Strategic Sample Selection*

Alfredo Di Tillio† Marco Ottaviani‡ Peter Norman Sørensen§

May 2, 2019

Abstract

This paper develops the notion of accuracy for multidimensional experiments to characterize the welfare impact of sample selection from a larger presample. Sample selection benefits or hurts a decision maker according to whether the reverse hazard rate of the data distribution is log-supermodular—as in location experiments with normal noise—or log-submodular. Our results characterize situations in which potential sources of selection bias can be exploited, with broad implications for the choice and design of selected experiments. In strategic settings, selection arises in equilibrium when the sample is chosen by a biased researcher. Applied to educational testing, we determine whether allowing an examinee to choose which questions to answer dominates randomly selecting the same number of questions. We also characterize when the common-law right of peremptory challenge improves the quality of judgment by eliminating jurors with extreme views on either side.

Keywords: Strategic selection; Comparison of experiments; Accuracy; Persuasion; Welfare

JEL codes: D82, D83, C72, C90

*Research funded by the European Research Council through Advanced Grant 29583 (EVALIDEA). We thank Matteo Camboni for outstanding research assistance. We also thank, without implication, Pierpaolo Battigalli, Francesco Corielli, Danielle Li, Nicola Limodio, Karl Schlag, and seminar participants at Bocconi, Bruxelles, Caltech, Carlos III, Chicago, Como, Copenhagen, Duke, East Anglia, EUI, Genova, Gerzensee, Graz, Helsinki, Innsbruck, Leicester, Lisbon, London Business School, Lund, Madrid, Milan, Naples, NBER, Northwestern, Nottingham, Paris, Rome, Southampton, Surrey, Stockholm, Tel Aviv, Toulouse, and Vienna for helpful comments.

†Department of Economics and IGIER, Bocconi University, Via Roberto Sarfatti 25, 20136 Milan, Italy. Phone: +39–02–5836–5422. E-mail: alfredo.ditillio@unibocconi.it.

‡Department of Economics and IGIER, Bocconi University, Via Roberto Sarfatti 25, 20136 Milan, Italy. Phone: +39–02–5836–3385. E-mail: marco.ottaviani@unibocconi.it.

§Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Building 26, DK–1353 Copenhagen K, Denmark. Phone: +45–3532–3056. E-mail: peter.sorensen@econ.ku.dk.

1 Introduction

Empirical and experimental data are often nonrandomly selected, due to choices made by subjects under investigation or sample inclusion decisions by data analysts.[1] Suppose a new treatment is given to the healthiest patients rather than to random patients in a group. Because of selection, favorable outcomes are weaker evidence that the treatment is effective. But once the evaluator adjusts for selection, does inference improve or worsen compared to random assignment? Similar questions arise in other contexts. For example, when testing a student in an exam, should the teacher ask questions at random or allow the student to select the most preferred questions out of a larger batch? When feeding consumer reviews to potential buyers with limited attention, should an e-commerce platform post random reviews or allow the merchant to cherry-pick them? And how does the right of peremptory challenge—by which the attorney on each side of a trial can strike down a number of jurors—affect judgment quality?

These comparisons are all instances of the same issue: assessing the welfare impact of sample selection in a decision problem. There is an unknown state θ, which may represent the difference in outcome improvement between two treatments, a student's ability, or a defendant's level of guilt. An evaluator must choose an action—assign a grade, choose a treatment for the next patient, or decide on a sentence—knowing that marginally increasing the action decreases payoff when θ is low, and increases it when θ is high. More precisely, we assume that the evaluator has preferences in the general interval dominance ordered (IDO) class introduced by Quah and Strulovici (2009), encompassing monotone decision problems (Karlin and Rubin, 1956) and single-crossing preferences (Milgrom and Shannon, 1994).

The evaluator decides after observing the realization of a statistical experiment, a random vector $X = (X_1, \ldots, X_n)$ whose distribution depends on θ. For instance, X_i may represent a patient's outcome under either treatment, a student's potential performance in a question on a certain topic, or a juror's opinion. Our basic question is, in which of the following scenarios does the evaluator make better decisions:

- **Random Experiment**. The sample observations are i.i.d. draws from a state-dependent cumulative distribution function $F(\cdot|\theta)$.
- **Selected Experiment**. The sample observations are selected—possibly strategically, by another party—as the n highest out of $k > n$ presampled i.i.d. draws from $F(\cdot|\theta)$.

The impact of selection on the evaluator's welfare is in general ambiguous. To fix ideas, suppose the evaluator faces a simple hypothesis testing problem: two states $\theta_H > \theta_L$ and two actions, rejection (the correct choice in state θ_L) and acceptance (the correct choice in state θ_H). In a unidimensional ($n = 1$) location experiment with noise drawn from a normal distribution F, the sample observation is normal with mean θ_L in the low state and θ_H in the high state, as illustrated by the blue graphs in the left panel of Figure 1. The evaluator's optimal decision is to accept when the

[1] For instance, Heckman (1979) refers from the outset to these two sources of selection.

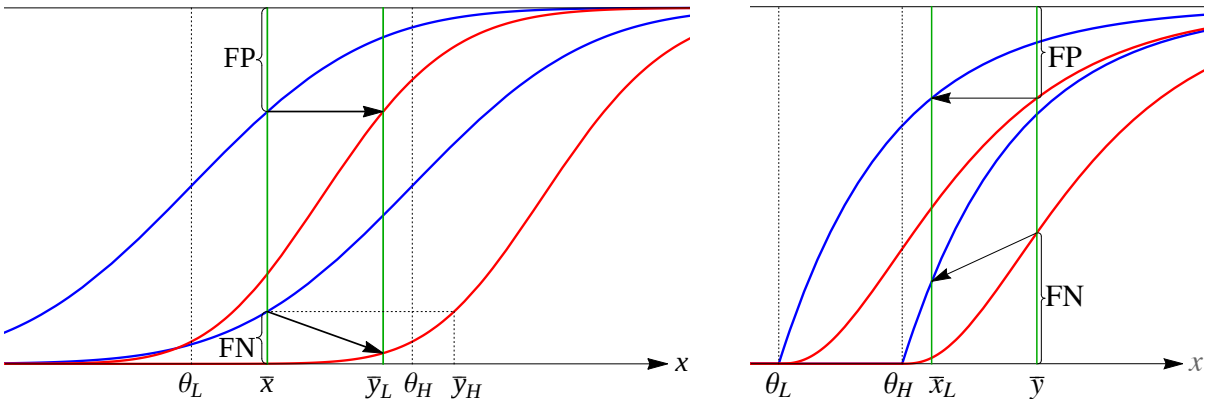

Figure 1: Selection provides more accurate information with normal noise (left) but less accurate information with exponential noise (right).

sample observation is above some cutoff $\bar{x}$. This is the familiar trade-off between the probability $1-F(\bar{x}-\theta_L)$ of a false positive (accepting in the low state, FP) and the probability $F(\bar{x}-\theta_H)$ of a false negative (rejecting in the high state, FN).[2]

How does a selected experiment compare? Selection changes the noise distribution to F^k, so the evaluator observes a random variable distributed according to $F^k(x-\theta_L)$ in the low state and $F^k(x-\theta_H)$ in the high state. This is illustrated by the red graphs in Figure 1. With normal noise, selection benefits the evaluator: by adopting the (possibly suboptimal) cutoff point $\bar{y}_L$ that induces as many false positives, the evaluator also induces fewer false negatives: $F^k(\bar{y}_L-\theta_H) < F(\bar{x}-\theta_H)$. But, as shown in the right panel of Figure 1, exactly the opposite is true with exponential noise: starting from any cutoff $\bar{y}$ for the selected experiment, the evaluator can match false positives, while lowering false negatives, by adopting the cutoff $\bar{x}_L$ in the random experiment.

What makes selection beneficial in some cases and harmful in others? To answer this question, we start from an observation of Lehmann (1988), who pointed out that an equivalent way to formulate the property that $\bar{y}_L$ induces as many false positives and fewer false negatives is to say that the selected experiment is *more accurate*.[3] This means that the cutoff point $\bar{y}_H$ inducing as many false *negatives*, defined by $F^k(\bar{y}_H-\theta_H) = F(\bar{x}-\theta_H)$, is larger than $\bar{y}_L$. By adopting the smaller cutoff $\bar{y}_L$ the evaluator necessarily induces more acceptance—and in particular more acceptance in the high state—than by adopting the larger cutoff $\bar{y}_H$. In effect, this is what happens in the normal case (and with opposite direction in the exponential case) depicted in Figure 1.

Our first main result identifies a necessary and sufficient condition for a larger presample size to increase or decrease accuracy in one-dimensional location experiments. Theorem 1 shows that

[2] A location problem with normal noise, like every other experiment considered in this paper, satisfies the monotone likelihood ratio property: given any two states, the higher the realization x, the higher the relative odds of the higher state. This property implies that the evaluator's optimal decision is increasing in x. With two actions and sample size $n=1$, this simply means choosing the higher action (acceptance) if and only if x is at least as large as some cutoff $\bar{x}$.

[3] The latter nomenclature is due to Persico (2000).

the evaluator welfare monotonically increases in k if the reverse hazard function of the noise distribution, $-\log F$, is logconcave, as with normal or logistic noise. Likewise, welfare monotonically decreases in k if and only if $-\log F$ is logconvex, as with exponential noise. There is only one case where the evaluator is indifferent to selection: noise drawn from the Gumbel distribution, $F(\varepsilon) = \exp(-\exp(\varepsilon))$. This is the only distribution with both logconcave and logconvex $-\log F$. Intuitively, our logconcavity criterion requires that neither the top tail of the distribution should be thicker than in the Gumbel distribution, nor the bottom tail should be thinner.

To characterize the impact of selection in the general case, with possibly non-additive noise and any sample size $n \geqslant 1$, we introduce a natural generalization of Lehmann's (1988) notion of accuracy. Our notion allows comparisons between any pair of experiments and shares the basic intuition with (for $n = 1$, it reduces to) Lehmann's (1988). To illustrate, consider two n-dimensional experiments X and Y and again a simple hypothesis testing setup. In experiment X the evaluator again adopts a cutoff strategy, but now the cutoff is a more complicated object, an $(n-1)$-dimensional hypersurface in $\mathbb{R}^n$.[4] Our notion requires a suitably defined cutoff hypersurface that induces as many false positives in Y as in X to lie below the one inducing as many false negatives. Much like in the unidimensional case, then, the evaluator must fare better with Y than with X.

Our notion of accuracy is the key technical tool needed to tackle the new issues arising in the multidimensional case. The main difficulty lies in the fact that sample observations are correlated with each other, even conditionally on the state. By disentangling the net value of information added by each observation, we can understand when selection adds or subtracts value to the evaluator's problem. Our main result, Theorem 2, shows that welfare monotonically increases or decreases in presample size, according to whether the reverse hazard rate $f(x|\theta)/F(x|\theta)$ is log-supermodular or log-submodular. In a location experiment, log-supermodularity reduces to logconcavity of the noise distribution's reverse hazard rate f/F. This condition strengthens the logconcavity criterion in Theorem 1. Intuitively, the noise distribution must be *increasingly* thinner at the top and thicker at the bottom, compared to the Gumbel distribution.

The analysis in this paper offers new insights to applied research. Selective sampling is typically viewed exclusively as a threat to the internal or external validity of an experiment. The main focus in the received literature is on identification issues—how to avoid or at least account for selection bias. Our results show that properly anticipated selection can have a beneficial impact on the quality of inference, and characterize precisely when it does, thus identifying a novel role for selective sampling.[5] The results can be used to discriminate situations where selection should indeed be avoided from those where instead it can be exploited. To a researcher choosing between two alternative datasets (one characterized by more selection than the other), as well as to a teacher deciding whether to allow examinee choice, the recommendation is immediate: check the posited data distribution for the conditions in Theorem 1 or Theorem 2.

[4] For example, with i.i.d. observations $x = (x_1, \ldots, x_n)$ from a location experiment with normal noise, the average observation is a sufficient statistic. In this case, the cutoff hypersurface has the form $\sum_i x_i/n = \tilde{x}$ for some $\tilde{x}$.

[5] As we discuss in the conclusion, selection may benefit even an unwary evaluator who does not anticipate it at all.

As a key advantage of our approach, the criteria we obtain do not depend on the specific decision problem, but rather hold for all preferences in the general IDO class. We also emphasize that, while sample selection often arises strategically—as suggested by the applications we mentioned—our comparative statics hold independently of the specific source of selection.

Our results open the way to new directions in the design of experiments. In a situation where more selection benefits the evaluator, presample size k is an additional channel of information that the evaluator can use to economize on sample size n.

Drawing on extreme value theory, we also analyze the impact of extreme selection, when presample size tends to infinity. Focusing on location experiments, in Theorem 3 we characterize the noise distributions such that the evaluator obtains the full-information payoff in the limit, and those such that information in the limit is less than full. As intuition suggests, if the noise distribution has support bounded above, extreme selection leads to full information—noise becomes concentrated around its upper bound. In the unbounded case, limit welfare is governed by the hazard rate of the noise distribution, $f(\varepsilon)/(1-F(\varepsilon))$: if and only if the hazard rate is unbounded—as for instance with normal noise—full information is approached in the limit. Returning to experiment design, we use Theorem 3 to show that when presampling costs are small, the evaluator always prefers to tolerate sample selection. We show that this conclusion holds whether presample size is optimally decided by the evaluator or strategically chosen by a sender.

Finally, our proof techniques suggest a general methodology that seamlessly accommodates other forms of selection. Before concluding the paper, we apply the general method of proof developed for Theorem 2 to characterize the impact of sampling from a truncated distribution (Theorem 4). The welfare impact of this other common type of selection is different in important cases. For example, with normal or logistic noise maximal selection benefits while truncation hurts the evaluator. We also analyze median selection, where the evaluator observes the median observation in a presample (Theorem 5). This analysis, based on a symmetric application of the argument of proof used for Theorem 2, allows us to derive conclusions for peremptory challenge.

Literature Contribution. Concerns about data selection and manipulation have long been voiced by the science and medicine literature and have led to important policy responses.[6] However, there are few economic models in the area.[7] An early exception is Blackwell and Hodges (1957), who analyze how an evaluator should optimally design a sequential experiment to minimize *selection bias*, a term they coined to represent the fraction of times a strategic researcher is able to correctly forecast the treatment assignment.[8] However, they did not model the information available to the

[6] See, for example, the analysis of Schulz, Chalmers, Hayes, and Altman (1995) and the CONSORT statement, http://www.consort-statement.org. See also Allcott (2015) for recent empirical evidence.

[7] Glaeser (2008) discusses a number of issues in this regard. Di Tillio, Ottaviani, and Sørensen (2017) compare different types of selection in the context of an illustrative model with binary noise, which violates the logconcavity assumption maintained in this paper.

[8] Blackwell and Hodges (1957) argue that selection bias is minimized by a truncated binomial design, according to which the initial allocations to treatment and control are selected independently with a fair coin, until half of the subjects are allocated to either treatment or control; from that point on, allocation is deterministic. Efron (1971), in-

researcher at the assignment stage. The ensuing literature focused on exogenous selection bias and on how to adjust for it, rather than on its strategic origin and its impact on the quality of inference, on which we focus. Once we explicitly model information, we characterize situations in which selection actually benefits the evaluator, contrary to what Blackwell and Hodges (1957) stipulate.

To the literature on stochastic orderings of order statistics, we contribute the characterization of noise distributions for which the maximum of k i.i.d. draws is more or less dispersed than a random draw, as explained in Section 3. Building on Lehmann's (1988) univariate notion of accuracy, we contribute a tool for comparing general multidimensional experiments with arbitrary correlation patterns. This tool allows us to nail down the welfare impact of maximal selection and flesh out the common logic behind the comparison of other forms of selection such as truncation, previously considered by Goel and DeGroot (1992), as explained in Section 5. Our analysis of extreme selection in Section 4 offers a novel economic application of the theoretical framework pioneered by Fisher and Tippett (1928) and Gnedenko (1943) as well as new insights for experiment design.

Relative to work on optimal persuasion following Rayo and Segal (2010) and Kamenica and Gentzkow (2011), in our strategic presampling game information acquisition is costly and information manipulation is naturally constrained by the need of reporting observations selected from the presample. With sample size $n = 1$, our sender discloses a single observation, as in the limited-attention model first proposed by Fishman and Hagerty (1990).[9] Thus, we have a signal-jamming model of equilibrium persuasion through presample collection and then sample selection. The researcher's choice of the size k of the presample is akin to the agent's effort choice in Holmström's (1999) classic career concern model. The wrinkle here is that this effort results in private information, which the researcher then uses to select the reported information.

In a complementary approach to modeling conflicts of interest in statistical testing, Banerjee, Chassang, Monteiro, and Snowberg (2017) propose a theory of an ambiguity-averse researcher facing an adversarial evaluator; see also Kasy (2016). In another complementary approach, Tetenov (2016) analyzes an evaluator's optimal commitment to a decision rule when privately informed researchers select into costly testing. Instead, we focus on the impact of the sender's selection on the welfare of an uncommitted evaluator.

2 Setup

An evaluator chooses an action $a \in A \subseteq \mathbb{R}$ under uncertainty about a state $\theta \in \Theta \subseteq \mathbb{R}$, where Θ is either a finite set or a (possibly unbounded) interval. The prior is represented by a density

stead, characterizes the selection bias resulting from a biased coin design, according to which the probability of current assignment to treatment is higher if previous randomizations resulted in excess balance of controls over treatments.

[9] See also Henry (2009), Dahm, Gonzàlez, and Porteiro (2009), Felgenhauer and Schulte (2014), Hoffmann, Inderst, and Ottaviani (2014), and Herresthal (2017) for persuasion models with endogenous information acquisition. Henry and Ottaviani (2019) analyze a dynamic model of persuasion with costly information acquisition à la Wald (1945), where information is truthfully reported at the time of application.

function $\pi(\theta)$ and the payoff function is $u : \Theta \times A \to \mathbb{R}$. For now we take the action set to be finite $A = \{a_1, \ldots, a_J\}$ with $a_1 < \ldots < a_J$ and in Appendix B we give an extension to continuous actions.

Preferences. The family of functions $\{u(\theta, \cdot)\}_{\theta \in \Theta}$ is assumed to be an interval dominance ordered (IDO) family (Quah and Strulovici, 2009). This means that for all states $\theta' > \theta$ and actions $a'' > a'$,

$$u(\theta, a'') \geqslant (>)\, u(\theta, a') \quad \Longrightarrow \quad u(\theta', a'') \geqslant (>)\, u(\theta', a') \tag{1}$$

whenever $u(\theta, a'') \geqslant u(\theta, a)$ for all actions a such that $a' \leqslant a \leqslant a''$. Equivalently, if action a'' is the best action in the interval $[a', a''] \cap A$ when the state is θ, then the (weak or strict) preference of a'' over each action in the interval continues to hold at every higher state θ'. As pointed out by Quah and Strulovici (2009), the IDO class includes both single-crossing preferences (Milgrom and Shannon, 1994) and monotone preferences à la Karlin and Rubin (1956).[10]

Experiments and Welfare. Before deciding, the evaluator observes the realization of an experiment: a random vector X in $\mathbb{R}^n$ having state-dependent distribution $G(\cdot|\theta)$ and density $g(\cdot|\theta)$ with monotone likelihood ratio (MLR): if $x' \geqq x$ then $g(x'|\theta)/g(x|\theta)$ is increasing in θ.[11] An important consequence of IDO and MLR is that the evaluator can without loss adopt a monotone strategy, where the action increases in the realization.[12] Thus, the evaluator partitions $\mathbb{R}^n$ into a sequence of sets $(E_1, \ldots, E_J)$ such that, for all j, the set $\bar{E}_j = E_j \cup \cdots \cup E_J$ is an upper set, and chooses a_j when the realization belongs to E_j.[13] The evaluator welfare, $\int_\Theta \sum_j \Pr_\theta(X \in E_j) u(\theta, a_j) \pi(\theta) d\theta$, can then be rewritten, summing by parts and disregarding constants, as

$$U(X) := \int_\Theta \sum_{j<J} \Pr_\theta(X \in \bar{E}_{j+1}) \big[u(\theta, a_{j+1}) - u(\theta, a_j)\big] \pi(\theta) d\theta.$$

Example: Simple Hypothesis Testing. The simplest instance of our setup has two states $\theta_H > \theta_L$ and two actions, rejection a_L and acceptance $a_H > a_L$. The evaluator optimally accepts when $g(x|\theta_H)/g(x|\theta_L) \geqslant r$, where r depends on parameters.[14] In the unidimensional case this strategy takes a familiar form: accept if and only if $x \geqslant \bar{x}$, for some cutoff $\bar{x}$. In general, with $n \geqslant 1$, the acceptance region is an upper set $\bar{E}$. Given this, welfare rewrites (disregarding constants) as

[10] Single-crossing requires (1) to hold even if $u(\theta, a'') < u(\theta, a)$ for some a such that $a' \leqslant a \leqslant a''$. Monotonicity requires (1) only for adjacent actions, that is, $a' = a_j$ and $a'' = a_{j+1}$ for some $j < J$, but in addition requires that the state, say θ_j, where the difference $u(\theta, a_{j+1}) - u(\theta, a_j)$ changes sign, is increasing in j.

[11] Here and in the remainder of the paper, given two vectors $x = (x_1, \ldots, x_n)$ and $x' = (x'_1, \ldots, x'_n)$, we say that x' is larger than x, and write $x' \geqq x$, to indicate that $x'_i \geqslant x_i$ for every i.

[12] By Bayes' rule, MLR implies that the posterior belief on the state increases with the observed realization x in the likelihood ratio order—for all x and $x' \geqq x$ the ratio $\pi(\theta|x')/\pi(\theta|x)$ increases with θ. Thus, the evaluator cannot lose by increasing the action in response to a higher realization (Quah and Strulovici, 2009, Theorem 2).

[13] Recall that $E \subseteq \mathbb{R}^n$ is an *upper set* if it contains every point of $\mathbb{R}^n$ that is larger than some point of E.

[14] The conditional probability of θ_H given that $X = x$ equals $\pi(\theta_H) g(x|\theta_H) / [\pi(\theta_L) g(x|\theta_L) + \pi(\theta_H) g(x|\theta_H)]$, that is, $1/[(\pi(\theta_L)/\pi(\theta_H))(g(x|\theta_L)/g(x|\theta_H)) + 1]$. Thus, the expected payoff difference between acceptance and rejection is nonnegative if and only if $g(x|\theta_H)/g(x|\theta_L) \geqslant r := [\pi(\theta_L)/\pi(\theta_H)][u(\theta_L, a_L) - u(\theta_L, a_H)]/[u(\theta_H, a_H) - u(\theta_H, a_L)]$.

$-r\Pr_{\theta_L}(X \in \bar{E}) - \Pr_{\theta_H}(X \notin \bar{E})$, a negatively weighted sum of the probability of a false positive (accepting in θ_L) and that of a false negative (rejecting in θ_H), with r serving as relative weight.

Selected Experiments. In a typical scenario of statistical decision theory, the evaluator observes a random sample from a univariate distribution $F(\cdot|\theta)$ with density $f(\cdot|\theta)$ satisfying MLR. In this case, $G(x|\theta) = F(x_1|\theta)\cdots F(x_n|\theta)$, and (for fixed sample size n) welfare depends on the family of distributions $F(\cdot|\theta)$ only. In this paper we are interested in experiments involving *selected* rather than random observations. In this scenario, $G(\cdot|\theta)$ generally takes a different form, and welfare is a function of both the family $F(\cdot|\theta)$ and an additional parameter depending on the type of selection we consider. Our main focus is on *maximally selected* experiments, where $X_1, X_2, \ldots, X_n$ are the highest, second highest, $\ldots$, nth highest of $k \geqslant n$ random draws. Thus, the first observation is drawn from distribution $F^k(\cdot|\theta)$, and for $i > 1$, conditional on $X_1 = x_1, \ldots, X_{i-1} = x_{i-1}$, the ith observation is drawn from distribution $F^{k-i+1}(\cdot|\theta)$ right-truncated at x_{i-1}. Letting $< i$ denote the indices $1, \ldots, i-1$ to save on notation, for every x we have

$$G_1(x_1|\theta) = F^k(x_1|\theta) \quad \text{and} \quad G_i(x_i|\theta, x_{<i}) = \frac{F^{k-i+1}(x_i|\theta)}{F^{k-i+1}(x_{i-1}|\theta)} \quad \text{for all } i > 1.^{15} \tag{2}$$

We refer to k as the *presample size* of the experiment, and if $k = n$ we call the experiment *random*, because it is informationally equivalent to n random draws from $F(\cdot|\theta)$.[16] Note that, while it is natural to think of k as a natural number, a maximally selected experiment X is well defined for real presample sizes $k \geqslant n$ as well. Moreover, the distribution of X changes smoothly with k. This technical observation will prove useful in the sequel, when comparing selected experiments with different presample sizes.

Strategic Selection. One motivation for studying selected experiments is that they arise endogenously in strategic settings. In particular, maximal selection is an equilibrium phenomenon when the sample is provided by a strategic *sender* with private information on presample data. Consider the following game: First, the sender privately observes k random draws $x_1, \ldots, x_k$ from $F(\cdot|\theta)$ and chooses a subset $I \subseteq \{1, \ldots, k\}$ of size n. Second, the evaluator observes $(x_i)_{i \in I}$ and chooses an action. Assume the sender's payoff is a strictly increasing function of the evaluator's action. The following immediate observation provides a strategic foundation for maximal selection.[17]

Proposition 0. *For all n and $k \geqslant n$ there is a Bayes Nash equilibrium where the sender always chooses maximal selection. Moreover, this is the unique equilibrium where the sender always selects the same set of order statistics.*[18]

[15] The corresponding joint density is $g(x|\theta) = [k!/(k-n)!]F^{k-n}(x_n|\theta)f(x_1|\theta)\cdots f(x_n|\theta)$. This density satisfies MLR because log-supermodularity is preserved by integration (see e.g. Proposition 4 in Milgrom, 1981) and products of log-supermodular functions are log-supermodular.

[16] Knowing in advance that observations are sorted so that $X_1 \geqslant \cdots \geqslant X_n$ is clearly of no value for the evaluator.

[17] We use Bayes Nash equilibrium because the sender has private information. On the other hand, the sender reports observations in the support, so there is no reason to discuss refinements of off-path beliefs.

[18] There are cases where the sender would wish to reveal the entire presample, which arise when the hidden part

Comparing Multidimensional Experiments by Accuracy. Our main goal in this paper is to assess the welfare impact of selection, be it strategic or originating from any other source. This requires a tool for comparing experiments. In the rest of this section, we develop a natural multidimensional generalization of Lehmann's (1988) notion of accuracy. Our notion can be used to compare any two experiments (not necessarily selected experiments) with the same dimension n.

Let $G(t,\cdot|\theta)$ be a family of state-dependent distributions on $\mathbb{R}^n$ parametrized by $t \in [0,1]$ (and each with MLR density) such that, denoting by $\Pr_\theta(t,\cdot)$ the corresponding measure on $\mathbb{R}^n$, $\Pr_\theta(t,E)$ is continuously differentiable with respect to t for all E. Let $X(t)$ denote the corresponding family of experiments. In our application of accuracy to maximal selection, $X(0)$ and $X(1)$ will be selected experiment with equal sample size n but different presample sizes k and m respectively, while $X(t)$ will denote the experiment with real presample size $tk+(1-t)m$.

For all s,t in $[0,1]$ we define $\varphi_{s,t}(\cdot|\theta) : \mathbb{R}^n \to \mathbb{R}^n$ as follows: $\varphi_{s,t}(x_1,\dots,x_n|\theta) = (z_1,\dots,z_n)$, where $z_1,\dots,z_n$ are defined recursively as

$$z_1 = (G_1(t,\cdot|\theta))^{-1}(G_1(s,x_1|\theta)) \quad \text{and} \quad z_i = (G_i(t,\cdot|\theta,z_{<i}))^{-1}(G_i(s,x_i|\theta,x_{<i})) \quad \text{for } i>1. \quad (3)$$

We say the family $X(t)$ is *ordered by accuracy* if $\varphi_{s,t}(x|\theta') \geqq \varphi_{s,t}(x|\theta)$ whenever $\theta' > \theta$ and $t > s$, for all x. Note that for $n=1$ our definition reduces to Lehmann's (1988).

Theorem 0. *If the family $X(t)$ is ordered by accuracy, then welfare $U(X(t))$ is increasing in t.*

In the unidimensional case this result was proved by Lehmann (1988) for monotone preferences, by Persico (2000) and Jewitt (2007) for single-crossing preferences, and by Quah and Strulovici (2009) for IDO preferences. The latter paper assumes that Θ and A are compact and the distribution of $X(t)$ has the same compact support for all t and in each state. Theorem 0 extends the result to multidimensional experiments and allows unbounded or non-constant supports (as is necessarily the case e.g. in location experiments). We prove the theorem, and discuss further the difference among IDO, single-crossing and monotone preferences, in Appendix B.

The intuition for why accuracy increases welfare is essentially the same as in Lehmann (1988). Consider a simple hypothesis testing problem, and let $\bar{E}$ be any acceptance set the evaluator may choose in experiment $X(s)$. Suppose that as the experiment changes to $X(t)$ the evaluator moves the boundary of the acceptance set, point by point, via the mapping $x \mapsto \varphi_{s,t}(x|\theta_L)$. This means adopting $\varphi_{s,t}(\bar{E}|\theta_L)$ as acceptance set. By definition of the function $\varphi_{s,t}(\cdot|\theta_L)$, in state θ_L the random vector $\varphi_{s,t}(X(s)|\theta_L)$ has the same distribution as $X(t)$. Thus, false positives remain the same: $\Pr_{\theta_L}(X(t) \in \varphi_{s,t}(\bar{E}|\theta_L)) = \Pr_{\theta_L}(X(s) \in \bar{E})$. False negatives, on the other hand, decrease: $\Pr_{\theta_H}(X(t) \in \varphi_{s,t}(\bar{E}|\theta_L)) \geqslant \Pr_{\theta_H}(X(t) \in \varphi_{s,t}(\bar{E}|\theta_H)) = \Pr_{\theta_H}(X(s) \in \bar{E})$. The equality follows from $\varphi_{s,t}(X(s)|\theta_H)$ and $X(t)$ having the same distribution in state θ_H. The inequality, from $\bar{E}$ being an upper set—the image of an upper set under an increasing function is smaller than the set itself.

of the presample contains favorable observations. This unraveling effect is well known in the literature on strategic disclosure, at least since Grossman (1981) and Milgrom (1981). If sample size is exogenously fixed, such unraveling cannot occur, of course. However, as we shall demonstrate below, even when sample size is endogenous the evaluator may want to *commit* not to look at presample data—to strengthen the sender's incentive to collect enough of it.

3 Monotone Impact of Selection

In this section we characterize the families of distributions $F(\cdot|\theta)$ for which the following monotone comparative statics hold: for fixed sample size n, the larger the presample size, the higher (or the lower) the evaluator's welfare.

3.1 Unidimensional Location Experiments

We begin with the simple case of unidimensional location experiments. This case is of independent interest and provides a useful starting point for introducing the characterization in the general case. In a location experiment, the distributions $F(\cdot|\theta)$ are all shifted versions of the same distribution F, that is, $F(x|\theta) = F(x-\theta)$ for all θ and x. In this case, MLR means that F, the *noise distribution*, admits a logconcave density f, and $X = \theta + \varepsilon$, with ε the highest of $k \geqslant 1$ random draws from F.

Theorem 1. *Fixing the sample size to $n = 1$, an increase in presample size increases (decreases) welfare in a location experiment if the reverse hazard function of the noise distribution, $-\log F(\varepsilon)$, is logconcave (logconvex) in ε.*[19]

Proof. Fix two presample sizes k and m, and for every $t \in [0,1]$ denote by $X(t)$ the selected experiment with presample size $k_t := tk + (1-t)m$. The family $X(t)$ is ordered by accuracy if for $s < t$ the function $\varphi_{s,t}(x|\theta) = (F^{k_t})^{-1}\big(F^{k_s}(x-\theta)\big) + \theta$ is increasing in θ for every x. Taking the derivative with respect to θ, we must therefore have

$$k_t F^{k_t-1}(\varphi_{s,t}(x|\theta)-\theta) f(\varphi_{s,t}(x|\theta)-\theta) \geqslant k_s F^{k_s-1}(x-\theta) f(x-\theta). \tag{4}$$

Change variable from x to $u = F^{k_s}(x-\theta)$. Then (4) says that for every $u \in [0,1]$ the slope of F^{k_t} at $\varphi_{s,t}(x|\theta) - \theta = (F^{k_t})^{-1}(u)$ is greater than the slope of F^{k_s} at $x - \theta = (F^{k_s})^{-1}(u)$. Applying the strictly increasing $u \mapsto \lambda(u) = -\log(-\log u)$ to both F^{k_t} and F^{k_s}, this is in turn equivalent to the slope of $\lambda(F^{k_t}(\cdot))$ at $\varphi_{s,t}(x|\theta) - \theta$ being greater than the slope of $\lambda(F^{k_s}(\cdot))$ at $x - \theta$. But the transformations are vertical shifts of each other: $\lambda(F^{k_s}(\cdot)) + \log k_s = \lambda(F(\cdot)) = \lambda(F^{k_t}(\cdot)) + \log k_t$. Thus, (4) holds for $\lambda(F(\cdot))$ convex and $k > m$ (as $k > m$ implies $k_t > k_s$ and hence $\varphi_{s,t}(x|\theta) \geqslant x$) or $\lambda(F(\cdot))$ concave and $m > k$ (as $m > k$ implies $k_s > k_t$ and $\varphi_{s,t}(x|\theta) \leqslant x$). □

The proof of the theorem is illustrated in Figure 2 for a standard normal noise distribution. Since the function $u \mapsto -\log(-\log u)$ is strictly convex, welfare increases (because accuracy increases) as the presample size increases from m to $k > m$.[20]

Accuracy, Dispersion, and Converse Result. Bickel and Lehmann (1979) define a distribution G as *less dispersed* than another distribution F if the quantile difference $G^{-1}(u) - F^{-1}(u)$ is decreasing in u. This notion appeared in our proof of Theorem 1, when we asked whether the slope of F^{k_t}

[19] Marshall and Olkin (2007) define the reverse hazard function as $\log F$. Since F ranges between zero and one, $\log F$ is necessarily negative. Our definition uses a minus sign, so that logconcavity of the function makes sense.

[20] The plots are drawn for $k = 8$ and $m = 1$. Of course, any $k > m \geqslant 1$ give the same qualitative result.

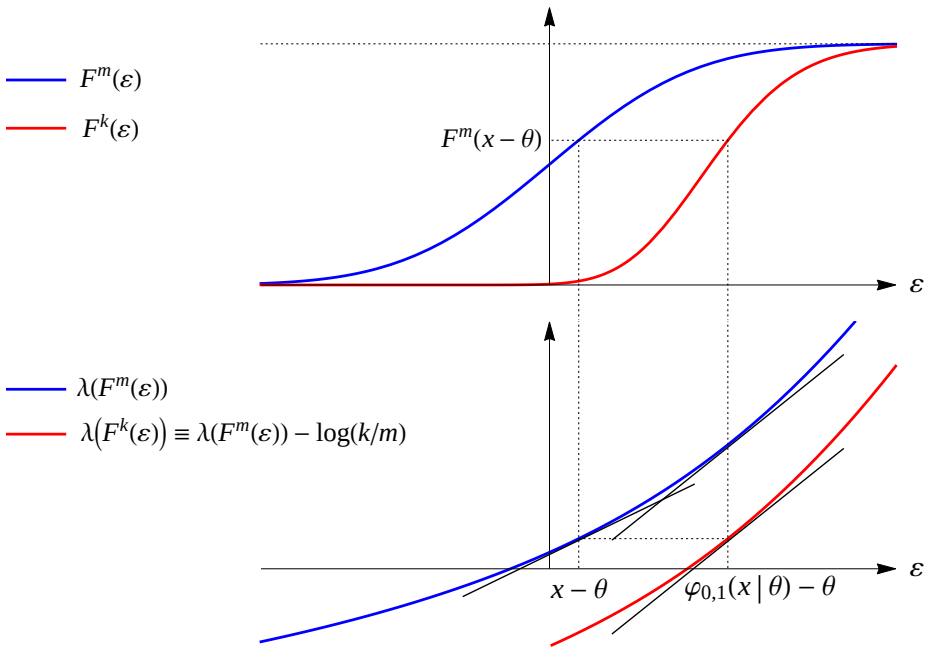

Figure 2: Normal unidimensional location experiment: double-log transformation.

is greater than the slope of F^{k_s} at corresponding quantiles. Applying the double-log transformation to both distributions distills out the effect of k and m on their slopes, revealing the key condition identified in Theorem 1. As shown in Lehmann (1988), for location experiments accuracy and dispersion are equivalent—a family of location experiments is ordered by accuracy if and only if the corresponding noise distributions are inversely ordered by dispersion. Finally, Lehmann (1988) shows that being more accurate is also a *necessary* condition for a unidimensional experiment to give higher welfare in every decision problem with Karlin and Rubin's (1956) monotone payoffs. This statement of course holds also for IDO payoffs. Thus, if we allow presample size to be a real rather than a natural number, the converse of Theorem 1 is true.[21]

Gumbel Noise Distribution. The only noise distribution F such that $-\log F$ is both logconcave and logconvex (i.e. loglinear) is the Gumbel extreme value distribution, $F(\varepsilon) = \exp(-\exp(-\varepsilon))$. In this case every maximally selected experiment gives the evaluator the same welfare. With presample size k the noise distribution is $F^k(\varepsilon) = \exp(-k\exp(-\varepsilon)) = F(\varepsilon - \log k)$. Since selection only inflates noise by a constant ($\log k$), selection has no impact on welfare.

Logistic, Exponential and Shifted Gompertz Noise Distributions. Besides the normal case, one instance where more selection benefits the evaluator is with logistic noise, $F(\varepsilon) = 1/(1+e^{-\varepsilon})$; we prove this and the following claims in footnotes 30 and 31 below. Our main example of the opposite case, where $-\log F$ is logconvex and hence more selection hurts the evaluator, is exponentially

[21] If instead we insist on k being a natural number, then with $k > m$ (resp. $k < m$) the function $\lambda(F(\cdot))$ could be nonconvex (resp. nonconcave) on a short interval, while (4) still holds.

distributed noise, $F(\varepsilon) = 1 - e^{-\varepsilon}$.[22] Finally, another case where more selection hurts is with shifted Gompertz noise, $F(\varepsilon) = (1 - e^{-\varepsilon}\exp(-\eta e^{-\varepsilon}))$.[23]

Contribution to Stochastic Ordering of Order Statistics. Previous results in the literature on stochastic ordering of order statistics only covered noise distributions with decreasing hazard rate. Notably, Khaledi and Kochar (2000, Theorem 2.1) showed that for any distribution with decreasing hazard rate higher order statistics are more dispersed.[24] Given that logconcavity implies increasing hazard rate by Prekopa's theorem, the only noise distribution with logconcave density for which Khaledi and Kochar's (2000) result applies is the exponential (loglinear) distribution, which has constant hazard rate.[25] The novel characterization in Theorem 1 applies more generally to noise distributions with logconcave densities.

3.2 General Multidimensional Experiments

Extending our analysis to general (not necessarily location type) experiments with sample size $n \geqslant 1$ poses two related challenges. First, individual comparisons between order statistics do not provide our desired characterization. For example, in some cases the evaluator *is* better off with maximal selection than with a random experiment, and yet an intermediate order statistic—say, the second or third highest—is not, in isolation, more informative than a random draw.[26] Second, order statistics are correlated, creating further ambiguity about the marginal value of information added by a single order statistic.[27] Multidimensional accuracy allows us to characterize when and how this correlation adds or subtracts value to the evaluator's problem as presample size increases.

Theorem 2. *For fixed sample size $n \geqslant 1$, an increase in presample size increases (decreases) welfare if the reverse hazard rate $f(\cdot|\theta)/F(\cdot|\theta)$ is log-supermodular (log-submodular, with support of $f(\cdot|\theta)$ unbounded above for all θ), that is, if for all states θ and $\theta' > \theta$ the ratio*

$$\frac{f(\cdot|\theta')/F(\cdot|\theta')}{f(\cdot|\theta)/F(\cdot|\theta)}$$

is increasing (resp. decreasing).

[22] More generally, given any $a < -1$, the distribution $F(\varepsilon) = \exp([(1-\exp(-\varepsilon))^{1+a} - 1]/(1+a))$ is such that $-\log F$ is logconvex. The exponential distribution is the special case $a \to -1$.

[23] This distribution was introduced by Bemmaor (1992).

[24] According to Khaledi and Kochar (2000, Theorem 2.1), if the variables X_i are i.i.d. with decreasing hazard rate, then $X_{i:n}$ is less dispersed than $X_{j:m}$ whenever $i \leqslant j$ and $n - i \geqslant m - j$. Setting $i = n = 1$ and $j = m = k$, we have that the maximum of k i.i.d. variables with decreasing hazard rate is more dispersed than the original variable.

[25] Theorem 1 also covers distributions with decreasing hazard rate, where $-\log F$ is necessarily logconvex.

[26] We present an instance of this fact after sketching the proof of Theorem 2.

[27] In the statistical literature on comparisons of multidimensional experiments, Shaked and Tong (1990, 1993) identify conditions under which an experiment with correlated draws is less informative than an i.i.d. experiment, assuming equal marginal distributions. Their results do not apply to our context.

Before discussing the proof of the result, a remark is in order. For location experiments, log-supermodularity of the reverse hazard rate is the same as logconcavity of the noise distribution's reverse hazard rate. This fact immediately implies the following:

Corollary 1. *For fixed sample size $n \geqslant 1$, an increase in presample size increases (decreases) welfare if the reverse hazard rate of the noise distribution, $f(\varepsilon)/F(\varepsilon)$, is logconcave (logconvex, with support of f unbounded above) in ε.*

The hypotheses in Corollary 1, logconcavity or logconvexity of the reverse hazard rate, are stronger than the corresponding conditions in Theorem 1.[28] Still, the corollary applies to all examples discussed earlier. Again, the Gumbel distribution is sandwiched between the noise distributions for which more selection benefits and those for which more selection hurts.[29] More selection benefits with normal or logistic noise, for these distributions have logconcave reverse hazard rates.[30] Finally, more selection hurts with exponential or shifted Gompertz noise.[31]

General Method of Proof. The proof of Theorem 2, like that of Theorem 1, consists in showing that a larger presample size increases (or decreases) accuracy. But the assumptions in Theorem 2 afford us a stronger argument, based on a method that is applicable (and we apply, in Section 5) to other forms of selection and more general comparisons as well. Before sketching a proof of Theorem 2 we thus find it instructive to present the method in its general form. This will also shed further light on our notion of accuracy, and provide a couple of simple ways to check for it.

The method builds on two immediate observations. Consider a family of distributions $G(t,\cdot|\theta)$ and corresponding experiments $X(t)$. Suppose that for each t and θ the variables in $X(t)$ are *associated* in the following sense: for each $i > 1$, conditioning on larger values of $X_{<i}(t)$ induces a first-order stochastic dominance increase in $X_i(t)$. That is, $G_i(t,x_i|\theta,x_{<i})$ is decreasing in $x_{<i}$ for all x_i. Our first observation is that if association holds then $X(t)$ is ordered by accuracy, provided that for all $t > s$ and $\theta' > \theta$, defining $z = \varphi_{s,t}(x|\theta)$ as in (3), we have

$$\frac{G_1(t,z_1|\theta')}{G_1(s,x_1|\theta')} \leqslant 1 \quad \text{and} \quad \frac{G_i(t,z_i|\theta',z_{<i})}{G_i(s,x_i|\theta',x_{<i})} \leqslant 1 \quad \text{for all } i > 1. \tag{5}$$

Our second observation is that (5) must hold as long as for every $i \geqslant 1$ the corresponding ratio is monotonically increasing in x_i. Applying the implicit function theorem to $z = \varphi_{s,t}(x|\theta)$ in

[28] The reverse hazard function is the right-sided integral of the reverse hazard rate: $-\log F(\varepsilon) = \int_{\varepsilon}^{\infty} (f(\varepsilon)/F(\varepsilon))d\varepsilon$. Thus, the reverse hazard function inherits logconcavity (and logconvexity, if the support of f is unbounded above) of the reverse hazard rate (An, 1998, Lemma 3).

[29] In the Gumbel case, $f(\varepsilon)/F(\varepsilon) = \exp(-\varepsilon)$, a loglinear function.

[30] In the normal case, the reciprocal of the reverse hazard rate, $F(\varepsilon)/f(\varepsilon) = \int_{-\infty}^{x} e^{\varepsilon^2/2} e^{-t^2/2} dt = \int_{-\infty}^{0} e^{-u^2/2} e^{-u\varepsilon} du$, is logconvex because $e^{-u\varepsilon}$ is logconvex, and logconvexity is preserved under mixtures (An, 1998, Proposition 3). In the logistic case, the reverse hazard rate is $f(\varepsilon)/F(\varepsilon) = 1/(e^{\varepsilon}+1)$, which is easily seen to be logconcave.

[31] In the generalized exponential case, $f(\varepsilon)/F(\varepsilon) = (1-e^{-\varepsilon})^a e^{-\varepsilon}$, which is logconvex because $a < -1$. In the shifted Gompertz case, the second derivative of $\log(f(\varepsilon)/F(\varepsilon))$ is positive, having the same sign as $e^{3\varepsilon} + \eta(e^{2\varepsilon} - 1)$.

order to compute the derivative of z_i with respect to x_i for each i, we can write this monotonicity requirement more revealingly in terms of reverse hazard rates:

$$\frac{g_i(t,z_i|\theta',z_{<i})/G_i(t,z_i|\theta',z_{<i})}{g_i(t,z_i|\theta,z_{<i})/G_i(t,z_i|\theta,z_{<i})} \geqslant \frac{g_i(s,x_i|\theta',x_{<i})/G_i(s,x_i|\theta',x_{<i})}{g_i(s,x_i|\theta,x_{<i})/G_i(s,x_i|\theta,x_{<i})} \quad \text{for all } i \geqslant 1 \tag{6}$$

(where the conditioning on $x_{<i}$ and $z_{<i}$ is vacuous when $i=1$). We conclude that (6) is a general sufficient condition for any family $X(t)$ satisfying association to be ordered by accuracy. This condition goes a long way in characterizing the impact of selection—maximal and otherwise.

Sketch of Proof of Theorem 2. Consider a selected experiment with presample size k from distribution $F(\cdot|\theta)$. Recall from (2) that the first observation has distribution $F^k(\cdot|\theta)$ and the ith has distribution $F^{k-i+1}(\cdot|\theta)$ right-truncated at x_{i-1}. A simple but crucial fact is that the reverse hazard rate of a distribution is unaffected (except for a multiplicative constant) when we take powers or right-truncate the distribution. In particular, for every $i \geqslant 1$ the reverse hazard rate of the ith observation is $k-i+1$ times the reverse hazard rate of a random draw:

$$\frac{(k-i+1)F^{k-i}(\cdot|\theta)f(\cdot|\theta)}{F^{k-i+1}(\cdot|\theta)} = (k-i+1)\frac{f(\cdot|\theta)}{F(\cdot|\theta)}.$$

But, taking the ratio between reverse hazard rates at different states θ and θ', the multiplicative constant $k-i+1$ disappears. In other words, with k playing the role of t and another presample size $m<k$ playing the role of s, condition (6) depends on t and s only through z. Thus, (6) is simply log-supermodularity of $f(\cdot|\theta)/F(\cdot|\theta)$, because $m<k$ implies $z \geqq x$. Order statistics are associated, so (6) can in fact be used. The argument for the log-submodular case is symmetric.

Exponential Distribution. Corollary 1 shows that maximal selection has a negative impact in location experiments with exponential noise. Remarkably, outside the class of location experiments selection is beneficial when *observations*, rather than noise terms, are drawn from the exponential distribution: $F(x|\theta) = 1-e^{-x/\theta}$ (for $x \geqslant 0$). In this case, for all x and $\theta' > \theta$ we have

$$\frac{f(\cdot|\theta')/F(\cdot|\theta')}{f(\cdot|\theta)/F(\cdot|\theta)} = \frac{\theta\left(e^{x/\theta}-1\right)}{\theta'\left(e^{x/\theta'}-1\right)},$$

which is easily seen to be increasing in x. By Theorem 2, more selection is beneficial in this case.

Intermediate Order Statistics: Unidimensional vs Multidimensional Accuracy. Our notion of accuracy reveals welfare rankings that are not captured by unidimensional comparisons between order statistics. Consider, for instance, a location experiment with *positive* exponential noise: $F(x|\theta) = F(x-\theta) = \exp(x-\theta)$ (for $x \leqslant \theta$). Let X be a random draw, and let Y_1 and Y_2 be the first and second highest of k draws. By Theorem 1, the evaluator is better off with Y_1 than with X, for $-\log F(\varepsilon)) = -\varepsilon$ is strictly logconcave. However, Y_2 and X are not comparable.[32]

[32] Letting $X(0)=X$ and $X(1)=Y_2$, the reciprocal of $\varphi_{0,1}(\cdot|\theta)$ is $\log(k\exp((k-1)(y-\theta)) - (k-1)\exp(k(y-\theta))) + \theta$, which is a bell-shaped function of θ.

But $f(\varepsilon)/F(\varepsilon) = 1$ is loglinear and hence satisfies the logconcavity condition in Corollary 1 (but notably *not* the logconvexity condition, because of the bounded-above support) so the evaluator is in fact better off with experiment (Y_1, Y_2) than with two random draws.

Minimal Selection. Our results have symmetric counterparts for when the evaluator observes the n smallest rather than largest presample draws. This is equivalent to maximal selection in the dual problem where states, actions and realizations have reversed sign. Log-supermodularity (log-submodularity, with support unbounded above) of the reverse hazard rate in the dual problem is equivalent to log-supermodularity (log-submodularity, with support unbounded below) of the hazard rate in the original problem. Thus, in location experiments with sample size $n = 1$ minimal selection increases (decreases) welfare if and only if the hazard function $-\log(1 - F(\varepsilon))$, is logconcave (logconvex). In the general case, minimal selection increases (decreases) welfare if the hazard rate $f(\cdot|\theta)/[1 - F(\cdot|\theta)]$ is log-supermodular (log-submodular, with support of $f(\cdot|\theta)$ unbounded below for every θ). We report these claims as Theorems 1* and 2* in Appendix A.

3.3 Applications of Maximal Selection

Our results offer criteria for comparing the value of experiments with the same sample size n but different presample sizes $k > m$. For example, the evaluator might have to decide between cities with subject pool of different sizes k and m, and believes subjects in each city self-select into the experiment by efficient rationing. Theorem 2 provides immediate guidance on which dataset is better for the evaluator: the first or second according to whether $f(\cdot|\theta)/F(\cdot|\theta)$ is log-supermodular or log-submodular. Equivalently, the same conclusion holds when the two datasets are collected by two independent but strategic researchers (Proposition 0). Next, we discuss other environments where our setup and results apply.

Application: Treatment Effects. In a location experiment maximal selection on the noise terms $\varepsilon_1, \ldots, \varepsilon_k$ is equivalent to maximal selection on the realizations $x_1, \ldots, x_k$.[33] This immediate observation allows us to interpret our model in terms of potential outcomes, following Neyman (1923) and Rubin (1974, 1978). The state θ may represent the homogeneous effect of a treatment, and ε_i an individual's untreated outcome. Given that the untreated outcome distribution F is known, the evaluator does not benefit from requiring a control group in a randomized trial. Suppose now that the experiment is carried out by a strategic researcher who privately observes the untreated outcome of $k > 2n$ individuals, and on this basis (i) selects $2n$ individuals and (ii) assigns n individuals to each treatment. Out of the k presampled individuals, the researcher assigns the n individuals with the highest value of ε to the treatment group, and the n with the lowest ε to the control group (immediate extension of Proposition 0). By Corollary 1, under logconcavity of f/F selection benefits the evaluator directly—the treatment group is more informative than without selection. But it also

[33] In particular, Proposition 0 is robust to a modification of the game where the sender observes (and hence selects sample units based upon) the noise terms $\varepsilon_1, \ldots, \varepsilon_k$ rather than the realizations $x_1, \ldots, x_k$.

benefits indirectly—the untreated outcomes of the untreated individuals are correlated with and hence informative about the counterfactual untreated outcomes of the treated.

Application: Examinee Choice. When testing students, examiners often give examinees the possibility of picking questions from a larger set of questions. Denote a student's ability by θ, and suppose that time permits a test with n questions. The examiner must decide (e.g. as per schoolwide policy) between the following two off-the-shelf test formats: (i) ask n questions at random; (ii) present $k \geqslant n$ questions and commit to grading only n questions picked by the student. From the examiner's perspective, the student's performance in any given question is a random variable with distribution $F(\cdot|\theta)$.[34] The examiner has IDO preferences depending on student's ability and grade assigned to the student, whereas the student's payoff is increasing in the grade. By Theorem 2 and Proposition 0, the examiner prefers format (i) when $f(\cdot|\theta)/F(\cdot|\theta)$ is log-submodular, and format (ii) when $f(\cdot|\theta)/F(\cdot|\theta)$ is log-supermodular.

Design of Selected Experiments: Optimal Presampling. In some settings the evaluator may not only know, but also have control over, both sample size n and presample size k. This is the case, for instance, when a teacher is directly involved in deciding the number of total (k) and required (n) questions in an exam. The determination of the optimal sample size for a given evaluator payoff function and sample size cost—say, linear, with marginal cost c_S—is standard in statistical decision theory; see Berger (1985) for a basic treatment. Our results suggest that selection should be factored in when presampling is possible—say, also at linear cost, with marginal cost c_P.

Denoting the evaluator optimal expected (gross) payoff by $U(k,n)$, the optimal experiment format maximizes $U(k,n) - kc_P - nc_S$. Clearly, $U(k,n)$ increases in n and increases or decreases in k according to Theorems 1 and 2 (paired with Proposition 0 in the strategic case). Thus, under logconvexity of the noise reverse hazard function or log-submodularity of the data reverse hazard rate the evaluator trade-off is trivial: no presampling. As long as $c_P > 0$, at the optimal solution the evaluator must set $k = n$, and the optimal sample size (equal to presample size) solves $\max_{n \geqslant 0}[U(n,n) - nc_S]$, where $n = 0$ corresponds to the no-information payoff.[35] The problem reduces to the standard optimal sample size determination.

When instead selection benefits, the trade-off is nontrivial. The evaluator values sample size but also values sample selection: n and k are two goods. The exact trade-off depends on both c_P and c_S and the specific family of distributions $F(\cdot|\theta)$ under consideration. For an extreme example, in a location experiment with positive exponential noise, $U(k,n) = U(k,1)$ for every k and $n \leqslant k$. The evaluator's posterior belief about the state only depends on the largest observed realization, which for fixed k does not depend on n. Thus, at the optimum, sample size is $n = 1$ and presample size solves $\max_{k \geqslant 1}[U(k,1) - c_S - kc_P]$. Returning to our application to examinee choice, we next illustrate the less extreme trade-off arising with normal noise.

[34] Wainer and Thissen (1994) emphasized the difficulties arising when different examinees' choices generate subsets of questions of unequal difficulty. Our assumption that presample data are i.i.d. assumes away this effect.

[35] That is, the evaluator's expected payoff at the prior, $\max_a \int_\Theta u(\theta,a)\pi(\theta)d\theta$.

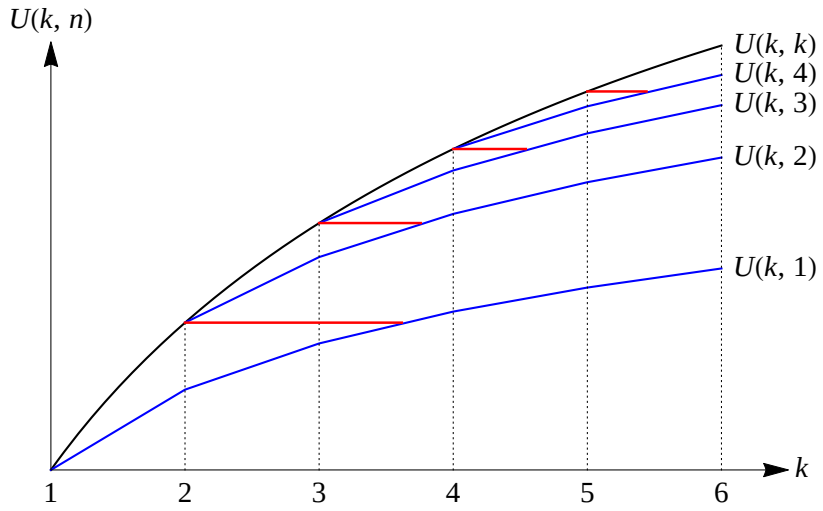

Figure 3: Welfare as a function of presample and sample size: normal location experiment.

Application: Examinee Choice—Continued. Here c_P and c_S represent the marginal costs of preparing and grading questions. Figure 3 illustrates the trade-off with normal noise in the context of simple hypothesis testing—assessing the (binary) ability of a student by means of a pass-fail test. The total number of questions (presample size k) is measured on the horizontal axis. The evaluator gross payoff, normalized so that $U(1,1) = 0$, on the vertical axis. Each blue function corresponds to a fixed number of required questions (sample size n) and is increasing in the total number of questions k, by Theorem 2. Their upper envelope, in black, represents welfare without examinee choice ($k = n$). Note that each red segment is shorter than two: starting from any test format with $k = n \geqslant 2$, the examiner has an incentive to decrease required questions by one, while increasing total questions by two: $U(n+2, n-1) > U(n,n)$ for every $n \geqslant 2$. Thus, no test format with $k = n \geqslant 2$ can be optimal if $c_S > 2c_P$ (or, in a class with N students, if $Nc_S > 2c_P$).

4 Extreme Selection

Complementing the monotone comparative statics results of Section 3, in this section we analyze extreme selection, where presample size grows unbounded. The main result here, Theorem 3, characterizes the corresponding limit welfare. Besides being of independent interest, the theorem will allow us to derive robust propositions for experiment design. For simplicity, here we restrict attention to location experiments.

Our analysis draws on the fundamental result in extreme value theory, which characterizes the limit distribution of the maximum of k i.i.d. random variables, properly normalized for location and scale inflation. Take a noise distribution F and suppose that, for some nondegenerate distribution $\bar{F}$ and some sequence of numbers $\alpha_k > 0$ and β_k, for every continuity point ε of $\bar{F}$ we have

$$F^k(\beta_k + \alpha_k \varepsilon) \to \bar{F}(\varepsilon) \quad \text{as} \quad k \to \infty.$$

The fundamental result of extreme value theory says that $\bar{F}$ must be Gumbel, Extreme Weibull or Frechet.[36] Our maintained assumption that F admits a logconcave density f implies $\bar{F}$ is, in fact, either Gumbel or Extreme Weibull, and always Gumbel if the support of f is unbounded above.[37]

Characterization of Limit Welfare. A larger presample induces a first-order stochastic dominance increase in the noise distribution, hence the location normalization sequence β_k is growing. But the evaluator adjusts for any such inflation without effect on welfare. The limit impact of selection therefore hinges on the behavior of the scale normalization sequence α_k. If this sequence converges to zero, noise becomes more and more concentrated around β_k and the evaluator perfectly learns the state. Otherwise, we can choose a constant sequence $\alpha_k = \alpha$ and an extremely selected experiment is welfare equivalent to a random experiment based on $\bar{F}(\cdot/\alpha)$. The limit behavior of α_k in turn depends on whether the support of f is bounded or unbounded above. In the bounded case, necessarily $\alpha_k \to 0$. An extremely selected observation is arbitrarily close to the upper bound, thus revealing the state. In the unbounded case, the limit behavior of α_k instead further depends on whether the noise distribution satisfies the unbounded hazard rate (UHR) property: $f(\varepsilon)/[1-F(\varepsilon)]$ tends to infinity as ε tends to the upper bound of the support of f.

Theorem 3. *In a unidimensional selected location experiment, as presample size grows without bound welfare converges to the full information payoff if and only if at least one of the following holds: (i) the support of the noise distribution is bounded above; (ii) the noise distribution satisfies UHR. If neither holds, then the limit welfare is the welfare from a unidimensional experiment with Gumbel noise distribution* $\bar{F}(\cdot/\alpha) = \exp(-\exp(-\varepsilon/\alpha))$*, where* $\alpha = \lim_{\varepsilon\to\infty}[1-F(\varepsilon)]/f(\varepsilon)$.[38]

Pairing this result with Theorem 1, we immediately conclude that welfare *monotonically* converges to the full information payoff whenever $-\log F(\varepsilon)$ is logconcave *and* either (or both) of conditions (i) and (ii) in Theorem 3 holds. The hypotheses in the two theorems are overlapping but distinct. For example, UHR holds for normal noise but not for logistic noise. Conditions (i) and (ii) in Theorem 3 cover some distributions with logconcave reverse hazard function, for instance the positive exponential. But the conditions also cover many distributions without logconcave reverse hazard function. For example, in the bounded case, all beta distributions with logconcave density, including uniform. In the unbounded case, beyond normal (or half-normal, which has the same right tails), also every distribution in the exponential power family $f(\varepsilon) = [s/\Gamma(1/s)]\exp(-|\varepsilon|^s)$ with shape parameter $s > 1$. Strikingly, the Laplace distribution ($s = 1$), which has exponential right tails, is the *only* member of this family for which $\alpha_k \not\to 0$. The negative impact of selection with exponential noise discussed earlier becomes fragile under extreme selection—an arbitrarily close exponential power distribution reverses the conclusion; as s approaches 1, however, convergence to full information becomes slower.

[36] See e.g. Leadbetter, Lindgren, and Rootzén (1983) for a primer on extreme value theory.

[37] See Müller and Rufibach (2008).

[38] As we have remarked earlier, logconcavity implies increasing hazard rate, so this limit exists when (ii) is violated.

Experiment Design with Small Presampling Costs. Theorem 3 has immediate implications for experiment design when presampling costs are small. As intuition suggests, under optimal presampling the evaluator exploits selection to economize on sample size. The optimal experiment format is *always* characterized by sample selection under either criterion (i) or (ii) in Theorem 3:

Proposition 1. *In a location experiment design problem with optimal presampling and noise distribution having either support bounded above or the UHR property, for every $c_S > 0$ there exist $\bar{c}_P > 0$ such that if $c_P \leqslant \bar{c}_P$ then every optimal experiment format (k,n) is such that $k > n$.*

The proof of the proposition uses Theorem 3 to show that there is an experiment format $(k,1)$ that dominates the format (n,n) uniformly over all $n \geqslant 1$. Since $c_S > 0$, the design problem under the constraint $k = n$ has an optimal solution $(\bar{n},\bar{n})$. Thus, it suffices to choose k sufficiently large that $(k,1)$ gives higher (gross) payoff than $(\bar{n},\bar{n})$, and then send c_P to zero. Note that Figure 3 provides an illustration of Proposition 1. In that example, simple hypothesis testing with normal noise, whenever $c_P < \bar{c}_P = c_s/2$ the optimal experiment format is always such that $k > n$.

Design of Selected Experiments: Strategic Presampling. In some strategic settings the sender is not only in charge of sample selection, but also has the opportunity to privately choose the presample size. For example, a researcher could choose how much presample data to obtain endogenously, best-replying to a sample size fixed by the evaluator. Consider a game where in the first stage the evaluator chooses a sample size n and decides whether sample selection is allowed or not—we allow this option in order to discuss experiment design. In the second stage, the sender privately chooses $k \geqslant n$ (with $k = n$ if sample selection is not allowed) or opts out of the game. Following these two stages, if the sender has not opted out, the game proceeds as in Proposition 0. If the sender does not opt out, payoffs in state θ are $u(\theta,a) - nc_S^e$ for the evaluator and $v(a) - nc_S^s - kc_P$ for the sender, with $v(\cdot)$ strictly increasing. If the sender opts out, the sender's payoff is zero, while the evaluator gets the no-information payoff.[39] Thus, we again assume that each sample unit costs $c_S > 0$, but here we allow this cost to be split arbitrarily between sender (who pays $c_S^s \geqslant 0$) and evaluator (who pays $c_S^e = c_S - c_S^s \geqslant 0$). In this scenario, the evaluator can again exploit selection to save on sample costs, relying on the fact that the sender will indeed choose a large presample size when this is not too costly—the optimal policy always allows sample selection.

Proposition 2. *In a location experiment design problem with strategic presampling and noise distribution having either support bounded above or the UHR property, for every $c_S > 0$ there exist $\bar{c}_P > 0$ such that if $c_P \leqslant \bar{c}_P$ then it is optimal for the evaluator to allow sample selection.*

The proof of this proposition is similar to that of Proposition 1, but the argument is less immediate. First, we use Theorem 3 to find a selected experiment that dominates all random experiments that are individually rational for both evaluator and sender. Second, we show that, given any presample size conjectured by the evaluator, the sender's expected payoff is increasing and concave

[39] Without loss of generality, we have normalized the sender's payoff so that it is zero when the evaluator chooses the no-information optimal action, $\arg\max_a \int_\Theta u(\theta,a)\pi(\theta)d\theta$.

in the actual presample size chosen by the sender. This fact allows us to conclude that every sufficiently small presample cost c_P induces a sender's optimal presample size that is both large enough and equal to the evaluator's conjectured presample size.

5 Other Forms of Selection

Maximal (or minimal) selection is but one instance of lack of randomness in a statistical sample. In this section we discuss two other forms of selection. For each case, we establish a result using the general method of proof adopted for Theorem 2.

5.1 Truncation

One type of selection that is often relevant involves independent observations from a *truncated distribution*. Here we review this kind of selection and contrast it with the form of selection analyzed earlier. Given a random variable X with distribution $F(\cdot|\theta)$ and density $f(\cdot|\theta)$ satisfying MLR, and given two truncation points $-\infty \leqslant a < b < \infty$, define the left-truncated variables $Y_a := X|X \geqslant a$ and $Y_b := X|X \geqslant b$. Similarly, define the right-truncated variables $W_a := X|X \leqslant a$ and $W_b := X|X \leqslant b$. By variants of the arguments used in the proof of Theorem 2, we obtain:

Theorem 4. *If the hazard rate $f(x|\theta)/[1-F(x|\theta)]$ is log-supermodular, then more left-truncation decreases welfare: $U(Y_b) \leqslant U(Y_a)$. If the reverse hazard rate $f(x|\theta)/F(x|\theta)$ is log-supermodular, then more right-truncation decreases welfare: $U(W_a) \leqslant U(W_b)$.*

This result was known for distributions featuring unbounded likelihood ratio and monotone preferences—see Goel and DeGroot (1992). Our novel proof also covers distributions with bounded likelihood ratio, as in location experiments with logistic noise. Moreover, our result applies more generally to IDO preferences.

The theorem compares unidimensional experiments. The extension of the result to an arbitrary number of independent observations, with exogenous and possibly observation-specific truncation points, is immediate. This is because combining more accurate mutually independent experiments (in this case, more accurate unidimensional experiments) results in a more accurate experiment.[40]

Before discussing the proof of the theorem and its analogy with that of Theorem 2, we contrast the welfare implications of truncation with those of maximal and minimal selection.

Truncation vs Maximal and Minimal Selection. Left-truncation bears a resemblance to maximal selection: with both forms of selection, probability mass is moved toward the upper tail of the distribution. Similarly, right-truncation resembles minimal selection, as both move mass toward the lower tail. However, the theorem shows that as far as accuracy—and hence welfare—is concerned,

[40] More precisely, take two families of experiments $X(t)$ and $X'(t)$, both ordered by accuracy, such that $X(t)$ is independent of $X'(t)$. Then $(X(t), X'(t))$ is also ordered by accuracy.

the right analogy to make is different. More left-truncation (moving from Y_a to Y_b) hurts the evaluator when the hazard rate is log-supermodular, so its impact on welfare is analogous to less minimal selection (Theorem 1*) rather than to more maximal selection. Similarly, more right-truncation (moving from W_b to W_a) hurts the evaluator when the reverse hazard rate is log-supermodular, so its effect is analogous to less maximal selection (Theorem 1). The welfare consequences of the two types of selection are strikingly different. Take a normal or logistic location experiment. Hazard rate and reverse hazard rate of the noise distribution are both logconcave, so more maximal *or* minimal selection benefits the evaluator (Corollary 1). However, more truncation hurts both ways: experiments Y_b and W_a are respectively worse than the less truncated Y_a and W_b.

General Method of Proof—Alternative Version. The proof of the second claim in Theorem 4 is based on the general method we used for Theorem 2. The proof of the first claim is based on an alternative version of that method, which we present here (and also use in the proof of Theorem 5 below). First, note that (5) is identical to

$$\frac{1-G_1(t,z_1|\theta')}{1-G_1(s,x_1|\theta')} \geqslant 1 \quad \text{and} \quad \frac{1-G_i(t,z_i|\theta',z_{<i})}{1-G_i(s,x_i|\theta',x_{<i})} \geqslant 1 \quad \text{for all } i>1. \tag{7}$$

As before, this holds if for all $i \geqslant 1$ the corresponding ratio is monotonically increasing in x_i. But now the implicit function theorem reveals a role for hazard rates, rather than reverse hazard rates:

$$\frac{g_i(t,z_i|\theta',z_{<i})/[1-G_i(t,z_i|\theta',z_{<i})]}{g_i(t,z_i|\theta,z_{<i})/[1-G_i(t,z_i|\theta,z_{<i})]} \leqslant \frac{g_i(s,x_i|\theta',x_{<i})/[1-G_i(s,x_i|\theta',x_{<i})]}{g_i(s,x_i|\theta,x_{<i})/[1-G_i(s,x_i|\theta,x_{<i})]} \quad \text{for all } i \geqslant 1. \tag{8}$$

Condition (8) is another sufficient condition for a family of experiments $X(t)$ satisfying association to be ordered by accuracy. Remarkably, while (7) and (5) are identical, (8) is *not* the same as (6). For example, our main result on maximal selection (Theorem 2), which we proved via (6), cannot be proved using (8). (Its counterpart for minimal selection, Theorem 2*, can be proved via (8).)

Sketch of Proof of Theorem 4. To establish the second claim in the theorem, we use (8) to prove that the family of experiments $Y(t)$, where $Y(t) = X|X \geqslant a_t$ and $a_t = b - t(b-a)$, is ordered by accuracy. Since left-truncation does not affect the hazard rate of a distribution, the first inequality in (8), which is all we need given that the experiments $Y(t)$ are unidimensional, is the following:

$$\frac{f(z|\theta')/\big[1-F(z|\theta')\big]}{f(z|\theta)/\big[1-F(z|\theta)\big]} \leqslant \frac{f(x|\theta')/\big[1-F(x|\theta')\big]}{f(x|\theta)/\big[1-F(x|\theta)\big]}$$

where z is defined by

$$\frac{F(z|\theta)-F(a_t|\theta)}{1-F(a_t|\theta)} = \frac{F(x|\theta)-F(a_s|\theta)}{1-F(a_s|\theta)}.$$

The inequality follows immediately from log-supermodularity of the hazard rate, because $s<t$ implies $z \leqslant x$. The proof of the first part of the theorem is similar: we use (6) to prove that the family of experiments $W(t)$, where $W(t) = X|X \leqslant b_t$ and $b_t = a + t(b-a)$, is ordered by accuracy.

5.2 Median Selection

Finally, we extend our analysis of selection in a new direction, considering central rather than maximal or minimal selection. Call *median selected* an experiment with sample size $n = 1$ where the evaluator observes the rth highest of k random draws from a distribution $F(\cdot|\theta)$, where k is odd and $r = (k+1)/2$. This is the random variable with cumulative distribution function given by

$$\hat{F}(\cdot|\theta) = \sum_{i=r}^{k} \binom{k}{i} F^i(\cdot|\theta)[1 - F(\cdot|\theta)]^{k-i}.$$

Note that median selection can be viewed as a form of sequential selection: first maximal selection of the r largest of the k presample observations, then minimal selection of the smallest among the r maximally selected observations. Our last theorem shows that under monotone preferences à la Karlin and Rubin (1956) median selection turns out to be beneficial precisely when both maximal and minimal selection are beneficial.

Theorem 5. *Assume that preferences are monotone. If the hazard rate $f(\cdot|\theta)/[1 - F(\cdot|\theta)]$ and the reverse hazard rate $f(\cdot|\theta)/F(\cdot|\theta)$ are both log-supermodular, then median selection increases welfare over a random experiment.*

Sketch of Proof of Theorem 5. The proof of the result uses both versions of our general method of proof. Since the median and random experiment are not ranked by first-order stochastic dominance, neither version alone suffices. To grasp intuition, consider once again simple hypothesis testing. In the random experiment the evaluator optimally accepts when the observation exceeds a cutoff $\bar{x}$. Similarly to maximal selection, when $\bar{x}$ is below the median of $F(\cdot|\theta_L)$ the cutoff $\bar{z}$ defined by $\hat{F}(\bar{z}|\theta_L) = F(\bar{x}|\theta_L)$ is larger than $\bar{x}$. This fact, together with log-supermodularity of the reverse hazard rate, guarantees that our sufficient condition (6) for accuracy holds, and hence that (5) holds, that is, $\hat{F}(\bar{z}|\theta_H) \leqslant F(\bar{x}|\theta_H)$. If instead $\bar{x}$ is above the median of $F(\cdot|\theta_L)$, then $\bar{z}$ is smaller than $\bar{x}$. In this case log-supermodularity of the hazard rate guarantees the alternative sufficient condition for accuracy, namely (8).

Application: Peremptory Challenge. Theorem 5 has an immediate application to the analysis of peremptory challenge, a common-law right of the attorneys on each side of a trial to reject a certain number of jurors. Here we consider the problem of a judge who must order a sentence based on the opinion of one juror. The judge does not know the defendant's level of guilt θ, but knows that conditional on θ the jurors' estimates are independently distributed according to $F(\cdot|\theta)$.[41] Assume

[41] There are few formal analyses of peremptory challenge in law and economics. Flanagan (2015) discusses how peremptory challenges necessarily increase the probability of biased juries or affect the expected conviction rate. Schwartz and Schwartz (1996) use a spatial model to highlight the role of peremptory challenge in eliminating jurors with extreme preferences. Earlier analyses of peremptory challenge appear in Brams and Davis (1978) and in Roth, Kadane, and Degroot (1977) and Degroot and Kadane (1980), who analyze optimal strategies for sequential processes of elimination. In all these models, jurors' opinions are uncorrelated with and hence uninformative about guilt, that is, in the language of this paper, $F(\cdot|\theta)$ does not depend on θ.

that both the prosecuting attorney—the one who desires the judge to take higher actions—and the defense attorney—the one who desires the judge to take smaller actions—each has the right to strike down $(k-1)/2$ jurors from an initial set of k jurors. Assume that both attorneys anticipate each juror i's opinion of the defendant's level of guilt. Proposition 0 immediately generalizes to this two-sender setup: as long as the judge adopts a monotone strategy, the prosecuting attorney will strike down the $(k-1)/2$ jurors with the lower opinions, while the defense attorney will strike down the $(k-1)/2$ jurors with the higher opinions.[42] This implies that peremptory challenge leads the judge to decide based on the opinion of the juror with the median opinion. Theorem 5 provides a prior-free criterion to assess whether peremptory challenge provides the judge with more or less accurate information, relative to a randomly chosen juror.

6 Conclusion

Our analysis assumes that the evaluator perfectly predicts the extent of selection, for example because selection is under the evaluator's control, or because the sender's preferences and presampling costs are known. This is the most optimistic scenario when evaluating the impact of selection. Relaxing this assumption, we can consider scenarios where the evaluator may be uncertain about the presample size k, or even fail altogether to anticipate any selection.

Uncertain Selection. In some settings, assuming the evaluator is uncertain about presample size k may be natural. For instance, uncertainty arises with strategic sample selection when the evaluator does not know precisely the sender's preferences. Such uncertainty tends to harm the evaluator,[43] and this can be an important caveat in some contexts. But our main results are only partly overturned by uncertainty. First, our results are robust to small amounts of uncertainty—the evaluator can behave as if k is known, and expected payoffs are continuous in k. Second, when selection strictly benefits an evaluator who perfectly anticipates k, it can also strictly benefit under nontrivial uncertainty about k. For example, in simple hypothesis testing with a normal basic noise distribution, sample size $n = 1$ and equal chance of $k = 1$ and $k = 2$, the evaluator fares strictly better than in a random experiment in the realistic case where the evaluator a priori strongly favors rejection.

Unanticipated Selection. Consider an unwary evaluator who wrongly anticipates a smaller presample size than true. The evaluator is clearly worse off by being unwary than being rational. More interestingly, if a rational evaluator benefits from selection then it is ambiguous whether an unwary evaluator gains or loses when the true presample size is larger than expected. In an important benchmark case, we find that the unwary evaluator is exactly indifferent to an increase of selection from $k = 1$ to $k = 2$. Consider simple hypothesis testing and suppose that the evaluator

[42] This immediately follows from the fact that, given any strategy of the defense (prosecuting) attorney, by eliminating the jurors with the highest (lowest) opinions the prosecuting (defense) attorney induces a first-order stochastic dominance increase (decrease) in the realization observed by the judge.

[43] This is particularly evident in a location experiment with Gumbel noise. In this case anticipated selection leaves the evaluator indifferent, so any uncertainty on k makes the evaluator strictly worse off.

is a priori indifferent between accepting and rejecting. Assume also a noise distribution F symmetric around zero, so that $F(\varepsilon) = 1 - F(-\varepsilon)$. Start from the acceptance cutoff that is optimal in the random experiment, namely $\bar{x} = (\theta_L + \theta_H)/2$, and consider how selection with $k = 2$ affects an unwary evaluator who maintains the acceptance standard unchanged at $\bar{x}$. The probability of acceptance increases in both states, and the resulting change in welfare equals

$$-\pi(\theta_L)\underbrace{\left[F(\bar{x}-\theta_L) - F^2(\bar{x}-\theta_L)\right]}_{\text{increase in false positives}}\left[u(\theta_L,a_1) - u(\theta_L,a_2)\right]$$
$$+\pi(\theta_H)\underbrace{\left[F(\bar{x}-\theta_H) - F^2(\bar{x}-\theta_H)\right)}_{\text{reduction in false negatives}}\left[u(\theta_H,a_2) - u(\theta_H,a_1)\right].$$

By ex ante indifference, $(1-p)[u(\theta_L,a_1) - u(\theta_L,a_2)] = p[u(\theta_H,a_2) - u(\theta_H,a_1)]$. By symmetry, $F(\bar{x}-\theta_L) + F(\bar{x}-\theta_H) = 1$. It follows that the loss due to the increase in false positives exactly offsets the gain due to the reduction in false negatives. The unwary evaluator, who anticipates no selection, is indifferent between no selection and selection with $k = 2$.

A Proofs

Proof of Proposition 0. Assume without loss of generality that the random vector $(X_1,\dots,X_k)$ is ordered so that $X_1 \geqslant \cdots \geqslant X_k$. It suffices to show that given any set of indices $1 \leqslant i_1 < \cdots < i_n \leqslant k$ the random vector $(X_{i_1},\dots,X_{i_n})$ satisfies the MLR property, for this implies that the evaluator's best response is a monotone strategy, and the maximally selected experiment $(X_1,\dots,X_n)$ stochastically dominates $(X_{i_1},\dots,X_{i_n})$ in each state. The density function of $(X_{i_1},\dots,X_{i_n})$ can be written as

$$cF^{k-i_n}(x_{i_n}|\theta)\left[F(x_{i_{n-1}}|\theta) - F(x_n|\theta)\right]^{i_n - i_{n-1}-1} \times \cdots$$
$$\times \left[F(x_{i_1}|\theta) - F(x_{i_2}|\theta)\right]^{i_2-i_1-1}\left[1 - F(x_{i_1}|\theta)\right]^{i_1-1} f(x_{i_1}|\theta)\cdots f(x_{i_n}|\theta),$$

where c is a constant depending on $i_1,\dots,i_n$ and k (see e.g. David and Nagaraja, 2003). Since log-supermodularity is preserved by integration, and products of log-supermodular functions are log-supermodular, the result follows. □

Proof of Theorem 2. Fix two presample sizes k and m, and for every $t \in [0,1]$ denote by $X(t)$ the selected experiment with presample size $k_t := tk + (1-t)m$. Fix $s < t$ and $\theta < \theta'$, and write $\varphi_{s,t}(x|\theta) = z$ and $\varphi_{s,t}(x|\theta') = z'$ for brevity. As a preliminary observation, note that in state θ the support of $X(t)$ is the support of $f(\cdot|\theta)$, and hence does not depend on t. Thus, as x_1 converges to the upper bound of this support, so does z_1. Similarly, for every $i = 2,\dots,n$ and every x_{i-1}, as x_i converges to x_{i-1} (its largest possible value), z_i converges to z_{i-1}. We must prove that under either condition in the theorem ($m \geqslant k$ and the reverse hazard rate is log-supermodular, or $m \leqslant k$ and the reverse hazard rate is log-submodular) for every x we have $z' \geqq z$, or equivalently

$$F^{k_s}(z_1|\theta') \leqslant F^{k_s}(z_1'|\theta') \quad \text{and} \quad F^{k_s-i+1}(z_i|\theta') \leqslant F^{k_s-i+1}(z_i'|\theta') \quad \text{for } i = 2,\dots,n.$$

Plugging the definition of z', we can rewrite these inequalities as

$$F^{k_s}(z_1|\theta') \leqslant F^{k_t}(x_1|\theta') \quad \text{and} \quad \frac{F^{k_s-i+1}(z_i|\theta')}{F^{k_s-i+1}(z'_{i-1}|\theta')} \leqslant \frac{F^{k_t-i+1}(x_i|\theta')}{F^{k_t-i+1}(x_{i-1}|\theta')} \quad \text{for } i = 2,\dots,n.$$

But, for every $i = 2,\dots,n$, if $(z'_1,\dots,z'_{i-1}) \geqq (z_1,\dots,z_{i-1})$ then the denominator of the left-hand side of the second inequality becomes smaller, and hence the left-hand side of the inequality larger, if we replace z'_{i-1} with z_{i-1}. Rearranging terms, we conclude that it suffices to prove that

$$\frac{F^{k_s}(z_1|\theta')}{F^{k_t}(x_1|\theta')} \leqslant 1 \quad \text{and} \quad \frac{F^{k_s-i+1}(z_i|\theta')}{F^{k_t-i+1}(x_i|\theta')} \leqslant \frac{F^{k_s-i+1}(z_{i-1}|\theta')}{F^{k_t-i+1}(x_{i-1}|\theta')} \quad \text{for } i = 2,\dots,n. \tag{9}$$

By the preliminary observation, as x_1 tends to the upper bound of the support of the density associated to $F(\cdot|\theta)$, so does z_1. Thus, under either condition in the theorem ($m \geqslant k$, or $m \leqslant k$ and the support of $F(\cdot|\theta)$ is unbounded above), the left-hand side of the first inequality in (9) tends to a number no greater than one. This implies that the first inequality in (9) holds if the left-hand side of the inequality increases with x_1, that is, differentiating with respect to x_1 and dropping the positive denominator in the derivative,

$$k_s F^{k_s-1}(z_1|\theta') f(z_1|\theta') \frac{dz_1}{dx_1} F^{k_t}(x_1|\theta') \geqslant k_t F^{k_t-1}(x_1|\theta') f(x_1|\theta') F^{k_s}(z_1|\theta'). \tag{10}$$

But, by definition of z,

$$\frac{dz_1}{dx_1} = \frac{k_t F^{k_t-1}(x_1|\theta) f(x_1|\theta)}{k_s F^{k_s-1}(z_1|\theta) f(z_1|\theta)}.$$

Plugging the latter in (10) and simplifying, we conclude that the first inequality in (9) holds if

$$\frac{f(z_1|\theta')/F(z_1|\theta')}{f(z_1|\theta)/F(z_1|\theta)} \geqslant \frac{f(x_1|\theta')/F(x_1|\theta')}{f(x_1|\theta)/F(x_1|\theta)},$$

which in turn follows from log-supermodularity (resp. log-submodularity) of the reverse hazard rate when $m \geqslant k$ (resp. $m \leqslant k$), because $m \geqslant k$ implies $z_1 \geqslant x_1$ (resp. $m \leqslant k$ implies $z_1 \leqslant x_1$).

Again by the preliminary observation, for every $i = 2,\dots,n$ and every x_{i-1}, as x_i converges to x_{i-1}, z_i converges to z_{i-1}. Thus, as before, under either condition in the theorem the left-hand side of the second inequality in (9) tends to a number no greater than the right-hand side. The second inequality in (9) then holds if its left-hand side increases with x_i. Differentiating with respect to x_i and simplifying, as before, we obtain

$$\frac{f(z_i|\theta')/F(z_i|\theta')}{f(z_i|\theta)/F(z_i|\theta)} \geqslant \frac{f(x_i|\theta')/F(x_i|\theta')}{f(x_i|\theta)/F(x_i|\theta)},$$

which again follows from log-supermodularity (resp. log-submodularity) of the reverse hazard rate when $m \geqslant k$ (resp. $m \leqslant k$), because $m \geqslant k$ implies $z_i \geqslant x_i$ (resp. $m \leqslant k$ implies $z_i \leqslant x_i$). □

Minimal Selection. Call *minimally selected* an experiment $X = (X_1, \ldots, X_n)$ where X_1 is the smallest of $k \geqslant n$ random draws from $F(\cdot|\theta)$, X_2 the second smallest, and so on.[44] Consider the dual problem with state space $\tilde{\Theta} = \{-\theta : \theta \in \Theta\}$, action space $\tilde{A} = \{-a : a \in A\}$, payoff function $\tilde{u}(\tilde{\theta}, \tilde{a}) = u(-\tilde{\theta}, -\tilde{a})$, and signal distribution $\tilde{F}(\tilde{x}|\tilde{\theta}) = 1 - F(-\tilde{x}| - \tilde{\theta})$, with density $\tilde{f}(\tilde{x}|\tilde{\theta}) = f(-\tilde{x}| - \tilde{\theta})$. For location experiments, the cumulative noise distribution is therefore $\tilde{F}(\tilde{\varepsilon}) = 1 - F(-\tilde{\varepsilon})$. Having changed sign to both states and actions, the dual problem is also monotone, and having changed sign to both states and signals, the MLR property holds. Finally, action a is optimal given a realization x in the original problem if and only if $-a$ is optimal given realization $-x$ in the dual problem. Since the linear transformation $\varepsilon \mapsto -\varepsilon$ does not change the logconcavity (or logconvexity) of $-\log(\tilde{F}(\cdot))$, this property is equivalent to logconcavity (or logconvexity) of $-\log(1 - F(\cdot))$. Thus, Theorem 1 has the following counterpart:

Theorem 1*. *Fixing the sample size to $n = 1$, an increase in the presample size increases (decreases) welfare if and only if the hazard function of the noise distribution, $-\log(1 - F(\varepsilon))$, is logconcave (logconvex) in ε.*

For general multidimensional experiments, note that the support of $\tilde{f}(\cdot|\tilde{\theta})$ is unbounded above if and only if the support of $f(\cdot| - \tilde{\theta})$ is unbounded below. Moreover, the reverse hazard rate satisfies $\tilde{f}(\tilde{x}|\tilde{\theta})/\tilde{F}(\tilde{x}|\tilde{\theta}) = f(-\tilde{x}| - \tilde{\theta})/[1 - F(-\tilde{x}| - \tilde{\theta})]$, and the switch of sign in both arguments does not affect the log-supermodularity (or log-submodularity) of the function. Thus, Theorem 2 has the following counterpart:

Theorem 2*. *For a fixed sample size $n \geqslant 1$, an increase in the presample increases (decreases) welfare if the hazard rate $f(\cdot|\theta)/[1 - F(\cdot|\theta)]$ is log-supermodular (log-submodular, with support of $f(\cdot|\theta)$ unbounded below for every θ), that is, if for all states θ and $\theta' > \theta$ the hazard rate ratio*

$$\frac{f(\cdot|\theta')/[1 - F(\cdot|\theta')]}{f(\cdot|\theta)/[1 - F(\cdot|\theta)]}$$

is increasing (resp. decreasing).

Proof of Theorem 3. We start by showing that under either condition (i) or (ii) in the theorem we have $\alpha_k \to 0$ as $k \to \infty$. Let $\bar{\varepsilon}$ denote the upper bound of the support of f. If $\bar{\varepsilon} < \infty$ then, as shown in Müller and Rufibach (2008), the limiting distribution is either extreme Weibull or Gumbel. In the first case, we can take $\alpha_k = \bar{\varepsilon} - F^{-1}(1 - 1/k)$ by Proposition 1.13 in Resnick (2008), whence $\alpha_k \to 0$ follows. In the second case, or, by Lemma 3.5 in Müller and Rufibach (2008), if $\bar{\varepsilon} = \infty$, the limiting distribution is Gumbel. It follows from Proposition 1.9 in Resnick (2008) that we can take α_k to be the mean residual life evaluated at $\varepsilon_k := F^{-1}(1 - 1/k)$, that is, $\alpha_k = k \int_{\varepsilon_k}^{\bar{\varepsilon}} \varepsilon f(\varepsilon) d\varepsilon$. As shown in Calabria and Pulcini (1987) and Bradley and Gupta (2003), the limiting behavior of

[44] Thus, in each state θ the support of X is contained in $\{x \in \mathbb{R}^n : x_1 \leqslant \cdots \leqslant x_n\}$, the conditional cumulative distribution function of X_1 is $1 - [1 - F(\cdot|\theta)]^k$, and for every $i = 2, \ldots, n$ the cumulative distribution function of X_i given that $X_1 = x_1, \ldots, X_{i-1} = x_{i-1}$ is $1 - ([1 - F(x_i|\theta)]/[1 - F(x_{i-1}|\theta)])^{k-i+1}$.

the mean residual life is the same as the limiting behavior of the inverse of the hazard rate.[45] Thus, using the fact that $\varepsilon_k \to \bar{\varepsilon}$ as $k \to \infty$, we again obtain $\lim_{k\to\infty} \alpha_k = \lim_{\varepsilon\to\bar{\varepsilon}}[1-F(\varepsilon)]/f(\varepsilon) = 0$.

We now show that if $\alpha_k \to 0$ then the evaluator's payoff converges to the full information payoff, $\overline{U} := \int_\Theta \max_a u(\theta,a)\pi(\theta)d\theta$, as $k \to \infty$. Recall that, by IDO, for every $1 \leqslant j < J$ there exists a state θ_j such that $u(\theta,a_j) - u(\theta,a_{j+1})$ is nonnegative for $\theta \leqslant \theta_j$ and nonpositive for $\theta \geqslant \theta_j$. As we noted in the proof of Theorem 0, one consequence of this observation is that if $\theta_j < \theta_{j-1}$ then action a_j can be removed from A without affecting the IDO property. Another consequence is that action a_j is never optimal at any state, and hence it is never used under full information. Thus, we may assume without loss of generality that $\theta_j \geqslant \theta_{j-1}$ for all $j > 1$. The full information payoff can then be written, summing by parts, as

$$\overline{U} = \int_\Theta \sum_{j<J} \mathbf{1}_{\{\theta\geqslant\theta_j\}} \big[u(\theta,a_{j+1}) - u(\theta,a_j)\big]\pi(\theta)d\theta.$$

Fix $\delta > 0$, and let $\eta > 0$ be such that

$$\int_\Theta \sum_{j<J} \mathbf{1}_{\{\theta_j-\eta\leqslant\theta<\theta_j\}} \big[u(\theta,a_{j+1}) - u(\theta,a_j)\big]\pi(\theta)d\theta \leqslant \frac{\delta}{2} \tag{11}$$

and, furthermore,

$$(1-\eta)\int_\Theta \sum_{j<J} \mathbf{1}_{\{\theta\geqslant\theta_j\}} \big[u(\theta,a_{j+1}) - u(\theta,a_j)\big]\pi(\theta)d\theta \geqslant \overline{U} - \frac{\delta}{2}. \tag{12}$$

Let $\bar{\varepsilon} > 0$ be such that $\hat{F}(\bar{\varepsilon}) - \hat{F}(-\bar{\varepsilon}) \geqslant 1-\eta/2$, and choose $\hat{k}$ so that, for all $k \geqslant \hat{k}$,

$$\alpha_k\bar{\varepsilon} < \eta, \qquad F^k(\alpha_k\bar{\varepsilon}+\beta_k) \geqslant \hat{F}(\bar{\varepsilon}) - \frac{\eta}{4}, \qquad \text{and} \qquad F^k(\alpha_k\bar{\varepsilon}+\beta_k) \leqslant \hat{F}(-\bar{\varepsilon}) + \frac{\eta}{4}.$$

Then, for each θ,

$$\begin{aligned}\Pr_\theta(\theta-\eta+\beta_k \leqslant X \leqslant \theta+\eta+\beta_k) &\geqslant \Pr_\theta(\theta-\alpha_k\bar{\varepsilon}+\beta_k \leqslant X \leqslant \theta+\alpha_k\bar{\varepsilon}+\beta_k)\\ &= F^k(\alpha_k\bar{\varepsilon}+\beta_k) - F^k(\alpha_k\bar{\varepsilon}+\beta_k)\\ &\geqslant \hat{F}(\bar{\varepsilon}) - \frac{\eta}{4} - \hat{F}(-\bar{\varepsilon}) - \frac{\eta}{4} \geqslant 1-\eta,\end{aligned} \tag{13}$$

so the distribution of X in state θ assigns at least probability $1-\eta$ to an η-neighborhood of $\theta+\beta_k$. Now consider the following strategy for the evaluator: choose a_1 if $X < \theta_1+\beta_k-\eta$, choose a_J if $X \geqslant \theta_{J-1}+\beta_k-\eta$, and for every $1 < j < J$, choose a_j if $\theta_{j-1}+\beta_k-\eta \leqslant X < \theta_j+\beta_k-\eta$. The corresponding payoff, again using summation by parts, is

$$\int_\Theta \sum_{j<J} \Pr_\theta(X \geqslant \theta_j+\beta_k-\eta)\big[u(\theta,a_{j+1}) - u(\theta,a_j)\big]\pi(\theta)d\theta.$$

[45] The latter two papers assume that the support of f is bounded below. However, for every x such that $0 < F(x) < 1$ the hazard rate of distribution F is the same as the hazard rate of the left-truncated distribution $F(\cdot)/[1-F(x)]$. Furthermore, the two distributions have the same right tails and hence the same limiting distribution $\bar{F}$.

By (11), (12) and (13), this payoff is at least as large as $\overline{U}-\delta$.

Finally, we show that when both conditions (i) and (ii) in the theorem fail, the limit welfare is the welfare from an experiment with Gumbel noise. This is immediate, because (as we have already argued earlier) violation of (i) implies that $\bar{F}$ is the Gumbel distribution $\exp(-\exp(-\varepsilon))$ and furthermore that $\lim_{k\to\infty}\alpha_k=\lim_{\varepsilon\to\bar{\varepsilon}}[1-F(\varepsilon)]/f(\varepsilon)=:\alpha$, whereas violation of (ii) implies $\alpha>0$. But then $F^k(\beta_k+\alpha\varepsilon)\to\bar{F}(\varepsilon)$, which means that (after the location normalization β_k) the noise distribution converges weakly to $\exp(-\exp(-\varepsilon/\alpha))$. □

Proof of Proposition 1. Let $\bar{n}=\arg\max_{n\geqslant 1}U(n,n)-(n-1)c_S$. Note that $\bar{n}$ exists because $U(n,n)$ is bounded above by the evaluator's full information payoff, $\overline{U}$, while $c_S>0$. Furthermore, $U(\bar{n},\bar{n})-(\bar{n}-1)c_S<\overline{U}$. By Theorem 3, $U(k,1)\to\overline{U}$ as $k\to\infty$. Thus, there exist $\bar{k}>1$ and $\delta>0$ such that

$$U(\bar{k},1)-c_S-\delta>U(n,n)-nc_S\qquad\text{for all } n\geqslant 1. \tag{14}$$

Letting $\bar{c}_P=\delta/\bar{k}$, we obtain the result. □

Proof of Proposition 2. Let $V(k,n)$ denote the sender's expected payoff with presample size k and sample size n. We start by constructing a sample size $\bar{n}$ such that if the evaluator does not allow sample selection, then in equilibrium the evaluator's payoff is no greater than $U(\bar{n},\bar{n})-\bar{n}c_S^e$. Let

$$\bar{n}^e=\arg\max_{n\geqslant 1}U(n,n)-nc_S^e\quad\text{ and }\quad\bar{n}^r=\max\left\{n\geqslant 1: V(n,n)-nc_S^r\geqslant 0\right\}$$

and note that either $\bar{n}^e$ or $\bar{n}^r$ must be finite, for $c_S=c_S^e+c_S^r>0$ whereas $U(n,n)$ and $V(n,n)$ are bounded above by $\bar{U}$ and $v(a_J)$, respectively. If both $\bar{n}^e<\infty$ and $\bar{n}^r<\infty$, let $\bar{n}=\max\{\bar{n}^e,\bar{n}^r\}$. Otherwise, let $\bar{n}=\min\{\bar{n}^e,\bar{n}^r\}$. Note that, in any case,

$$V(n,n)-nc_S^r\geqslant 0\quad\Longrightarrow\quad U(n,n)-nc_S^e\leqslant U(\bar{n},\bar{n})-\bar{n}c_S^e,\qquad\text{for all } n\geqslant 1.$$

Next, we construct a presample size $\bar{k}$ and a presample cost $\bar{c}_P$ such that, if $c_P\leqslant\bar{c}_P$ and the evaluator allows sample selection, then there exists an equilibrium with presample size $k\geqslant\bar{k}$ where the evaluator's payoff, $U(k,1)-c_S^e$, is strictly larger than $U(\bar{n},\bar{n})-\bar{n}c_S^e$, thus concluding the proof.

By Theorem 3, there exists $\bar{k}>1$ such that

$$U(k,1)-c_S>U(\bar{n},\bar{n})-\bar{n}c_S^e\qquad\text{for all } k\geqslant\bar{k}. \tag{15}$$

Now for each $k\geqslant 1$ define $c(k)=-\int_\Theta\sum_{j<J}\log(F(\bar{x}_j(k)-\theta))F^k(\bar{x}_j(k)-\theta)\pi(\theta)d\theta$, where $\bar{x}_1(k)$, $\ldots$, $\bar{x}_{J-1}(k)$ are the cutoffs defining the evaluator's optimal strategy when the sample size is $n=1$ and the presample size is k. Consider the set $C=\left\{c(k)\ :\ k\geqslant\bar{k}\right\}$. Since $c(k)\to 0$ as $k\to\infty$, we have $C=(0,\bar{c}_P]$ for some $\bar{c}_P>0$. Given any presample cost $c_P\in C$ and any $\hat{k}\geqslant\bar{k}$ such that $c(\hat{k})=c_P$, the sender's best response to $\bar{x}_2(\hat{k}),\ldots,\bar{x}_J(\hat{k})$ when sample selection is allowed is the optimal solution to

$$\max_{k\geqslant 1}\int_\Theta\sum_{j<J}\left[1-F^k(\bar{x}_{j+1}(\hat{k})-\theta)\right]\left[v(a_{j+1})-v(a_j)\right]\pi(\theta)d\theta-kc_P,$$

where we used summation by parts and disregarded constants to rewrite the sender's payoff. The objective function is concave, since $v(a_{j+1}) - v(a_j) \geqslant 0$ and $d^2\left[1 - F^k(\bar{x}_{j+1}(\hat{k}) - \theta)\right]/dk^2 = -[\log(F(\bar{x}_{j+1}(\hat{k}) - \theta))]^2 F^k(\bar{x}_{j+1}(\hat{k}) - \theta) < 0$ for every θ. Since $\hat{k}$ satisfies the first order condition, namely $c(\hat{k}) = c_P$, we are done. □

Proof of Theorem 4. Consider first the family of experiments $Y(t)$, where $Y(t) = X|X \geqslant a_t$ and $a_t = b - t(b-a)$. In each state θ the distribution of $Y(t)$ is $[F(y|\theta) - F(a_t|\theta)]/[1 - F(a_t|\theta)]$, for $y \geqslant a_t$. Now fix $s < t$ and consider the function $\varphi_{s,t}(\cdot|\theta)$, which is defined as follows:

$$[F(y|\theta) - F(a_s|\theta)]/[1 - F(a_s|\theta)] = [F(\varphi_{s,t}(y|\theta)|\theta) - F(a_t|\theta)]/[1 - F(a_t|\theta)].$$

for $y \geqslant a_s$. We must show that if $\theta' > \theta$ then $\varphi_{s,t}(y|\theta') \geqslant \varphi_{s,t}(y|\theta)$ for every $y \geqslant a_s$, and using the definition of $\varphi_{s,t}(\cdot|\theta')$ it suffices to show that

$$[F(y|\theta') - F(a_s|\theta')]/[1 - F(a_s|\theta')] \geqslant [F(\varphi_{s,t}(y|\theta)|\theta') - F(a_t|\theta')]/[1 - F(a_t|\theta')].$$

This inequality holds in the limit as y decreases to the lower bound a_s, as both sides converge to one, so we must prove that the ratio between right-hand and left-hand side decreases with y, or

$$\frac{f(y|\theta')/[1 - F(y|\theta')]}{f(y|\theta)/[1 - F(y|\theta)]} \geqslant \frac{f(\varphi_{s,t}(y|\theta)|\theta')/[1 - F(\varphi_{s,t}(y|\theta)|\theta')]}{f(\varphi_{s,t}(y|\theta)|\theta)/[1 - F(\varphi_{s,t}(y|\theta)|\theta)]}.$$

The latter inequality holds when the hazard rate is log-supermodular, given that $\varphi_{s,t}(y|\theta) \leqslant y$ by the fact that $Y(s)$ first-order stochastically dominates $Y(t)$.

Next, consider the family of experiments $W(t)$, where $W(t) = X|X \leqslant b_t$ and $b_t = a + t(b-a)$. In state θ the distribution of $W(t)$ is $F(w|\theta)/F(b_t|\theta)$, for $w \leqslant b_t$. Fix $s < t$ and consider the function $\varphi_{s,t}(\cdot|\theta)$ defined as follows: for every $w \leqslant b_s$,

$$F(w|\theta)/F(b_s|\theta) = F(\varphi_{s,t}(w|\theta)|\theta)/F(b_t|\theta).$$

We must show that if $\theta' > \theta$ then $\varphi_{s,t}(w|\theta') \geqslant \varphi_{s,t}(w|\theta)$ for all $w \leqslant b_s$. But, by definition of $\varphi_{s,t}(\cdot|\theta')$, we have $F(\varphi_{s,t}(w|\theta')|\theta')/F(b_t|\theta') = F(w|\theta')/F(b_s|\theta')$, so it suffices to show that

$$F(w|\theta')/F(b_s|\theta') \geqslant F(\varphi_{s,t}(w|\theta)|\theta')/F(b_t|\theta').$$

The inequality holds (with equality) in the limit as w increases to the upper bound b_s, because both sides converge to one. Thus, all we need to prove is that the ratio between the right-hand side and the left-hand side of the inequality increases with w. Taking derivatives, this condition says that

$$\frac{f(w|\theta')/F(w|\theta')}{f(w|\theta)/F(w|\theta)} \leqslant \frac{f(\varphi_{s,t}(w|\theta)|\theta')/F(\varphi_{s,t}(w|\theta)|\theta')}{f(\varphi_{s,t}(w|\theta)|\theta)/F(\varphi_{s,t}(w|\theta)|\theta)}.$$

This holds when the reverse hazard rate is log-supermodular, given that $\varphi_{s,t}(w|\theta) \geqslant w$ by the fact that $W(t)$ first-order stochastically dominates $W(s)$. □

Proof of Theorem 5. Since payoffs satisfy Karlin and Rubin's (1956) monotonicity, Theorem 0 holds for families of experiments $X(t)$ where t is an index from an arbitrary ordered set T, as shown in Appendix B. Thus, in order to prove the theorem it suffices to take $T = \{0,1\}$, with $X(0)$ the random experiment and $X(1)$ the median experiment. The density function of $X(1)$ is

$$cF^{r-1}(\cdot|\theta)[1-F(\cdot|\theta)]^{r-1}f(\cdot|\theta),$$

where c depends only on k. The cumulative distribution and survival functions can be written as

$$F^r(\cdot|\theta)\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}[1-F(\cdot|\theta)]^j \qquad \text{and} \qquad [1-F(\cdot|\theta)]^r\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}F^j(\cdot|\theta),$$

respectively. For each θ the function $\varphi_{0,1}(x|\theta)$ is defined by the equality

$$F(x|\theta) = F^r(\varphi_{0,1}(x|\theta)|\theta)\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}[1-F(\varphi_{0,1}(x|\theta)|\theta)]^j. \tag{16}$$

Now fix two states $\theta' > \theta$ and let $z = \varphi_{0,1}(x|\theta)$ and $z' = \varphi_{0,1}(x|\theta')$ for brevity. Let x_m denote the median of $F(\cdot|\theta)$, that is, $F(x_m|\theta) = 1/2$. Note that $z \geqslant x$ when $x \leqslant x_m$ and $z \leqslant x$ when $x \geqslant x_m$. Moreover, note that

$$\frac{dz}{dx} = \frac{f(x|\theta)}{cF^k(z|\theta)[1-F(z|\theta)]^k f(z|\theta)}. \tag{17}$$

We must show that $z' \geqslant z$ or, equivalently that

$$\frac{F^r(z|\theta')\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}[1-F(z|\theta')]^j}{F(x|\theta')} \leqslant 1, \tag{18}$$

which is the same as

$$\frac{[1-F(z|\theta')]^r\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}F^j(z|\theta')}{1-F(x|\theta')} \geqslant 1. \tag{19}$$

Suppose first that $x \leqslant x_m$, so that $z \geqslant x$. Since $F(\cdot|\theta')$ first-order stochastically dominates $F(\cdot|\theta)$, condition (18) holds at $x = x_m = z$. Thus, it suffices to show that the left-hand side of (18) increases in x when $x \leqslant z$. The derivative of the left-hand side is nonnegative if and only if

$$cF^{r-1}(z|\theta')[1-F(z|\theta')]^{r-1}f(z|\theta')\frac{dz}{dx}F(x|\theta') - F^r(z|\theta')\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}(1-F(z|\theta'))^j f(x|\theta') \geqslant 0$$

Plugging in (17) and using (16), the latter inequality is the same as

$$\frac{f(z|\theta')/F(z|\theta')}{f(z|\theta)/F(z|\theta)} \times \underbrace{\frac{[1-F(z|\theta')]^{r-1}/\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}[1-F(z|\theta')]^j}{[1-F(z|\theta)]^{r-1}/\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}[1-F(z|\theta')]^j}}_{\geqslant 1 \text{ because } F(\cdot|\theta) \geqslant F(\cdot|\theta')} \geqslant \frac{f(x|\theta')/F(x|\theta')}{f(x|\theta)/F(x|\theta)}$$

which is true by log-supermodularity of the reverse hazard rate, since $z \geqslant x$.

Suppose now that $x \geqslant x_m$, so that $z \leqslant x$. Since (19) is the same as (18), it holds at $x = x_m = z$, so it suffices to show that its left-hand side increases in x when $x \geqslant z$. The derivative of the left-hand side of (19) is nonnegative if and only if

$$-cF^{r-1}(z|\theta')[1-F(z|\theta')]^{r-1}f(z|\theta')\frac{dz}{dx}[1-F(z|\theta')]^{r}\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}F^{j}(z|\theta')f(x|\theta') \geqslant 0$$

Plugging in (17) and using (16), the latter inequality is the same as

$$\frac{f(z|\theta')/[1-F(z|\theta')]}{f(z|\theta)/[1-F(z|\theta)]} \times \underbrace{\frac{F^{r-1}(z|\theta')/\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}F^{j}(z|\theta')}{F^{r-1}(z|\theta)/\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}F^{j}(z|\theta)}}_{\leqslant 1 \text{ because } F(\cdot|\theta) \geqslant F(\cdot|\theta')} \leqslant \frac{f(x|\theta')/[1-F(x|\theta')]}{f(x|\theta)/[1-F(x|\theta)]}$$

which is true by log-supermodularity of the hazard rate, since $z \leqslant x$. □

B Accuracy and Welfare

In this appendix we prove Theorem 0 and provide an extension of the result to the continuous-action case. The case of preferences satisfying Karlin and Rubin's (1956) monotonicity affords us a much simpler argument, so we find it instructive to start with an independent proof for this case. After discussing the difficulty with single-crossing and IDO preferences, we provide a proof for the general IDO case.

B.1 Monotone Preferences

Recall that preferences are monotonic in the sense of Karlin and Rubin (1956) if there exist states $\theta_1 \leqslant \cdots \leqslant \theta_{J-1}$ such that, for every $j < J$, the difference $u(\theta, a_{j+1}) - u(\theta, a_j)$ is nonpositive for $\theta \leqslant \theta_j$ and nonnegative for $\theta \geqslant \theta_j$.

Proof of Theorem 0—Monotone Preferences. Let $X(t)$ be a family of experiments ordered by accuracy. Fix $s < t$, let $(E_1(s), \ldots, E_J(s))$ be the evaluator's optimal partition of $\mathbb{R}^n$ for experiment $X(s)$, and $\bar{E}_j(s) := E_j(s) \cup \cdots E_J(s)$. The evaluator's welfare is

$$\int_{\Theta} \sum_{j<J} \Pr_\theta\big(X(s) \in \bar{E}_{j+1}(s)\big)\big[u(\theta, a_{j+1}) - u(\theta, a_j)\big]\pi(\theta)d\theta.$$

To prove the result it suffices to exhibit nested upper sets $\bar{E}'_2 \supseteq \cdots \supseteq \bar{E}'_J$ such that, for every $j < J$ and every state θ, the difference

$$\Pr_\theta(X(t) \in \bar{E}'_{j+1}) - \Pr_\theta(X(s) \in \bar{E}_{j+1}(s)) \tag{20}$$

is nonpositive for $\theta \leqslant \theta_j$ and nonnegative for $\theta > \theta_j$. Indeed, this implies that the evaluator can achieve a higher expected payoff in experiment $X(t)$ by adopting the following strategy: choose a_1 when $X(t) \notin \bar{E}'_2$, choose a_2 when $X(t) \in \bar{E}'_2 \setminus \bar{E}'_3$, and so on. Define $\bar{E}'_{j+1} = \varphi_{s,t}(\bar{E}_{j+1}|\theta_j)$ for every $j < J$. Then we can rewrite the difference in (20) as

$$\Pr_\theta(X(t) \in \varphi_{s,t}(\bar{E}_{j+1}|\theta_j)) - \Pr_\theta(X(t) \in \varphi_{s,t}(\bar{E}_{j+1}|\theta)).$$

For $\theta \leqslant \theta_j$ the difference is nonpositive, because $\varphi_{s,t}(\cdot|\theta) \leqslant \varphi_{s,t}(\cdot|\theta_j)$ in this case. For $\theta > \theta_j$ it is nonnegative, because then $\varphi_{s,t}(\cdot|\theta) \geqslant \varphi_{s,t}(\cdot|\theta_j)$. □

Note that the above proof does not use the fact that t is a continuous parameter. The family of experiments $X(t)$ could be indexed in an arbitrary ordered set T rather than the interval $[0,1]$. This fact is used in the proof of Theorem 5.

B.2 IDO Preferences

The above proof does not extend immediately to IDO preferences or even only single-crossing preferences. The IDO property does imply that the difference $u(\theta,a_{j+1}) - u(\theta,a_j)$ exhibits single crossing, but does not require the crossing points θ_j to be increasing in j. This makes the upper sets $\bar{D}'_2,\dots,\bar{D}'_J$ non-nested and hence the proposed strategy for experiment $X(t)$ ill-defined.

To deal with this difficulty, we adopt a different strategy of proof, similar in spirit to the argument used by Jewitt (2007) for single-crossing preferences and unidimensional experiments. Our proof hinges on a crucial observation: any action a_j such that the crossing points θ_j and θ_{j-1} are not ordered in Karlin and Rubin's (1956) sense (i.e. such that $\theta_j < \theta_{j-1}$) can be removed from the action set without affecting IDO. In particular, we can remove any such action that, in addition, is not used under the optimal strategy, without affecting the evaluator's welfare, either.

Proof of Theorem 0—IDO preferences. Let $\{D_1(t),\dots,D_J(t)\}$ be the optimal partition of $\mathbb{R}^n$ for experiment $X(t)$, with $D_j(t) \cup \dots \cup D_J(t)$ an upper set and action a_j chosen when $X(t) \in D_j(t)$. Let $\Pr_\theta(t,\cdot)$ denote the measure on $\mathbb{R}^n$ induced by $X(t)$, and define $E_j(t) = \mathbb{R}^n \setminus (D_{j+1}(t) \cup \dots \cup D_J(t))$ for all t and $j < J$. Then the evaluator's welfare is

$$U(X(t)) = \int_\Theta \sum_{j<J} \left[1 - \Pr_\theta(t,E_j(t))\right]\left[u(\theta,a_{j+1}) - u(\theta,a_j)\right]\pi(\theta)d\theta.$$

Now take any t and $u > t$ in $[0,1]$. Applying Theorem 2 in Milgrom and Segal (2002), we obtain

$$U(X(u)) - U(X(t)) = \int_t^u \int_\Theta \sum_{j<J} \frac{\partial \Pr_\theta(s,E_j(s))}{\partial t}\left[u(\theta,a_j) - u(\theta,a_{j+1})\right]\pi(\theta)d\theta ds, \qquad (21)$$

and we have to show that the expression in (21) is nonnegative. We do this in four steps.

Step 1—Use IDO to rewrite the payoff difference. We start by rewriting, for each s, the summation inside the integral in (21), as follows. Recall that, by IDO, for every $1 \leqslant j < J$ there exists

a state θ_j such that the difference $u(\theta,a_j)-u(\theta,a_{j+1})$ is nonnegative for $\theta \leqslant \theta_j$ and nonpositive for $\theta \geqslant \theta_j$. An immediate consequence of this observation is that if $\theta_j < \theta_{j-1}$ then action a_j can be removed from A without affecting the IDO property.[46] By using this fact (repeatedly, if necessary) together with the fact that $E_j(s) = E_{j-1}(s)$ whenever $D_j(s) = \varnothing$, we conclude that for every s there exists a list of indices $1 \leqslant j(s,1) < \cdots < j(s,I_s) \leqslant J$ of some length $I_s \leqslant J$, and a list of states $\theta_1(s),\ldots,\theta_{I_s}(s)$, with the following properties. First,

$$U(X(u))-U(X(t)) = \int_t^u \int_\Theta \sum_{i<I_s} \frac{\partial \Pr_\theta\left(s,E_{j(s,i)}(s)\right)}{\partial t}\left[u(\theta,a_{j(s,i)})-u(\theta,a_{j(s,i+1)})\right]\pi(\theta)d\theta ds. \quad (22)$$

Second,

$$u(\theta,a_{j(s,i)})-u(\theta,a_{j(s,i+1)}) \gtrless 0 \quad \text{for } \theta \lessgtr \theta_i(s). \quad (23)$$

Third,

$$\theta_i(s) \geqslant \theta_{i-1}(s) \qquad \text{for all } i \in \{2,\ldots,I_s\} \text{ such that } D_{j(s,i)}(s) = \varnothing. \quad (24)$$

Step 2—Use accuracy to set a lower bound on the payoff difference. Take any θ, s and $i < I_s$, and consider the corresponding derivative appearing inside the summation in (22). We have

$$\begin{aligned}
\frac{\partial \Pr_\theta\left(s,E_{j(s,i)}(s)\right)}{\partial t} &= \lim_{\delta\to 0} \frac{\Pr_\theta\left(s+\delta,E_{j(s,i)}(s)\right)-\Pr_\theta\left(s,E_{j(s,i)}(s)\right)}{\delta} \\
&= \lim_{\delta\to 0} \frac{\Pr_\theta\left(s,\varphi_{s+\delta,s}(E_{j(s,i)}(s)|\theta)\right)-\Pr_\theta\left(s,E_{j(s,i)}(s)\right)}{\delta} \\
&\gtrless \lim_{\delta\to 0} \frac{\Pr_\theta\left(s,\varphi_{s+\delta,s}(E_{j(s,i)}(s)|\theta_i(s))\right)-\Pr_\theta\left(s,E_{j(s,i)}(s)\right)}{\delta} \quad \text{for } \theta \lessgtr \theta_i(s).
\end{aligned} \quad (25)$$

The second equality follows from the definition of the function $\varphi_{s+\delta,s}(\cdot|\cdot)$. The inequality, from $X(s+\delta)$ being more accurate than $X(s)$ for every $\delta > 0$ (as this means that $\varphi_{s+\delta,s}(x|\theta)$ is decreasing in θ for every x and $\delta > 0$) and $E_{j(s,i)}(s)$ being a lower set (the complement of an upper set). Letting $L(\theta,s,i)$ denote the right-hand side of (25), from (22), (23) and (25) we obtain

$$U(X(u))-U(X(t)) \geqslant \int_t^u \int_\Theta \sum_{i<I_s} L(\theta,s,i)\left[u(\theta,a_{j(s,i)})-u(\theta,a_{j(s,i+1)})\right]\pi(\theta)d\theta ds. \quad (26)$$

Step 3—Rewrite the lower bound. In this and the next step we prove that, for every s,

$$\int_\Theta \sum_{i<I_s} L(\theta,s,i)\left[u(\theta,a_{j(s,i)})-u(\theta,a_{j(s,i+1)})\right]\pi(\theta)d\theta \geqslant 0. \quad (27)$$

[46] That is, letting $\tilde{u}:\Theta\times A\setminus\{a_j\}\to\mathbb{R}$ denote the restriction of u to $\Theta\times A\setminus\{a_j\}$, the family $\{\tilde{u}(\theta,\cdot)\}_{\theta\in\Theta}$ is again an IDO family.

The result will then follow from (26) and (27). First note that, since $E_{j(s,i)}(s)$ is a lower set, for some function $\bar{x}_n : \mathbb{R}^{n-1} \to \mathbb{R} \cup \{-\infty, +\infty\}$ we have

$$\Pr_\theta\big(s, E_{j(s,i)}(s)\big) = \int_{\mathbb{R}^{n-1}} g_{<n}(s, x_{<n}|\theta) G_n(s, \bar{x}_n(x_{<n})|\theta, x_{<n}) dx_{<n},$$

where $g_{<n}(s,\cdot|\theta)$ is the density of $(X_1(s), \ldots, X_{n-1}(s))$ in state θ. Similarly, for every $\delta > 0$ and $i < I_s$ there is a function $\bar{x}_{n,i}(\delta,\cdot) : \mathbb{R}^{n-1} \to \mathbb{R} \cup \{-\infty, +\infty\}$ such that

$$\Pr_\theta\big(s, \varphi_{s+\delta,s}(E_{j(s,i)}(s)|\theta_i(s))\big) = \int_{\mathbb{R}^{n-1}} g_{<n}(s, x_{<n}|\theta) G_n(s, \bar{x}_{n,i}(\delta, x_{<n})|\theta, x_{<n}) dx_{<n}.$$

Taking limits, we conclude that, for every s and $i < I_s$,

$$L(\theta, s, i) = \int_{\partial E_{j(s,i)}(s)} g(s, x_{<n}, x_n|\theta) \underbrace{\left(\lim_{\delta \to 0} \frac{\bar{x}_{n,i}(\delta, x_{<n}) - \bar{x}_n(x_{<n})}{\delta} \right)}_{=: K(i, x_{<n})} dx_n dx_{<n},$$

where $\partial E_{j(s,i)}(s)$ denotes the boundary of $E_{j(s,i)}(s)$. Thus, the expression in (27) can be written as

$$\sum_{i<I_s} \int_{\partial E_{j(s,i)}(s)} K(i, x_{-n}) \int_\Theta g(s, x_{<n}, x_n|\theta) \big[u(\theta, a_{j(s,i)}) - u(\theta, a_{j(s,i+1)})\big] \pi(\theta) d\theta. \tag{28}$$

Step 4—Show that the lower bound is nonnegative. Since $\varphi_{s+\delta,s}(\theta, x)$ is decreasing in θ, for each $i < I_s$ such that $\theta_i(s) \leqslant \theta_{i+1}(s)$ we have $\bar{x}_{n,i}(\delta, x_{<n}) \geqslant \bar{x}_{n,i+1}(\delta, x_{<n})$, and hence $K(i, x_{<n}) \geqslant K(i+1, x_{<n})$ for all $x_{<n}$. Let $i_1 < \ldots < i_{H(s)}$ denote the set of indices $i < I_s$ such that $D_{j(s,i)}(s) \neq \varnothing$. Then, for every $h \in \{2, \ldots, H(s)\}$ and $i \in \{i_h + 1, \ldots, i_{h+1} - 1\}$, at each point in the boundary of $E_{j(s,i_h)}(s)$ the evaluator prefers $a_{j(s,i_{h+1})}$ to $a_{j(s,i)}$, and hence, using (24),

$$\begin{aligned}
&\sum_{i=i_h}^{i_{h+1}-1} \int_{\partial E_{j(s,i_h)}(s)} K(i, x_{<n}) \int_\Theta g(s, x_{<n}, x_n|\theta) \big[u(\theta, a_{j(s,i)}) - u(\theta, a_{j(s,i+1)})\big] \pi(\theta) d\theta \\
&\qquad \geqslant \int_{\partial E_{j(s,i_h)}(s)} K(i_h, x_{<n}) \int_\Theta g(s, x_{<n}, x_n|\theta) \big[u(\theta, a_{j(s,i_h)}) - u(\theta, a_{j(s,i_{h+1})})\big] \pi(\theta) d\theta.
\end{aligned}$$

It follows that the expression in (28) is at least as large as

$$\sum_{h<H(s)} \int_{\partial E_{j(s,i_h)}(s)} K(i_h, x_{<n}) \int_\Theta g(s, x_{<n}, x_n|\theta) \big[u(\theta, a_{j(s,i_h)}) - u(\theta, a_{j(s,i_{h+1})})\big] \pi(\theta) d\theta.$$

But all terms in this summation are zero, as $D_{j(s,i_h)}(s) \neq \varnothing$ and $D_{j(s,i_{h+1})}(s) \neq \varnothing$ imply that the evaluator is indifferent between $a_{j(s,i_h)}$ and $a_{j(s,i_{h+1})}$ at each point in the boundary of $E_{j(s,i_h)}(s)$. □

B.3 Continuous Actions

To deal with a continuous action set A we make two assumptions. First, we assume payoffs are continuous and bounded below (e.g. nonnegative). Second, we impose regularity on the family of functions $\{u(\theta,\cdot)\}_{\theta\in\Theta}$ by assuming that the family of their restrictions to every sufficiently large but finite subset of actions is also an IDO family. Note that the latter assumption is automatically satisfied with single-crossing or monotone preferences.[47] Moreover, it allows us to extend Theorem 0 by simply showing the following: for any fixed experiment X, the constrained welfare the evaluator obtains when restricted to choosing from a finite subset B of actions converges to the unconstrained welfare as B becomes large. We do this next.

Let $a(\cdot):\mathbb{R}^n\to A$ be the evaluator's (unconstrained) optimal strategy. Let $J=|B|$ and denote by $a_1<\ldots<a_J$ the elements of B. Define $a_B:\mathbb{R}^n\to B$ for the restricted problem as follows:

$$a_B(x)=a_1 \text{ if } a(x)\leqslant a_1, \quad a_B(x)=a_2 \text{ if } a_1<a(x)\leqslant a_2, \quad \ldots, \quad a_B(x)=a_J \text{ if } a(x)>a_{J-1}.$$

Then for every state θ, every B, and every b in B, we have

$$\Pr_\theta(a_B(X)<b)\leqslant\Pr_\theta(a(X)<b) \quad \text{and} \quad \Pr_\theta(a_B(X)\leqslant b)=\Pr_\theta(a(X)\leqslant b).$$

Thus, for every $\Theta'\subseteq\Theta$,

$$\int_{\Theta'}\Pr_\theta(a_B(x)<b)\pi(\theta)d\theta\leqslant\int_{\Theta'}\Pr_\theta(a(x)<b)\pi(\theta)d\theta$$

and

$$\int_{\Theta'}\Pr_\theta(a_B(x)\leqslant b)\pi(\theta)d\theta=\int_{\Theta'}\Pr_\theta(a(x)\leqslant b)\pi(\theta)d\theta.$$

This implies that for every c in the union of the B's we have

$$\limsup_B\int_{\Theta'}\Pr_\theta(a_B(x)<c)\pi(\theta)d\theta\leqslant\int_{\Theta'}\Pr_\theta(a(x)<c)\pi(\theta)d\theta$$

and

$$\liminf_B\int_{\Theta'}\Pr_\theta(a_B(x)\leqslant c)\pi(\theta)d\theta=\int_{\Theta'}\Pr_\theta(a(x)\leqslant c)\pi(\theta)d\theta.$$

Since Θ' is arbitrary and we can replace c with any a in A (because the union of the B's is dense in A), we conclude that the probability measure on states and actions induced by $a_B(\cdot)$ converges weakly to that induced by $a(\cdot)$. Thus, $\liminf_B\mathbb{E}_B(u)\geqslant\mathbb{E}(u)$, where $\mathbb{E}$ and $\mathbb{E}_B$ are the expectations with respect to the measures on $\Theta\times A$ induced by the optimal and B-constrained optimal strategy, respectively. Since $\mathbb{E}_B(u)\leqslant\mathbb{E}(u)$ for every B, we are done.

[47] In the continuous case, Karlin and Rubin's (1956) monotonicity means that every function $u(\theta,a)$ is (i) maximized at some $a(\theta)$ that is increasing in θ, and (ii) decreasing in a as a moves away from $a(\theta)$. Quah and Strulovici (2009) refer to these preferences as *quasi-concave with increasing peaks*.

References

Allcott, H. (2015): "Site Selection Bias in Program Evaluation," *Quarterly Journal of Economics*, 130(3), 1117–1165.

An, M. Y. (1998): "Logconcavity versus Logconvexity: A Complete Characterization," *Journal of Economic Theory*, 80(2), 350–369.

Banerjee, A. V., S. Chassang, S. Monteiro, and E. Snowberg (2017): "A Theory of Experimenters," NBER Working Paper No. 23867.

Bemmaor, A. C. (1992): "Modeling the Diffusion of New Durable Goods: Word-of-Mouth Effect Versus Consumer Heterogeneity," in *Research Traditions in Marketing. International Series in Quantitative Marketing*, ed. by G. Laurent, G. L. Lilien, and B. Pras, Springer, vol. 5.

Berger, J. O. (1985): *Statistical Decision Theory and Bayesian Analysis*, Springer, New York, NY, 2nd ed.

Bickel, P. J. and E. L. Lehmann (1979): "Descriptive Statistics for Nonparametric Models. IV," in *Contributions to Statistics, Jaroslav Hájek Memorial Volume*, ed. by J. Jurečková, Academia, Prague.

Blackwell, D. and J. L. Hodges (1957): "Design for the Control of Selection Bias," *Annals of Mathematical Statistics*, 28(2), 449–460.

Bradley, D. M. and R. C. Gupta (2003): "Limiting Behaviour of the Mean Residual Life," *Annals of the Institute of Statistical Mathematics*, 55(1), 217–226.

Brams, S. J. and M. D. Davis (1978): "Optimal Jury Selection: A Game-Theoretic Model for the Exercise of Peremptory Challenges," *Operations Research*, 26(6), 966–991.

Calabria, R. and G. Pulcini (1987): "On the Asymptotic Behaviour of the Mean Residual Life Function," *Reliability Engineering*, 19(3), 165–170.

Dahm, M., P. Gonzàlez, and N. Porteiro (2009): "Trials, Tricks and Transparency: How Disclosure Rules Affect Clinical Knowledge," *Journal of Health Economics*, 28(6), 1141–1153.

David, H. A. and H. N. Nagaraja (2003): *Order Statistics*, Wiley.

Degroot, M. H. and J. B. Kadane (1980): "Optimal Challenges for Selection," *Operations Research*, 28(4), 952–968.

Di Tillio, A., M. Ottaviani, and P. N. Sørensen (2017): "Persuasion Bias in Science: Can Economics Help?" *Economic Journal*, 127(605), F266–F304.

Efron, B. (1971): "Forcing a Sequential Experiment to be Balanced," *Biometrika*, 58(3), 403–417.

Felgenhauer, M. and E. Schulte (2014): "Strategic Private Experimentation," *American Economic Journal: Microeconomics*, 6(4), 74–105.

Fisher, R. and L. Tippett (1928): "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample," *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2), 180–190.

Fishman, M. J. and K. M. Hagerty (1990): "The Optimal Amount of Discretion to Allow in Disclosure," *Quarterly Journal of Economics*, 105(2), 427–444.

Flanagan, F. (2015): "Peremptory Challenges and Jury Selection," *Journal of Law and Economics*, 58(2), 385–416.

Glaeser, E. L. (2008): "Researcher Incentives and Empirical Methods," in *The Foundations of Positive and Normative Economics: A Hand Book*, ed. by A. Caplin and A. Schotter, New York: Oxford University Press, 300–319.

Gnedenko, B. (1943): "Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire," *Annals of Mathematics*, 44(3), 423–453.

Goel, P. K. and M. H. DeGroot (1992): "Comparison of Experiments for Selection and Censored Data Models," in *Bayesian Analysis in Statistics and Econometrics. Lecture Notes in Statistics*, ed. by P. K. Goel and N. S. Iyengar, Springer, New York, NY, vol. 75.

Grossman, S. (1981): "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics*, 24(3), 461–83.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–161.

Henry, E. (2009): "Strategic Disclosure of Research Results: The Cost of Proving Your Honesty," *Economic Journal*, 119(539), 1036–1064.

Henry, E. and M. Ottaviani (2019): "Research and the Approval Process: The Organization of Persuasion," *American Economic Review*, 109(3), 911–955.

Herresthal, C. (2017): "Hidden Testing and Selective Disclosure of Evidence," University of Cambridge mimeo.

Hoffmann, F., R. Inderst, and M. Ottaviani (2014): "Persuasion through Selective Disclosure: Implications for Marketing, Campaigning, and Privacy Regulation," Bocconi mimeo.

Holmström, B. (1999): "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 66(1), 169–182.

Jewitt, I. (2007): "Information Order in Decision and Agency Problems," Oxford mimeo.

Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101(6), 2590–2615.

Karlin, S. and H. Rubin (1956): "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio," *Annals of Mathematical Statistics*, 27(2), 272–299.

Kasy, M. (2016): "Why Experimenters Might not Always Want to Randomize, and What They Could Do Instead," *Political Analysis*, 24(3), 324–338.

Khaledi, B.-E. and S. Kochar (2000): "On Dispersive Ordering between Order Statistics in One-Sample and Two-Sample Problems," *Statistics & Probability Letters*, 46(3), 257–261.

Leadbetter, M. R., G. Lindgren, and H. Rootzén (1983): *Extremes and Related Properties of Random Sequences and Processes*, Springer.

Lehmann, E. L. (1988): "Comparing Location Experiments," *Annals of Statistics*, 16(2), 521–533.

Marshall, A. W. and I. Olkin (2007): *Life Distributions*, Springer.

Milgrom, P. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12(2), 380–391.

Milgrom, P. and I. Segal (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70(2), 583–601.

Milgrom, P. and C. Shannon (1994): "Monotone Comparative Statics," *Econometrica*, 62(1), 157–180.

Müller, S. and K. Rufibach (2008): "On the Max-Domain of Attraction of Distributions with Log-Concave Densities," *Statistics and Probability Letters*, 78(12), 1440–1444.

Neyman, J. S. (1923): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," Translated in *Statistical Science*, 5(4), 465–480.

Persico, N. (2000): "Information Acquisition in Auctions," *Econometrica*, 68(1), 135–148.

Quah, J. K.-H. and B. Strulovici (2009): "Comparative Statics, Informativeness, and the Interval Dominance Order," *Econometrica*, 77(6), 1949–1992.

Rayo, L. and I. Segal (2010): "Optimal Information Disclosure," *Journal of Political Economy*, 118(5), 949–987.

Resnick, S. I. (2008): *Extreme Values, Regular Variation, and Point Processes*, Springer.

Roth, A., J. B. Kadane, and M. H. Degroot (1977): "Optimal Peremptory Challenges in Trials by Juries: A Bilateral Sequential Process," *Operations Research*, 25(6), 901–919.

Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology*, 66(5), 688–701.

——— (1978): "Bayesian Inference for Causal Effects: The Role of Randomization," *Annals of Statistics*, 6(1), 34–58.

Schulz, K. F., I. Chalmers, R. J. Hayes, and D. G. Altman (1995): "Empirical Evidence of Bias: Dimensions of Methodological Quality Associated with Estimates of Treatment Effects in Controlled Trials," *Journal of the American Medical Association*, 273(5), 408–412.

Schwartz, E. P. and W. F. Schwartz (1996): "The Challenge of Peremptory Challenges," *Journal of Law, Economics, and Organization*, 12(2), 325–360.

Shaked, M. and Y. L. Tong (1990): "Comparison of Experiments for a Class of Positively Dependent Random Variables," *Canadian Journal of Statistics*, 18(1), 79–86.

——— (1993): "Comparison of Experiments of Some Multivariate Distributions with a Common Marginal," *Stochastic Inequalities*, 22, 79–86.

Tetenov, A. (2016): "An Economic Theory of Statistical Testing," Bristol mimeo.

Wainer, H. and D. Thissen (1994): "On Examinee Choice in Educational Testing," *Review of Educational Research*, 64(1), 159–195.

Wald, A. (1945): "Sequential Tests of Statistical Hypotheses," *Annals of Mathematical Statistics*, 16(2), 117–186.