i) = -\log(p)$ and the cost only depends on the overall capacity constraint. In such cases, the DM would be solely concerned with comparing the outcomes of one- and k-stage screening processes.

5 Summary

This study presents an analysis of dynamic screening, demonstrating that the quality of the screening process is not necessarily a monotone function of the number of stages. Specifically, a single screening stage can be added such that the overall quality of the screening process decreases, whereas a few additional stages can significantly improve the process.

A few natural questions arise from our analysis. First, what occurs when transitioning from a single-stage process to one with many stages? It seems reasonable that first-order stochastic dominance does not revert in a single stage, but a slow transition from one posterior distribution to another occurs. Identifying the posterior distribution's evolution as a function of the number of stages is a crucial follow-up question for future inquiry, particularly for applicative purposes.

Another rather difficult question to address in future research is related to the nature of the noise throughout the process. In our framework, noises are given exogenously, whereas in practice, noises are endogenously determined according to existing (technical and monetary) capacities and feasibility constraints. Because the technical complexity of such questions is potentially overwhelming, it might be essential (and even more interesting) to adopt a combined empirical–theoretical perspective with assumptions and theoretical analysis based on actual data.

References

Arrow, K. J., D. Blackwell, and M. A. Girshick (1949): "Bayes and minimax solutions of sequential decision problems," *Econometrica*, 17, 213–244, publisher: JSTOR.

Athey, S. and J. Levin (2018): "The value of information in monotone decision problems," *Research in Economics*, 72, 101–116, publisher: Academic Press.

Ben-Yashar, R. and J. Paroush (2000): "A nonasymptotic Condorcet jury theorem," *Social Choice and Welfare*, 17, 189–199, publisher: Springer.

Berend, D. and L. Sapir (2005): "Monotonicity in Condorcet Jury Theorem," *Social Choice and Welfare*, 24, 83–92.

Bikhchandani, S., D. A. Hirshleifer, O. Tamuz, and I. Welch (2021): "Information Cascades and Social Learning," *SSRN Electronic Journal*.

Chambers, C. P. and P. J. Healy (2011): "Reversals of signal-posterior monotonicity for any bounded prior," *Mathematical Social Sciences*, 61, 178–180, publisher: North-Holland.

Condorcet, M. d. (1785): *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, Paris: L'imprimerie Royale.

Courty, P. and L. Hao (2000): "Sequential Screening," *The Review of Economic Studies*, 67, 697–717.

Ganuza, J.-J. and J. S. Penalva (2010): "Signal Orderings Based on Dispersion and the Supply of Private Information in Auctions," *Econometrica*, 78, 1007–1030, publisher: The Econometric Society.

Lagziel, D. and E. Lehrer (2019): "A Bias of Screening," *American Economic Review: Insights*, 1, 343–356.

——— (2022): "Screening Dominance: A Comparison of Noisy Signals," *American Economic Journal: Microeconomics*, 14, 1–24.

Lehmann, E. L. (1988): "Comparing Location Experiments," *Annals of Statistics*, 16, 521–533, publisher: Institute of Mathematical Statistics.

Milgrom, P. R. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12, 380–391, publisher: RAND Corporation.

Pavan, A., I. Segal, and J. Toikka (2014): "Dynamic Mechanism Design: A Myersonian Approach," *Econometrica*, 82, 601–653, publisher: The Econometric Society.

Quah, J. K.-H. and B. Strulovici (2009): "Comparative Statics, Informativeness, and the Interval Dominance Order," *Econometrica*, 77, 1949–1992.

Wald, A. (1939): "Contributions to the Theory of Statistical Estimation and Testing Hypotheses," *Annals of Mathematical Statistics*, 10, 299–326, publisher: Institute of Mathematical Statistics.

——— (1947): "Foundations of a General Theory of Sequential Decision Functions," *Econometrica*, 15, 279–313, publisher: JSTOR.

A Appendices

A.1 Proof of Theorem 1

Proof. Fix an impact variable V, a noise variable N, and some capacity $p \in (0,1)$. Denote the one- and two-stage screening problems by $\text{SP}_k = (V, \{N_i\}_{i=1}^k, p)$, where $k = 1, 2$, and all noise variable are independent and distributed according to N (i.e., $N_i \sim N$ for every $i \geqslant 1$). Denote the screening strategy (and threshold) of SP_1 and SP_2 by $\tau_1 = (t_1)$ and $\tau_2 = (t_2, t_2)$, respectively. Recall that we consider a two-stage screening process with a fixed-threshold strategy. Note the following:

$$\begin{aligned} F_{V_1}(t) &= \Pr\left(V \leqslant t | V + N \geqslant t_1\right) \\ &= \tfrac{1}{p}\Pr\left(V \leqslant t, V + N \geqslant t_1\right) \\ &= \tfrac{1}{p}\int_{-\infty}^{t} f_V(x)\Pr\left(N \geqslant t_1 - x\right) dx. \end{aligned}$$

Therefore,

$$\begin{aligned} f_{V_1}(t) &= \tfrac{1}{p} f_V(t)\Pr\left(N \geqslant t_1 - t\right) \\ &= \begin{cases} \tfrac{1}{p} f_V(t) G\left(t_1 - t\right), & \text{for } t_1 - \overline{N} \leqslant t \leqslant \overline{V}, \\ 0, & \text{otherwise,} \end{cases} \end{aligned}$$

where $G(x) = \Pr(N \geqslant x)$ is a differentiable decreasing function, such that $G(x) = 0$ if $x \geqslant \overline{N}$, and $G(x) = 1$ if $x \leqslant \underline{N}$. Similarly,

$$\begin{aligned} f_{V_2}(t) &= \tfrac{1}{p} f_V(t)[G\left(t_2 - t\right)]^2 \\ &= \begin{cases} \tfrac{1}{p} f_V(t)[G\left(t_2 - t\right)]^2, & \text{for } t_2 - \overline{N} \leqslant t \leqslant \overline{V}, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

The capacity constraint implies that $t_2 < t_1$, and both values converge to $\overline{V} + \overline{N}$ as $p \to 0$. Moreover, it follows that $\underline{V_2} < \underline{V_1}$, which suggests that $f_{V_2}(t) > f_{V_1}(t)$ in some interval. Therefore, the two densities intersect at least once, otherwise $f_{V_2}(t) > f_{V_1}(t)$ for every t in both supports. We will prove that V_1 first-order stochastically dominates (FOSD) V_2 by demonstrating that a range of values close to $\overline{V} + \overline{N}$ exists (equivalently, a range for p close to zero), such that for every t_1 and every $t_2 < t_1$ in that interval, there exists a unique (interior) point $t \in (t_1 - \overline{N}, \overline{V})$ such that $G(t_1 - t) = [G(t_2 - t)]^2$; namely, a single crossing between the two densities.

Define the function

$$H(t|t_1, t_2) = G(t_1 - t) - [G(t_2 - t)]^2.$$

With some abuse of notation, we allow t_1 and t_2 to vary independent of p. Note that H is continuously differentiable and $H'(t|t_1, t_2) = -G'(t_1 - t) + 2G(t_2 - t)G'(t_2 - t)$. Assume, by contradiction, that for every interval $I = (\overline{V} + \overline{N} - c, \overline{V} + \overline{N})$ where $c > 0$, there exist a point $t_1 \in I$ and a point $t_2 \in (\overline{V} + \overline{N} - c,\, t_1)$, such that the equation $H(t|t_1, t_2) = 0$ admits at least two distinct solutions with respect to t in the interval $I_1 = (t_1 - \overline{N},\, \overline{V})$. By the mean-value theorem, there exists a point $t' \in I_1$ such that $H'(t'|t_1, t_2) = 0$. As H is continuously differentiable, the equality $H'(t'|t_1, t_2) = 0$ holds for $c \to 0$, or equivalently, for $t' \to t_1 - \overline{N}$ and $t_2 \to t_1$.[12] Therefore, we obtain the following:

$$H'(t_1 - \overline{N}|t_1, t_1) \quad = \quad -G'(\overline{N}) + 2G(\overline{N})G'(\overline{N}) = -G'(\overline{N}) > 0$$

where the inequality follows from the fact that G is strictly decreasing on $[\underline{N}, \overline{N}]$ and $G(\overline{N}) = 0$.

We conclude that $H'(t|t_1, t_2)$ is strictly positive as p tends toward zero (and $\{t_1, t_2\}$ are taken according to the capacity constraint), and there are no two distinct solutions given a sufficiently small capacity p. Subsequently, for every $t \in [t_2 - \overline{N}, \overline{V}]$, it follows that $G(t_1 - t) \lesseqgtr [G(t_2 - t)]^2$ and $f_{V_1}(t) \lesseqgtr f_{V_2}(t)$, for $t \lesseqgtr t_p$ where t_p is the single intersection of the chosen capacity. This establishes that V_1 FOSD V_2. ■

A.2 Unified proof for Theorems 2 and 3

The proofs of Theorems 2 and 3 follow a similar structure to that of Theorem 1. To enhance readability and improve clarity, we provide a unified proof for both theorems.

Proof. Fix $\varepsilon \in (0, 1/2)$ and two screening problems, $\text{SP}_1 = (V, \{N_1\}, p)$ and $\text{SP}_2 = (V, \{N_1, N_2\}, p)$, where $p < \varepsilon$ and N_1 and N_2 are independent noise variables. Denote the screening strategy (and threshold) of SP_1 and SP_2 by $\tau_1 = (t_1)$ and $\tau_2 = (t_2^1, t_2^2)$, respectively. Recall that V_k relates to the posterior distribution of values given the k-stage screening process (for $k = 1, 2$), and the support of V_k is denoted by $[\underline{V_k}, \overline{V_k}]$.

[12] Note that G is continuously differentiable on $(\underline{N}, \overline{N})$; therefore, in the stated limit we consider the LHS derivative of G, i.e., $G'(\overline{N}) = \lim_{x \to \overline{N}^-} \frac{G(x) - G(\overline{N})}{x - \overline{N}} < 0$ as G is strictly decreasing and with a strictly positive density on $[\underline{N}, \overline{N}]$.

The proof consists of three parts. We first prove that $\underline{V_2} < \underline{V_1}$, then we present the densities f_{V_1} and f_{V_2}, and finally we show that the densities intersect exactly once, establishing the stated FOSD property.

Part 1: Proving that $\underline{V_2} < \underline{V_1}$.

If N_1 and N_2 are identically distributed, then the stages are interchangeable and we can assume without a loss of generality that $t_2^1 \geqslant t_2^2$. Moreover, the capacity constraint implies that $t_2^1 = \max\{t_2^1, t_2^2\} < t_1$, which also implies that $\underline{V_2} < \underline{V_1}$, assuming that p is sufficiently small such that $\min\{\underline{V_1}, \underline{V_2}\} > \underline{V}$ (i.e., assuming that the posteriors' lower bounds are strictly above the minimal level $\underline{V}$).

Otherwise, N_1 and N_2 are not identically distributed but the strategy is fully ϵ-distinct; thus, $p_2 > \varepsilon$ and $p_1 \cdot p_2 = p$. Therefore, if $p \to 0$, then $p_1 \to 0$ and $t_2^1 \to \overline{V} + \overline{N_1}$. Thus, from some point onward (namely, as long as t_2^1 is sufficiently close to $\overline{V} + \overline{N_1}$), the support of $V|\{V + N_1 \geqslant t_2^1\}$ shrinks toward $\overline{V}$, and according to Lemma 1 below, this indicates that t_2^2 is bounded away from $\underline{X} + \overline{N_2}$, where $X \sim V|\{V + N_1 \geqslant t_2^1\}$. In other words, for every sufficiently small p, the lower bound $\underline{V_2}$ of V_2 is $t_2^1 - \overline{N_1}$. Recall that $t_2^1 < t_1$ (otherwise, $p_1 < p$ and the capacity constraint is violated); therefore, we conclude that $\underline{V_2} < \underline{V_1}$ for sufficiently small capacities as in the case of i.i.d. noise variables. Thus, we henceforth assume that p is sufficiently small to sustain this condition for every fully ε-distinct two-stage screening strategy.

Part 2: The densities of V_1 and V_2.

Similar to the proof of Theorem 1, we have the following:

$$f_{V_1}(t) \quad = \quad \begin{cases} \frac{1}{p} f_V(t) G_1\left(t_1 - t\right), & \text{for } t_1 - \overline{N_1} \leqslant t \leqslant \overline{V}, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$f_{V_2}(t) \quad = \quad \begin{cases} \frac{1}{p} f_V(t) G_1\left(t_2^1 - t\right) G_2\left(t_2^2 - t\right), & \text{for } t_2^1 - \overline{N_1} \leqslant t \leqslant \overline{V}, \\ 0, & \text{otherwise,} \end{cases}$$

where $G_i(x) = \Pr(N_i \geqslant x)$ for $i = 1, 2$ are differentiable and decreasing functions, such that $G_i(x) = 0$ if $x \geqslant \overline{N_1}$, and $G_i(x) = 1$ if $x \leqslant \underline{N_1}$.

As established in the proof of Theorem 1 and because $\underline{V_2} < \underline{V_1}$, the graphs $f_{V_1}(\cdot)$ and $f_{V_2}(\cdot)$ must intersect in at least one interior point. Otherwise, $f_{V_2}(t) > f_{V_1}(t)$ for every $t \in (\underline{V_2}, \overline{V})$, thus violating the capacity constraint of either V_1 or V_2. This leads us to the third part of the proof which demonstrates that exactly one intersection exists, confirming that V_1 FOSD V_2.

Part 3: Exactly one intersection of the densities.

Consider the function $H(t|t_1, t_2^1, t_2^2) = G_1(t_1 - t) - G_1(t_2^1 - t)G_2(t_2^2 - t)$. With some abuse of notation, we allow for t_1 and t_2^1 to vary independent of p, while t_2^2 maintains the relevant ε-distinct property. Assume, by contradiction, that for every interval $I = (\overline{V} + \overline{N_1} - c, \overline{V} + \overline{N_1})$ where $c > 0$, there exist: (i) a point $t_1 \in I$, (ii) a point $t_2^1 \in (\overline{V} + \overline{N_1} - c, t_1)$, and (iii) a point $t_2^2 < t_2^1$ satisfying the relevant ε-distinct property, such that the equation $H(t|t_1, t_2^1, t_2^2) = 0$ admits at least two distinct solutions with respect to t in the interval $I_1 = (t_1 - \overline{N_1}, \overline{V})$. Note that according to the described densities, $f_{V_1}(t) > 0$ implies that t ranges between $t_1 - \overline{N_1}$ and $\overline{V}$, whereas $t_1 \in I$.

By the mean-value theorem, there exists a point $t' \in I_1$ such that $H'(t'|t_1, t_2^1, t_2^2) = 0$. As H is continuously differentiable, the equality $H'(t'|t_1, t_2^1, t_2^2) = 0$ holds for $c \to 0$, or equivalently, for $t' \to t_1 - \overline{N_1}$ and $t_2^1 \to t_1$ as the capacity tends toward 0. Therefore, as in the proof of Theorem 1, we get the value $G_1'(\overline{N_1})$, which refers to the LHS derivative of G_1 at $\overline{N_1}$, and

$$\begin{aligned} H'(t_1 - \overline{N}|t_1, t_1, t_2^2) &= -G_1'(\overline{N_1}) + G_1'(\overline{N_1})G_2(t_2^2 - t_1 + \overline{N_1}) + G_1(\overline{N_1})G_2'(t_2^2 - t_1 + \overline{N_1}) \\ &= -G_1'(\overline{N_1}) + G_1'(\overline{N_1})G_2(t_2^2 - t_1 + \overline{N_1}), \end{aligned} \tag{3}$$

where we use the fact that $G_1(\overline{N_1}) = 0$. Whether the two-stage screening strategy is ε-distinct or fully ε-distinct, the bound on the capacity $p_2 = \Pr(X + N_2 > t_2^2) < 1 - \varepsilon$, where $X \sim V|\{V + N_1 \geqslant t_2^1\}$, holds. Therefore, given that p is sufficiently small, there exists $\delta > 0$ such that

$$t_2^2 > \underline{X} + \underline{N_2} + \delta = t_2^1 - \overline{N_1} + \underline{N_2} + \delta,$$

as $\underline{X} = t_2^1 - \overline{N_1}$, which implies that $t_2^2 + \overline{N_1} > t_2^1 + \underline{N_2} + \delta$; therefore, we obtain the following:

$$G_2(t_2^2 - t_1 + \overline{N_1}) < G_2(t_2^1 - t_1 + \underline{N_2} + \delta). \tag{4}$$

Combining Eqs. (3) and (4) yields

$$\begin{aligned} H'(t_1 - \overline{N}|t_1, t_1, t_2^2) &= -G_1'(\overline{N_1}) + G_1'(\overline{N})G_2(t_2^2 - t_1 + \overline{N_1}) \\ &> -G_1'(\overline{N_1}) + G_1'(\overline{N_1})G_2(t_2^1 - t_1 + \underline{N_2} + \delta) > 0, \end{aligned}$$

where the first inequality follows from Eq. (4) and the second inequality holds because $G_1'(\overline{N_1}) < 0$, as explained in Footnote 12, and t_2^1 is arbitrarily close to t_1 and $G_2(t_2^1 - t_1 + \underline{N_2} + \delta) < 1$. Specifically, $H'(\cdot|t_1, t_2^1, t_2^2) > 0$ remains strictly positive as p tends to zero, which contradicts the previous equality $H'(t'|t_1, t_2^1, t_2^2) = 0$ as p tends to 0. Therefore, in either case (and assuming the capacity is small), there exists a single intersection between the two densities f_{V_2} and f_{V_1}, and since $\underline{V_2} < \underline{V_1}$, we conclude that V_1 FOSD V_2. ■

Lemma 1. *For every noise variable N and every $\varepsilon > 0$, there exists $c > 0$ such that for every impact variable V supported on an interval I of length $c' < c$, the inequality $\Pr(V + N \geqslant t) > \varepsilon$ implies that $t < \underline{V} + \overline{N}$.*

Proof. Fix a noise variable N. Take $N_0 \in \text{Supp}(N)$ such that $\Pr(N \geqslant N_0) = \varepsilon$. Fix $c = \overline{N} - N_0$ and fix $a \in \mathbb{R}$. Consider the interval $I = [a, a + \overline{N} - N_0]$ and an impact variable V supported on I. Take $t = \underline{V} + \overline{N} = a + \overline{N}$ and compute $\Pr(V + N \geqslant t)$ as follows:

$$
\begin{aligned}
\Pr(V + N \geqslant a + \overline{N}) &= \int_a^{a+\overline{N}-N_0} f_V(x) \Pr\left(N \geqslant a + \overline{N} - x\right) dx \\
&< \int_a^{a+\overline{N}-N_0} f_V(x) \Pr\left(N \geqslant N_0\right) dx \\
&= \varepsilon \int_a^{a+\overline{N}-N_0} f_V(x) dx = \varepsilon.
\end{aligned}
$$

Therefore, for every V supported on an interval with a length of $\overline{N} - N_0$, the inequality $\Pr(V + N \geqslant t) > \varepsilon$ implies that $t < \underline{V} + \overline{N}$. Since the same computation holds for any interval of length smaller than $\overline{N} - N_0$, the statement holds. ∎

A.3 Proof of Theorem 4

Proof. Fix an impact variable V, a noise variable N, and a capacity $p > \Pr(V + N > \min\{\overline{V} + \underline{N}, \underline{V} + \overline{N}\})$. For every $k = 1, 2$, denote the k-stage screening problem by $\text{SP}_k = (V, \{N_i\}_{i=1}^k, p)$, where $N_1, N_2 \sim N$ are independent and identically distributed. Denote the screening strategy and threshold of SP_1 and SP_2 by $\tau_1 = (t_1)$ and $\tau_2 = (t_2, t_2)$, respectively. Note that for the given capacity (as well as higher capacities), $t_1 < \min\{\overline{V} + \underline{N}, \underline{V} + \overline{N}\}$. In addition, we know that $t_2 < t_1$ and both thresholds converge to $\underline{V} + \underline{N}$ as $p \to 1$.

Consider the construction presented in the proof of Theorem 1. Under SP_1 we obtain the following:

$$
f_{V_1}(t) = \begin{cases} \frac{1}{p} f_V(t) G\left(t_1 - t\right), & \text{for } \underline{V} \leqslant t \leqslant t_1 - \underline{N}, \\ \frac{1}{p} f_V(t), & \text{for } t_1 - \underline{N} \leqslant t \leqslant \overline{V}, \\ 0, & \text{otherwise,} \end{cases}
$$

where $G(x) = \Pr(N \geqslant x)$, and under SP_2 we obtain the following:

$$
f_{V_2}(t) = \begin{cases} \frac{1}{p} f_V(t) [G\left(t_2 - t\right)]^2, & \text{for } \underline{V} \leqslant t \leqslant t_2 - \underline{N}, \\ \frac{1}{p} f_V(t), & \text{for } t_2 - \underline{N} \leqslant t \leqslant \overline{V}, \\ 0, & \text{otherwise,} \end{cases}
$$

Since $t_2 < t_1$, along with the condition that densities are strictly positive, we can deduce that $f_{V_2}(t) \geqslant f_{V_1}(t)$ for every $t \in (t_2 - \underline{N}, \overline{V})$, and the inequality is strict for every $t \in (t_2 - \underline{N}, t_1 - \underline{N})$. Therefore, we can establish the stochastic dominance of V_2 over V_1 by showing that there exists a single intersection between the two densities at the $(\underline{V}, t_2 - \underline{N})$ interval.

The fact that at least one crossing exists is trivial due to the capacity constraint. Therefore assume, by contradiction, that there is more than one intersection independent of the high capacity p. That is, we assume that for every $t_1 < \min\{\overline{V} + \underline{N}, \underline{V} + \overline{N}\}$, there exists $t_2 \in (\underline{N} + \underline{V}, t_1)$ such that the equation $H(t|t_1, t_2) = G(t_1 - t) - [G(t_2 - t)]^2 = 0$ has two solutions in the interval $(\underline{V}, t_2 - \underline{N})$.

Similar to the proof of Theorem 1, the mean-value theorem ensures that there exists $t' \in (\underline{V}, t_2 - \underline{N})$ such that $H'(t|t_1, t_2) = -G'(t_1 - t) + 2G(t_2 - t)G'(t_2 - t) = 0$. By continuity, this holds for every t_1 and t_2 close to $\underline{V} + \underline{N}$; therefore, we can take the limit $t_i \to \underline{V} + \underline{N}$ and $t' \to \underline{V}$, which yields the following:

$$0 = -G'(\underline{N}) + 2G(\underline{N})G'(\underline{N}) = -G'(\underline{N}) + 2G'(\underline{N}) = G'(\underline{N}) < 0,$$

reaching a contradiction.

Therefore, there exists $p > \Pr(V + N > \min\{\overline{V} + \underline{N}, \underline{V} + \overline{N}\})$ such that the densities $f_{V_2}(t)$ and $f_{V_1}(t)$ intersect only once in the interval $(\underline{V}, t_2 - \underline{N})$, which establishes that V_2 FOSD V_1, as needed. ■

A.4 Proof of Theorem 5

Proof. Fix an infinite screening problem SP $= (V, N, p)$ with an increasing strategy $\tau = (t_1, t_2, \dots)$, such that $t_{k+1} \geqslant t_k$ for every $k \geqslant 1$. By definition, this strategy τ maintains the stated capacity $p \in (0, 1)$. Therefore, $t_k \leqslant \overline{V} + \overline{N}$ for every k, and the sequence $(t_1, t_2, \dots)$ monotonically converges to some $t_\infty \in \mathbb{R}$.

Consider the stage-k conditional distribution of V in some $t \in \mathbb{R}$, denoted $F_{V_k}(t)$. Clearly, the following holds:

$$\begin{aligned} F_{V_k}(t) &= \Pr\left(V_{k-1} \leqslant t | V_{k-1} + N_{k-1} \geqslant t_{k-1}\right) \\ &= \tfrac{1}{p_k} \Pr\left(V_{k-1} \leqslant t, V_{k-1} + N_{k-1} \geqslant t_{k-1}\right) \\ &= \tfrac{1}{p_k} \int_{-\infty}^{t} f_{V_{k-1}}(s) \Pr\left(N_{k-1} \geqslant t_{k-1} - s\right) ds \\ &= \tfrac{1}{p_k} \int_{-\infty}^{t} f_{V_{k-1}}(s) \Pr\left(N \geqslant t_{k-1} - s\right) ds. \end{aligned}$$

Therefore,

$$\begin{aligned} f_{V_k}(t) &= \tfrac{1}{p_k} f_{V_{k-1}}(t) \Pr\left(N \geqslant t_{k-1} - t\right) \\ &= \tfrac{1}{p_{k-1} p_k} f_{V_{k-2}}(t) \Pr\left(N \geqslant t_{k-2} - t\right) \Pr\left(N \geqslant t_{k-1} - t\right) \\ &= \tfrac{1}{\prod_{i=1}^{k} p_i} f_V(t) \prod_{i=1}^{k} \Pr\left(N \geqslant t_i - t\right), \end{aligned}$$

and

$$f_{V_{\mathrm{SP}(\tau)}}(t) = \lim_{k\to\infty} f_{V_k}(t) = \tfrac{1}{p} f_V(t) \prod_{i=1}^{\infty} \Pr\left(N \geqslant t_i - t\right).$$

Fix $t \in \mathbb{R}$ and recall that the increasing sequence $(t_i)_{i\in N}$ converges monotonically to t_∞. If $t < t_\infty - \underline{N}$, then there exists N_0 such that $t < t_i - \underline{N}$ for every $i > N_0$, i.e., $\underline{N} < t_i - t$ for every $i > N_0$, so $\Pr\left(N \geqslant t_i - t\right) < 1$. Therefore, $\prod_{i=1}^{\infty} \Pr\left(N \geqslant t_i - t\right) = 0$. Otherwise, $t \geqslant t_\infty - \underline{N}$, and $t \geqslant t_i - \underline{N}$ for every $i \in \mathbb{N}$. This implies that $\Pr\left(N \geqslant t_i - t\right) = 1$ for every $i \in \mathbb{N}$, and $\prod_{i=1}^{\infty} \Pr\left(N \geqslant t_i - t\right) = 1$. Therefore, we conclude the following:

$$f_{V_{\mathrm{SP}(\tau)}}(t) = \begin{cases} 0, & \text{for every } t < t_\infty - \underline{N}, \\ \frac{1}{p} f_V(t), & \text{for every } t \geqslant t_\infty - \underline{N}. \end{cases}$$

Clearly, the following holds:

$$\int_{\mathbb{R}} f_{V_{\mathrm{SP}(\tau)}}(t)dt = \int_{t_\infty - \underline{N}}^{\overline{V}} \tfrac{1}{p} f_V(t)dt = 1,$$

so $t_\infty - \underline{N} = v_p$, and τ induces a perfect screening. ■

A.5 Proof of Theorem 6

Proof. Fix an impact variable V, a noise variable N, and a capacity p. For every $k \in \mathbb{N}$, let $\mathrm{SP}_k = (V, \{N_i\}_{i=1}^{k}, p)$ denote a k-stage screening problem, where (V, N, p) are fixed, all noise variable are independent and distributed according to N, and consider the fixed-threshold strategy, $\tau = (t^k, \ldots, t^k)$. That is, t^k denotes the fixed threshold of τ in SP_k.

Recall that v_p denotes the $(1-p)$-quantile of V, i.e., $\Pr(V \geqslant v_p) = p$. We begin by establishing that $\{t^k\}_{k\geqslant 1}$ is a decreasing sequence that converges to $v_p + \underline{N}$ as $k \to \infty$. Note the distinction between the current set-up and the proof of Theorem 5, where we consider an increasing *infinite* strategy. In the case of an infinite screening problem, the capacity constraint only holds in the limit, and the capacity is strictly above p in every finite stage. However, in the k-stage screening process, the capacity constraint must be met at stage k. Therefore, if we take the fixed-threshold t^k (used in the k-stage screening problem) and apply an additional screening stage, using it as a threshold in the $k+1$-screening problem, the capacity would decrease below p, violating the constraint.

As shown in the proof of Theorem 5, for every $k \in \mathbb{N}$, the following holds:

$$f_{V_k}(t) \quad = \quad \tfrac{1}{p} f_V(t) \prod_{i=1}^{k} \Pr\left(N \geqslant t^k - t\right) = \tfrac{1}{p} f_V(t) \left[\Pr\left(N \geqslant t^k - t\right)\right]^k.$$

Therefore, if $t^k \leqslant t^{k+1}$, then the following holds:

$$f_{V_k}(t) \quad = \quad \tfrac{1}{p} f_V(t) \left[\Pr\left(N \geqslant t^k - t\right)\right]^k \geqslant \tfrac{1}{p} f_V(t) \left[\Pr\left(N \geqslant t^{k+1} - t\right)\right]^{k+1} = f_{V_{k+1}}(t),$$

and the inequality is strict for every t where $f_V(t)\Pr\left(N \geqslant t^k - t\right) > 0$ and $\Pr\left(N \geqslant t^k - t\right) < 1$. Thus, we reach a contradiction since at least one of the two variables, either V_{k+1} or V_k, is not normalized. Therefore, we conclude that $t^{k+1} < t^k$ for every $k \in \mathbb{N}$. In addition, $t^k \geqslant \underline{V} + \underline{N}$ for every k; otherwise, no screening will occur whatsoever and the capacity constraint would be violated. Thus, we deduce that $\{t^k\}_{k\geqslant 1}$ is a decreasing and bounded sequence, and subsequently converges. Denote the limit by $t^* = \lim_{k\to\infty} t^k$.

To show that $\{t^k\}_{k\geqslant 1}$ converges to $v_p + \underline{N}$, consider the infinite sequence $(t^*, t^*, \dots)$. The induced conditional distribution of V in the respective (infinite) screening problem SP, given that all threshold values equal t^*, is as follows:

$$f_{V_{\mathrm{SP}}}(t) \quad = \quad \tfrac{1}{p} f_V(t) \prod_{i=1}^{\infty} \Pr\left(N \geqslant t^* - t\right) = \begin{cases} 0, & \text{for every } t \text{ s.t. } \Pr\left(N \geqslant t^* - t\right) < 1, \\ \tfrac{1}{p} f_V(t), & \text{for every } t \text{ s.t. } \Pr\left(N \geqslant t^* - t\right) = 1. \end{cases}$$

To meet the capacity constraint p, we require that $\Pr\left(N \geqslant t^* - t\right) = 1$ if and only if $t \geqslant v_p$. In addition, $\Pr\left(N \geqslant t^* - t\right) = 1$ if and only if $t^* - t \leqslant \underline{N}$, i.e., if and only if $t \geqslant t^* - \underline{N}$. Combining the two necessary and sufficient conditions, $t \geqslant t^* - \underline{N}$ and $t \geqslant v_p$, we conclude that $t^* - \underline{N} = v_p$, as needed. Therefore, we conclude that $\lim_{k\to\infty} t^k = v_p + \underline{N}$.

Finally, we need to prove that V_k converges in distribution to $V|\{V \geqslant v_p\}$. Fix $t \in \mathbb{R}$. If $t > v_p$, then there exists $k_t \in \mathbb{N}$ such that $t^k - \underline{N} < t$ for every $k \geqslant k_t$, and $\Pr\left(N \geqslant t^k - t\right) = 1$. Therefore, $f_{V_k}(t) = \frac{1}{p} f_V(t)$ for every $t > v_p$ and every $k \geqslant k_t$. Furthermore, if $t < v_p$, then $\Pr(N \geqslant t^k - t) < \Pr(N \geqslant v_p + \underline{N} - t) < 1$ for every $k \in \mathbb{N}$, which yields the following:

$$f_{V_k}(t) = \tfrac{1}{p} f_V(t) \left[\Pr\left(N \geqslant t^k - t\right)\right]^k < \tfrac{1}{p} f_V(t) \left[\Pr(N \geqslant v_p + \underline{N} - t)\right]^k \to 0 \ \text{ as } k \to \infty,$$

as needed.

We next examine the fixed-capacity strategy. Again, for every $k \in \mathbb{N}$, let $\mathrm{SP}_k = (V, \{N\}_{i=1}^k, p)$ denote a k-stage screening problem, where (V, N, p) are fixed, and consider the fixed-capacity strategy, $\tau = (t_1^k, t_2^k, \dots, t_k^k)$. We can easily demonstrate that, upon a screening stage, the induced posterior distribution of the impact variable first-order stochastically dominates the prior distribution upon a screening stage (in fact, the two distributions sustain the monotone likelihood-ratio property). Therefore, to maintain a capacity of $p^{1/k}$ in every stage, the thresholds must strictly increase.

We begin by proving that for every $\varepsilon > 0$ and every $l \in \mathbb{N}$ there exists $K_{l,\varepsilon} \in \mathbb{N}$ such that for every $k > K_{l,\varepsilon}$ there are at least l^2 stages where the threshold is strictly higher than $v_p + \underline{N} - \varepsilon$, i.e., $|\{i : \ t_i^k > v_p + \underline{N} - \varepsilon\}| \geqslant l^2$.

Fix $\varepsilon > 0$, $l \in \mathbb{N}$, and let $l_k = |\{i : \ t_i^k > v_p + \underline{N} - \varepsilon\}|$ denote the number of stages that the threshold is above $v_p + \underline{N} - \varepsilon$. Note that the capacity up to stage $k - l_k$ is at least $\Pr(V > v_p - \varepsilon)$ since values above $v_p - \varepsilon$ pass the screening in these stages. Denote $p_\varepsilon = \Pr(V > v_p - \varepsilon)$ and note that $1 > p_\varepsilon > p$.

In the remaining l_k stages, the capacity is fixed to $p^{l_k/k}$ according to τ. Thus, when considering the overall capacity p, we obtain $p > p_\varepsilon \cdot p^{l_k/k}$, which translates to $\frac{l_k}{k} > 1 - \frac{\ln(p_\varepsilon)}{\ln(p)} > 0$. In other words, there exists $c > 0$, which is independent of k, such that $l_k > ck$. We can now fix $K_{l,\varepsilon} = \frac{l^2}{c}$ to obtain the needed result.

We next use this claim to prove that τ converges to a perfect screening. Fix $t < v_p$. Note $\delta = v_p - t$. Fix $l > 0$ so that $\Pr\left(N \geqslant \underline{N} + \frac{\delta}{2}\right) < 1 - \frac{1}{l}$. Take $K_{l,\delta/2}$ such that the previous claim holds, and consider $k > K_{l,\delta/2}$. Then,

$$\begin{aligned} f_{V_k}(t) &= \tfrac{1}{p} f_V(t) \prod_{i=1}^{k} \Pr\left(N \geqslant t^k - t\right) \\ &\leqslant \tfrac{1}{p} f_V(t) \left[\Pr\left(N \geqslant v_p + \underline{N} - \tfrac{\delta}{2} - t\right)\right]^{l^2} \\ &= \tfrac{1}{p} f_V(t) \left[\Pr\left(N \geqslant \underline{N} + \tfrac{\delta}{2}\right)\right]^{l^2} \\ &\leqslant \tfrac{1}{p} f_V(t) \left[1 - \tfrac{1}{l}\right]^{l^2} \to 0 \end{aligned}$$

as $l \to \infty$. Therefore, for every $t < v_p$, the density $f_{V_k}(t)$ converges to zero, which necessarily leads to a perfect screening. ∎