Strategic Sample Selection*

Alfredo Di Tillio[†] Marco Ottaviani[‡] Peter Norman Sørensen[§]

November 12, 2020

Abstract

Are the highest sample realizations selected from a larger presample more or less informative than the same amount of random data? Developing multivariate accuracy for interval dominance ordered preferences, we show that sample selection always benefits (or always harms) a decision maker if the reverse hazard rate of the data distribution is log-supermodular (or log-submodular), as in location experiments with normal noise. We find non-pathological conditions under which the information contained in the winning bids of a symmetric auction decreases in the number of bidders. Exploiting extreme value theory, we quantify the limit amount of information revealed when the presample size (number of bidders) goes to infinity. In a model of equilibrium persuasion with costly information, we derive implications for the optimal design of selected experiments when selection is made by an examinee, a biased researcher, or contending sides with the peremptory challenge right to eliminate a number of jurors.

Keywords: Accuracy; Comparison of experiments; Strategic selection; Auctions; Information aggregation; Persuasion; Welfare; Design of experiments; Examinee choice; Peremptory challenge.

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, Alp Atakan, Pierpaolo Battigalli, Francesco Corielli, Basile Grassi, Stephan Lauermann, Danielle Li, Nicola Limodio, Marco Pagnozzi, José Penalva, Marek Pycia, Karl Schlag, Matt Shum, the reviewers, 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

Economic data are often nonrandomly selected, due to choices made by subjects under investigation or sample inclusion decisions by data analysts (see e.g. Heckman, 1979). Are selected data more or less informative than the same amount of random data? For example, fixing a number of objects for sale, are the winning bids in a more competitive auction more or less revealing of market demand? When a new treatment is given to the healthiest patients rather than to random patients in a group, does inference improve or worsen? When testing a candidate, should the examiner ask questions at random or allow the candidate to select the most preferred questions out of a larger batch? And how does the common-law 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 one and the same. There is an unknown state θ representing a fundamental demand driver, an average treatment effect, or a candidate ability. An evaluator must choose an action—estimate a parameter, choose a treatment, or assign a grade—knowing that marginally increasing the action decreases payoff when θ is low and increases it when θ is high. More precisely, we assume 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 acts after seeing the realization of an experiment, that is, a random vector $X = (X_1, \dots, X_n)$ whose distribution depends on θ. For instance, X_i may represent a bid in an auction, an outcome under a treatment, a student potential performance in a question, or a juror opinion. Consider the following two scenarios:

- *Random Experiment*. The sample observations are i.i.d. draws from a state-dependent cumulative distribution function $F(\cdot|\theta)$.
- *Maximally 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 $F(\cdot|\theta)$.

In which scenario is the expected payoff of the evaluator higher (in every IDO problem)?

The comparison is generally ambiguous. To fix ideas, take a simple hypothesis testing problem: two states $\theta_H > \theta_L$ and two actions, rejection (correct choice in θ_L) and acceptance (correct choice in θ_H). With sample size $n = 1$ and additive noise drawn from a normal distribution F, the observation is normal with mean θ_L in the low state and θ_H in the high state—drawn in blue in the left-hand panel of Figure 1. The evaluator optimally accepts if and only if the observation is above some cutoff $\bar{x}$, 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).[1] How does a selected experiment compare? The observation of the maximum of k i.i.d. draws has distribution $F^k(x - \theta_L)$ in the low state and $F^k(x - \theta_H)$ in the high state—drawn in red. To see why maximal selection benefits the evaluator with normal noise, as displayed in the

[1] An experiment with additive normal noise, like every other experiment considered in the paper, satisfies monotone likelihood ratio: 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 $\bar{x}$.

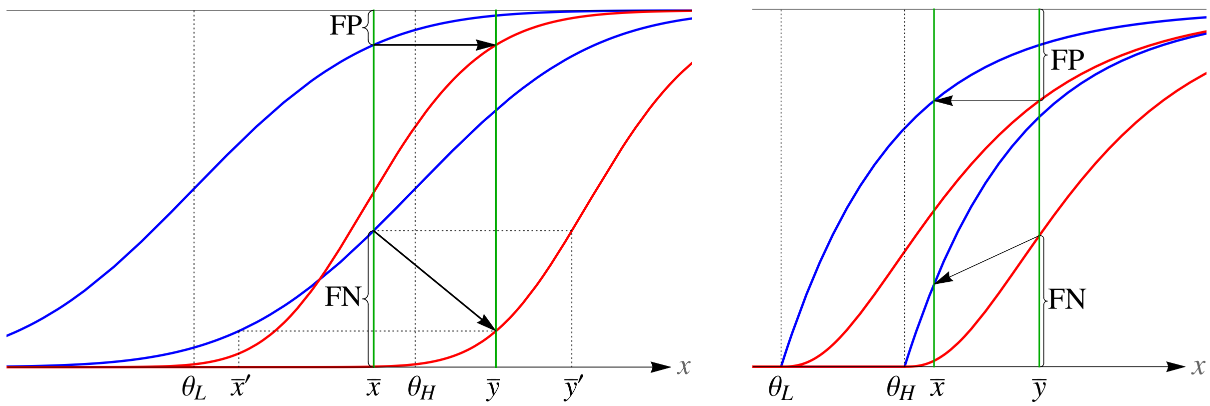

Figure 1: Maximal selection provides more accurate information with normal noise (left) but less accurate information with exponential noise (right).

left panel, note that by adopting the possibly suboptimal cutoff $\bar{y}$ defined so as to induce as many false positives, i.e. defined by $F^k(\bar{y}-\theta_L) = F(\bar{x}-\theta_L)$, the evaluator induces fewer false negatives: $F^k(\bar{y}-\theta_H) < F(\bar{x}-\theta_H)$. But, as shown in the right-hand panel of Figure 1, with exponential noise maximal selection harms the evaluator: given any cutoff $\bar{y}$ for the selected experiment, the evaluator can match false positives and lower false negatives by adopting cutoff $\bar{x}$ in the random experiment.

What makes selection beneficial in one case and harmful in the other? More generally, what is the welfare impact of selection with sample size $n \geqslant 1$ and possibly non-additive noise? To answer these questions, we start from Lehmann's (1988) notion of accuracy of an experiment (with $n = 1$), as illustrated in the left-hand panel of Figure 1. Recall that in the new (selected) experiment $\bar{y}$ induces by definition as many false positives as $\bar{x}$ induces in the original (random) experiment. Consider now setting the cutoff in the original experiment at $\bar{x}'$, so as to induce as many false *negatives* as $\bar{y}$ induces in the new experiment: $F(\bar{x}'-\theta_H) = F^k(\bar{y}-\theta_H)$. If $\bar{x}' < \bar{x}$, as it is in the figure, then $F(\bar{x}'-\theta_H) < F(\bar{x}-\theta_H)$, and we can conclude that $F^k(\bar{y}-\theta_H) < F(\bar{x}-\theta_H)$. Thus, in the new experiment $\bar{y}$ induces as many false positives and less false negatives than $\bar{x}$ induces in the original experiment. If the key inequality $\bar{x}' \leqslant \bar{x}$ holds for all possible values of θ_L and θ_H and every possible $\bar{x}$ and corresponding $\bar{x}'$, the new experiment is *more accurate* than the original experiment. Lehmann (1988) shows that more accuracy is necessary and sufficient to give higher welfare in every decision problem in Karlin and Rubin's (1956) class.[2]

To deal with general IDO problems and experiments with sample size $n \geqslant 1$, Section 3 develops a multidimensional version of accuracy, sharing the basic intuition with (and for $n = 1$, reducing

[2]An equivalent formulation (adopted e.g. by Persico, 2000, who coined the term *accuracy*) is the following: the new experiment is more accurate if, given any cutoff $\bar{x}$ for the original experiment, letting $\bar{y}$ be the cutoff that matches false positives in the new experiment, the cutoff that matches false negatives (denoted by $\bar{y}'$ in the figure) is larger than $\bar{y}$. As shown in Jewitt (2007), a unidimensional experiment satisfying monotone likelihood ratio is more accurate than another if and only if it is Blackwell more informative (Blackwell, 1951, 1953) in simple hypothesis testing problems.

Base	*Core*	*Applications*
Theorem 0 (accuracy)	Theorem 1 ($n = 1$, location)	Proposition 1 (auctions)
Proposition 0 (equilibrium)	Theorem 2 ($n \geqslant 1$, general)	Proposition 2 (optimal)
	Theorem 3 (extreme)	Propositions 3, 4 (delegated)

Extensions	
Theorem 4 (truncation)	Proposition 5 ($n = 1$, noisy)
Theorem 5 (median)	Proposition 6 ($n \geqslant 1$, noisy)

Table 1: Roadmap of results

to) Lehmann (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 an $(n-1)$-dimensional hypersurface. The evaluator accepts when the realization of X lies in the set $\bar{E}$ above the hypersurface.[3] Similarly to the unidimensional case, we define a hypersurface that induces as many false positives under Y as are induced by accepting in $\bar{E}$ under X. We say that Y is more accurate than X if the hypersurface we just defined also induces fewer false negatives—in particular, as many false negatives as would be induced under X by accepting in a set larger than $\bar{E}$. Theorem 0 proves that in all IDO problems welfare increases with accuracy, extending previous results by Persico (2000), Jewitt (2007) and Quah and Strulovici (2009).

Section 4 reports our core results. Theorem 1 identifies a necessary and sufficient condition for more selection, i.e. an increase in presample size k, to increase or decrease accuracy, and hence welfare, in an experiment with additive noise (called a *location* experiment) with sample size $n = 1$. Increasing k monotonically benefits the evaluator if and only if the reverse hazard function of the noise distribution, $-\log F$, is logconcave, as with normal or logistic noise. Likewise, welfare decreases in k if and only if $-\log F$ is logconvex, as with exponential noise. Selection is neutral only in one case: noise drawn from the Gumbel extreme value distribution, $F(\varepsilon) = \exp(-\exp(\varepsilon))$, the only distribution with both logconcave and logconvex reverse hazard function. This benchmark case provides intuition for our result. Deviating from Gumbel, maximal selection not only pushes realizations upward, but also changes the shape of the distribution. The pushed-up realizations are also more concentrated—improving accuracy and hence welfare—when the noise distribution is smaller than the Gumbel distribution in van Zwet's (1964) convex transform order. This means that noise is a concave transformation of, and hence has a thinner top tail than, a Gumbel distributed random variable.

Our notion of accuracy is the key tool needed to tackle the new issues arising in the multidimensional case with sample size $n \geqslant 1$. The main difficulty lies in the fact that selected 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. Our main result, Theorem 2, shows that welfare monotonically increases or decreases in

[3] 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 $\tilde{x}$.

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 reverse hazard rate f/F, strengthening the logconcavity criterion in Theorem 1.

Drawing on extreme value theory, we also quantify the impact of selection when presample size grows unboundedly large. Focusing on location experiments, Theorem 3 shows that an extremely selected sample gives the evaluator the full information payoff if and only if the hazard rate of the noise distribution $f(\varepsilon)/[1-F(\varepsilon)]$ is unbounded—for instance, with normal noise, or noise distributions with bounded-above support. With less than full information in the limit, welfare converges to the level corresponding to an experiment with scale parameter (proportional to variance) equal to the inverse of the limit hazard rate—building on Weissman (1978), we report a closed-form expression for the limit noise distribution.

Turning to applications, our theorems have immediate implications for the role of competition in aggregating private market information, complementing Wilson (1977) and Milgrom (1979). Section 5.1 considers an auction in which n identical objects are offered for sale to k symmetric bidders with interdependent values (Milgrom and Weber, 1982, 2000). As a direct corollary of our three core results, Proposition 1 characterizes when soliciting an additional bidder has a monotonic impact on the information revealed by the winning bids. In particular, in non-pathological cases in which bidders' signals have log-submodular reverse hazard rate, competition monotonically decreases information—overturning received wisdom. Using extreme value theory, we also quantify the amount of information revealed by the winning bids in the perfectly competitive limit.

Section 5.2 considers a different strategic source of selection, relevant for applications to educational testing and data collection: sample selection from the presample is delegated to a strategic sender (examinee or biased researcher) who wants to persuade the evaluator. Maximal selection arises in equilibrium (Proposition 0). Combining this observation with our core results, we characterize when allowing exam candidates to choose which questions to answer increases the informativeness of a test. By embedding the model into a potential outcomes framework (Neyman, 1923; Rubin, 1974) we then derive implications for the welfare impact of subversion of randomization in controlled trials (Schulz, 1995).

Taking a design perspective, what is the optimal experiment for the evaluator, when sampling and presampling are costly and endogenous? Section 5.3 illustrates how the evaluator can use presample size as an additional information channel, to economize on sample size when selection is beneficial. With sufficiently small presampling costs and unbounded hazard rate of the noise distribution, the optimal experiment must feature sample selection (Proposition 2).

We then turn to a general model of equilibrium persuasion with costly information where the choice of presample size is delegated to the sender. Relative to optimal persuasion (Rayo and Segal, 2010; Kamenica and Gentzkow, 2011), in our setting information acquisition is costly and manipulation is constrained by presample data. The choice of presample size is akin to agent effort in Holmström's (1999) career concern model.[4] The sender's incentive to collect more presample

[4]See also Glaeser (2008) for a broad discussion of incentives and biases in data collection and analysis.

data results in additional information that indirectly benefits or harms the evaluator through maximal selection. Smaller presampling costs strengthen this sender incentive to acquire information. An evaluator who indirectly benefits from this information can exploit this channel to save on sampling costs, as illustrated by Proposition 3. If instead the sender cannot commit not to disclose all information collected, full unraveling arises in equilibrium, with the sender effectively disclosing the whole presample data as in Grossman (1981) and Milgrom (1981). Similarly to Propositions 2 and 3, under unbounded hazard rate of the noise distribution the evaluator has then an incentive to block unraveling by committing to a fixed sample size, again to save on sampling costs (Proposition 4).

The general method of proof developed for Theorem 2 allows us to flesh out the common logic behind the comparison of other forms of selection such as truncation, previously considered by Goel and DeGroot (1992). According to Theorem 4, in important cases maximal selection and truncation lead to very different conclusions. For example, with normal or logistic noise maximal selection benefits while truncation harms the evaluator. In Section 6 we also analyze median selection, where the evaluator observes the median observation in a presample (Theorem 5), deriving conclusions for peremptory challenge. Finally, we extend our results to allow for noisy observation of presample as well as sample data (Propositions 5 and 6).

2 Setup

An evaluator with payoff function $u : \Theta \times A \to \mathbb{R}$ chooses an action $a \in A \subseteq \mathbb{R}$ under uncertainty about a state $\theta \in \Theta \subseteq \mathbb{R}$, where Θ is a finite set or a (possibly unbounded) interval. The prior is represented by a density (or mass) function $q(\theta)$ with cumulative distribution $Q(\theta)$. 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 A 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).[5]

Experiments and Welfare. Before deciding, the evaluator sees the realization of an *experiment*, a random vector X in $\mathbb{R}^n$ with state-dependent distribution $G(\cdot|\theta)$ and density $g(\cdot|\theta)$ satisfying

[5]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.

the following two properties. First, the set $\{x \in \mathbb{R}^n : g(x|\theta) > 0\}$, denoted by $S(X|\theta)$, is an open, convex set—in particular, $S(X|\theta)$ is also the interior of the support of X in state θ. Second, increasing the state increases the distribution of X in the likelihood ratio (LR) order,[6] implying in particular the monotone likelihood ratio (MLR) property: if $\theta' > \theta$ and $x' \geqq x$ then $g(x|\theta)g(x'|\theta') \geqslant g(x'|\theta)g(x|\theta')$.[7]

An important consequence of IDO and MLR is that the evaluator can without loss adopt a monotone strategy, where the action increases with the realization.[8] Thus, recalling 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, 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. The evaluator welfare, $\int_\Theta \sum_j \Pr_\theta(X \in E_j)u(\theta, a_j)dQ(\theta)$, can then be rewritten, summing by parts and disregarding constants, as

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

In the special case of a *location experiment*, observations have the form $X_i = \theta + \varepsilon_i$ and the noise vector $(\varepsilon_1, \ldots, \varepsilon_n)$ is drawn from some distribution G. The distributions $G(\cdot|\theta)$ are all shifted versions of G, with $G(x|\theta) = G(x_1 - \theta, \ldots, x_n - \theta)$ for all θ and x. Note that here MLR means that for every $\Delta > 0$ the noise density ratio $g(\varepsilon_1 + \Delta, \ldots, \varepsilon_n + \Delta)/g(\varepsilon_1, \ldots, \varepsilon_n)$ is decreasing in ε. With $n = 1$, this is simply logconcavity of g.[9]

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 $r = [q(\theta_L)/q(\theta_H)][u(\theta_L, a_L) - u(\theta_L, a_H)]/[u(\theta_H, a_H) - u(\theta_H, a_L)]$. 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 $-r\Pr_{\theta_L}(X \in \bar{E}) - \Pr_{\theta_H}(X \notin \bar{E})$, 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 (that is, i.i.d.) 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 a fixed sample size n welfare depends on the family of univariate distributions $F(\cdot|\theta)$ only. In this paper we are interested in experiments

---

[6]This means that for all $x, x' \in \mathbb{R}^n$ and $\theta' > \theta$, letting $x \vee x' = (\max\{x_1, x'_1\}, \ldots, \max\{x_n, x'_n\})$ and $x \wedge x' = (\min\{x_1, x'_1\}, \ldots, \min\{x_n, x'_n\})$, we have $g(x \vee x'|\theta')g(x \wedge x|\theta) \geqslant g(x|\theta)g(x'|\theta')$. See, for instance, Section 6.E in Shaked and Shanthikumar (2007).

[7]Given $x = (x_1, \ldots, x_n)$ and $x' = (x'_1, \ldots, x'_n)$ we say x' is larger than x, and write $x' \geqq x$, to mean $x'_i \geqslant x_i$ for all i.

[8]By Bayes' rule and MLR, the posterior belief on the state increases with the realization of X in the LR order: for all $\theta' \geqslant \theta$ and all $x' \geqq x$ in the set $\cup_{\theta \in \Theta} S(X|\theta)$, we have $q(\theta|x)q(\theta'|x') \geqslant q(\theta'|x)q(\theta|x')$. Thus, the evaluator cannot lose by increasing the action in response to a larger realization (Quah and Strulovici, 2009, Theorem 2).

[9]The ratio $g(\varepsilon + \Delta)/g(\varepsilon)$ decreases with ε for all $\Delta > 0$ if and only if $g'(\varepsilon + \Delta)/g(\varepsilon + \Delta) \leqslant g'(\varepsilon)/g(\varepsilon)$ for all ε and $\Delta > 0$, that is, if and only if $\log g$ is concave.

involving selected rather than random observations. In this scenario $G(\cdot|\theta)$ 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 focus is on *maximally selected* experiments, where $X_1 \geqslant X_2 \geqslant \cdots \geqslant X_n$ are the highest, second highest, ..., 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 from now on the indices $1, \ldots, i-1$ to save on notation, for every x we have

$$G_1(x_1|\theta) = F^k(x_1|\theta) \quad \text{and} \quad G_i(x_i|\theta, x_{<i}) = F^{k-i+1}(x_i|\theta)/F^{k-i+1}(x_{i-1}|\theta) \quad \text{for all } i > 1.^{10} \tag{2}$$

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 $F(\cdot|\theta)$: knowing in advance that observations are sorted so that $X_1 \geqslant \cdots \geqslant X_k$ is clearly of no value for the evaluator. Note that in the special case of a location experiment we can equivalently view selection as occurring on noise terms rather than on observations. Thus, with noise distribution F, we assume that F admits a logconcave density (so that the experiment satisfies LR ordering) and we have $X_i = \theta + \varepsilon_i$ for all i, with ε_1 drawn from F^k and ε_i drawn from $F^{k-i+1}(\cdot)/F^{k-i+1}(\varepsilon_{i-1})$. Finally, note that the distributions (2) are well defined for every real $k \geqslant n$ and continuously differentiable in k.

3 Multidimensional Accuracy

To assess the welfare impact of selection we develop a natural multidimensional generalization of Lehmann's (1988) notion of accuracy. Our notion can be used to compare two experiments (not necessarily selected experiments) with the same dimension by viewing them as boundary points of a parametrized family of experiments. Let $X(t)$ be a family of experiments indexed by $t \in [0,1]$, and let $G(t,\cdot|\theta)$ and $g(t,\cdot|\theta)$ be the corresponding state-dependent distributions and densities. Let $\Pr_\theta(t,\cdot)$ denote the measure on $\mathbb{R}^n$ induced by $X(t)$, and assume that for every event E the function $\Pr_\theta(\cdot, E)$ is continuously differentiable. In the application to maximal selection, we compare selected experiments with presample sizes k and m by viewing them as boundary points of $X(t)$, the experiment with presample size equal to the real number $tk + (1-t)m$.

Given any state θ and pair of indices s, t we define a bijection $\varphi_\theta : S(X(t)|\theta) \to S(X(s)|\theta)$ as follows: $\varphi_\theta(x_1, \ldots, x_n) = (z_1, \ldots, z_n)$, where $z_1, \ldots, z_n$ are defined recursively as

$$z_1 = (G_1(s,\cdot|\theta))^{-1}(G_1(t,x_1|\theta)) \quad \text{and} \quad z_i = (G_i(s,\cdot|\theta,z_{<i}))^{-1}(G_i(t,x_i|\theta,x_{<i})) \quad \text{for } i > 1.^{11} \tag{3}$$

[10] The corresponding joint density is $g(x|\theta) = [k!/(k-n)!]F^{k-n}(x_n|\theta)f(x_1|\theta)\cdots f(x_n|\theta)$ and the interior of the support is $S(X|\theta) = \{x \in \mathbb{R}^n : \underline{x}_\theta < x_n < \cdots < x_1 < \overline{x}_\theta\}$, where $\underline{x}_\theta$ and $\overline{x}_\theta$ are respectively the infimum and supremum of the set $\{x \in \mathbb{R} : f(x|\theta) > 0\}$. The density is LR-ordered with θ because log-supermodularity is preserved by integration (Karlin and Rinott, 1980) and products of log-supermodular functions are log-supermodular.

[11] Our notation omits the dependence of φ_θ on s and t for simplicity. Note that z_i is well defined because $x \in S(X(t)|\theta)$ implies $x_{<i} \in S(X_{<i}(t)|\theta)$ and hence $z_{<i} \in S(X_{<i}(s)|\theta)$. To see why φ_θ is a bijection (and is therefore invertible), note first that the map $x_1 \mapsto z_1$ is a strictly increasing bijection from $S(X_1(t)|\theta)$ onto $S(X_1(s)|\theta)$. Next, for

By definition, in state θ the random vector $\varphi_\theta(X(t))$ has the same distribution as $X(s)$. Thus, suppose that the evaluator accepts in an upper set $\bar{E}$ when testing a simple hypothesis $\theta = \theta_L$ vs. $\theta = \theta_H$ under $X(s)$. Then the evaluator can achieve the same false positives under $X(t)$ by (a) accepting when the realization x of $X(t)$ is such that $\varphi_{\theta_L}(x) \in \bar{E}$. Similarly, the evaluator can match false negatives by (b) accepting when $\varphi_{\theta_H}(x) \in \bar{E}$. Note that since $\bar{E}$ is an upper set, if $\varphi_{\theta_H}(\cdot) \leqq \varphi_{\theta_L}(\cdot)$ then the evaluator accepts more often with strategy (a) than with (b). But this means that under $X(t)$ strategy (a) gives as many false positives and fewer false negatives, compared to accepting in $\bar{E}$ under $X(s)$.

This heuristic argument motivates the following definition and provides an intuition for the result below. (Our argument is heuristic because $\varphi_\theta(x)$ is undefined when x lies outside the set $S(X(t)|\theta)$. Appendix A gives the formal argument.)

Definition. The family of experiments $X(t)$ is *ordered by accuracy* if $\varphi_{\theta'}(x) \leqq \varphi_\theta(x)$ for all states $\theta' > \theta$, all indices $t > s$, and all $x \in S(X(t)|\theta) \cap S(X(t)|\theta')$.

Note that the inequality $\varphi_{\theta'}(\cdot) \leqq \varphi_\theta(\cdot)$ is analogous (and for sample size $n = 1$, reduces) to the inequality $\bar{x}' \leqslant \bar{x}$ discussed in the introduction: were strategy (a) be such that false negatives rather than false positives are matched, then the set $\bar{E}$ should be larger—analogous to accepting above $\bar{x}'$ rather than above $\bar{x}$. In fact, it is clear 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 with Θ and A compact and $S(X(t)|\theta)$ constant across all t and θ. 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 Theorem 0, and discuss further the difference among IDO, single-crossing and monotone preferences, in Appendix A. The proofs for all other results in the paper are in Appendix B.

Unidimensional Location Experiments: Accuracy and Dispersion. When $X(t)$ is a family of unidimensional location experiments with noise distributions $G(t,\cdot)$, we can equivalently appeal to the notion of dispersion (Bickel and Lehmann, 1979). A univariate distribution G is *less dispersed* than another, F, if the quantile difference $G^{-1}(\cdot) - F^{-1}(\cdot)$ is decreasing. Equivalently, G is steeper than F at corresponding quantiles: $g(G^{-1}(\cdot)) \geqslant f(F^{-1}(\cdot))$. Lehmann (1988) shows that: (a) the family $X(t)$ is ordered by accuracy (for every possible choice of Θ) if and only if $G(t,\cdot)$ becomes less dispersed as t increases; (b) having a less dispersed noise distribution is necessary and sufficient for a location experiment to give higher welfare than another in every decision problem in

every $i > 1$ and $x_{<i} \in S(X_{<i}(t)|\theta)$ the map $x_i \mapsto z_i$ is a strictly increasing bijection from $S(X_i(t)|\theta, x_{<i})$, the interior of the support of $X_i(t)$ conditional on θ and $X_{<i}(t) = x_{<i}$, onto $S(X_i(s)|\theta, z_{<i})$, the interior of the support of $X_i(s)$ conditional on θ and $X_{<i}(s) = z_{<i}$.

Karlin and Rubin's (1956) class. It follows from (a), (b), and Theorem 0, that a family of unidimensional location experiments is ordered by increasing welfare in every IDO decision problem if and only if the corresponding noise distributions are ordered by decreasing dispersion.

4 Impact of Maximal Selection

This section characterizes 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. In particular, we answer our basic question as to when a selected experiment ($k > n$) improves or decreases welfare compared to a random experiment ($k = n$).

4.1 Unidimensional Location Experiments

We begin with unidimensional location experiments $X = \theta + \varepsilon$, where ε is the highest of $k \geqslant 1$ random draws from a noise distribution F admitting a logconcave density.

Theorem 1. *Fixing sample size to $n = 1$, an increase in presample size k increases (decreases) welfare in a selected location experiment if the noise reverse hazard function $-\log F$ is logconcave (logconvex). Conversely, allowing presample size to be any real number $k \geqslant 1$, if $-\log F$ is not logconcave (logconvex), then there is a payoff function in the IDO class such that welfare is not increasing (decreasing) in presample size.*[12]

Figure 2 illustrates the proof with normal noise—our main example of welfare-improving maximal selection. Based on the equivalence between accuracy and dispersion, given two presample sizes $k > m$ we ask what makes F^k steeper at any quantile ε_k than F^m at the corresponding quantile ε_m defined by $F^m(\varepsilon_m) = F^k(\varepsilon_k)$. By monotonicity of $-\log(\cdot)$, equivalently, when is $-\log F^k$ steeper at ε_k than $-\log F^m$ at ε_m? Or, when is the base $\varepsilon_k' - \varepsilon_k$ below the tangent to $-\log F^k$ at ε_k smaller than the base $\varepsilon_m' - \varepsilon_m$ below the tangent to $-\log F^m$ at ε_m? Note that $-\log F^k$ and $-\log F^m$ only differ by a multiplicative constant, so we can compute $\varepsilon_k' - \varepsilon_k$ also as the base below the tangent to $-\log F^m$, as displayed in the figure. Given that the base is the inverse of the rate of decay, $\varepsilon_k' - \varepsilon_k$ is smaller than $\varepsilon_m' - \varepsilon_m$ whenever $-\log F$ decays at an increasing rate, i.e., $-\log F$ is logconcave. When k and m can be any real numbers, ε_k and ε_m can be arbitrarily close to each other, explaining necessity in that case. The argument for logconvex $-\log F$ is analogous.

Gumbel Noise. The only noise distribution F such that $-\log F$ is both logconcave and logconvex (loglinear) is the Gumbel extreme value distribution, $F_{\text{Gum}}(\varepsilon) = \exp(-\exp(-\varepsilon))$. Given that $F^k_{\text{Gum}}(\varepsilon) = \exp(-k\exp(-\varepsilon)) = F_{\text{Gum}}(\varepsilon - \log k)$, maximal selection inflates noise by a constant

[12] 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.

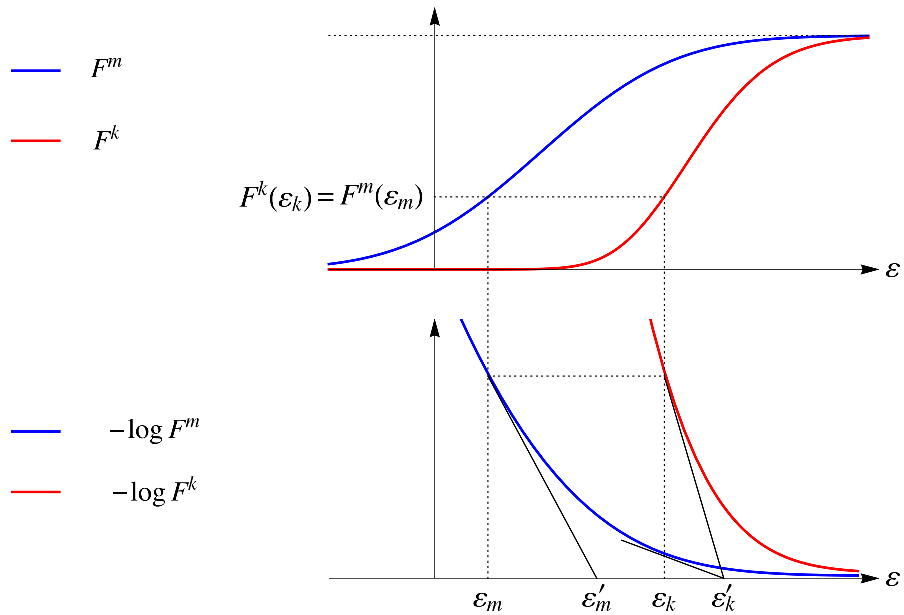

Figure 2: Normal noise: dispersion decreases with presample size (drawn for $k = 8$ and $m = 1$).

$(\log k)$ but does not change the shape of the distribution. Thus, every maximally selected experiment gives the same welfare—with Gumbel noise the reverse hazard function would be exponential (decaying at constant rate), with equal bases $\varepsilon'_k - \varepsilon_k = \varepsilon'_m - \varepsilon_m$.

Restatement in Terms of Convex Transform Order. Given two univariate distributions F and G, van Zwet (1964) defines F to be smaller than G in the *convex transform order*, denoted $F \leqslant_c G$, whenever $G^{-1}(F(\cdot))$ is convex. Given that $F_{\text{Gum}}^{-1}(F(\cdot)) = -\log(-\log F(\cdot))$, Theorem 1's logconcavity of $-\log F$ is equivalent to $F \leqslant_c F_{\text{Gum}}$. Equivalently, a random variable with logconcave $-\log F$ can be obtained through an increasing and concave transformation $F^{-1}(F_{\text{Gum}}(\cdot))$ of a random variable with Gumbel distribution. To gain intuition, visualize the random variable F^{-1} (on the vertical axis) as an increasing transformation of a Gumbel distributed random variable F_{Gum}^{-1} (on the horizonal axis) through a Q–Q plot. Concavity of $F^{-1}(F_{\text{Gum}}(\cdot))$ contracts the top tail, thus making it thinner than the top tail of the Gumbel distribution.

The restatement suggests an intuition for Theorem 1. Maximal selection raises realizations. The Gumbel distribution has exactly the shape that leaves dispersion unchanged after this operation. Deviating from Gumbel, dispersion is reduced (increased) when the quantile of the distribution F is more concave (convex) than the quantile of the Gumbel. Intuitively, concavity contracts the top tail above any percentile, thus making it thinner than the tail of the Gumbel distribution. When the tail is thinner (thicker) than Gumbel, the pushed up realizations in the top tail are more (less) informative than less selected realizations.

Examples of Welfare-Improving and Welfare-Decreasing Maximal Selection. Besides the normal case, one instance where more selection benefits is with logistic noise, $F(\varepsilon) = 1/(1+e^{-\varepsilon})$; we prove this and the following claims in Appendix B.2. Our main example of the opposite case, where $-\log F$ is logconvex and hence more selection harms, is exponential noise, $F(\varepsilon) = 1 - e^{-\varepsilon}$. 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 is the special case $a \to -1$. Finally, more selection harms with Bemmaor's (1992) shifted Gompertz, $F(\varepsilon) = (1 - \exp(-\varepsilon - \eta \exp(-\varepsilon)))$, as well as with the full support distribution $F(\varepsilon) = \exp(1 - \exp(\exp(-\varepsilon)))$ introduced by Noe (2020). Note also that in all these instances where more selection harms, the noise distribution is asymmetric. In effect, as shown below, this asymmetry is necessary for maximal selection to be harmful.

Minimal Selection. Theorem 1 has a symmetric counterpart for when the evaluator sees the lowest rather than the highest draw. Minimal selection increases (decreases) welfare if and only if the hazard function $-\log(1-F)$ is logconcave (logconvex). All results on maximal (as well as extreme) selection in the rest of the paper are also applicable to minimal selection, given that if $X_1 \leqslant \cdots \leqslant X_n$ are the n lowest of k random draws from $F(x|\theta)$, then $-X_1 \geqslant \cdots \geqslant -X_n$ are the n highest of k random draws from $\tilde{F}(x|\theta) = 1 - F(-x|\theta)$.

Fundamental Asymmetry between Beneficial and Harmful Selection. For symmetric distributions logconcavity (logconvexity) of the hazard function is equivalent to logconcavity (logconvexity) of the reverse hazard function. Thus, an immediate corollary of Theorem 1 is that with normal or logistic noise an increase in presample size is beneficial under both maximal and minimal selection.

But while more maximal and more minimal selection can both benefit, they *cannot* both harm. To see why, observe that a random experiment is informationally equivalent to the following: (i) two realizations are drawn; (ii) with equal chance the maximum or the minimum of the two is selected; (iii) the evaluator sees the selected draw, without knowing whether it is the maximum or the minimum. Now replace (iii) with (iii') the evaluator sees the selected draw, knowing whether it is the maximum or the minimum. If more maximal and more minimal selection were both harmful, the evaluator would be better off with a random experiment than with (i)–(iii'). This is impossible, because (i)–(iii') give more information than (i)–(iii).

Paired with Theorem 1 the above argument proves that there exists no noise distribution whose hazard function and reverse hazard function are both logconvex. In particular, in location experiments with $n = 1$ and symmetrically distributed noise *neither* maximal nor minimal selection can be harmful in every IDO problem.

Contribution to Stochastic Ordering of Order Statistics. The proof of Theorem 1 shows that $F <_c F_{\text{Gum}}$ is necessary and sufficient for F^k to become less dispersed as k increases. This complete characterization of the impact of maximal selection on dispersion of the highest order statistic appears to be new to the literature on stochastic orders. The closest result we could find in the literature, Khaledi and Kochar's (2000) Theorem 2.1, shows that if F has decreasing hazard rate then F^k becomes more dispersed as k increases. This result follows as a corollary of our characterization, given that F has decreasing hazard rate if and only if F is larger than the exponential distribution in the convex transform order, and the exponential distribution is in turn larger than the Gumbel distribution in this order.[13]

4.2 General Multidimensional Experiments

Extending our analysis to general (not necessarily location type) experiments with sample size $n \geqslant 1$ poses a challenge. Since order statistics are correlated, we cannot recover the accuracy ranking between selected experiments with (the same sample size n, but) different presample sizes by individually comparing the order statistics in the two experiments. Our multidimensional notion of accuracy allows us to characterize when the correlation structure in a maximally selected experiment benefits or harms the evaluator 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 upper bound of the support of $f(\cdot|\theta)$ independent of θ), that is, if for all states $\theta' > \theta$ the ratio*

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

is increasing (resp. decreasing).

As noticed by Noe (2020), with two states monotonicity of the reverse hazard ratio is equivalent to geometric dominance.[14]

General Method of Proof. The method of proof of Theorem 2 is applicable beyond selected experiments. Take any family $X(t)$ with respective distributions $G(t,\cdot|\theta)$ and suppose that for all t and θ the variables in $X(t)$ are *conditionally increasing in sequence* (CIS, Veinott, 1965): 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, $G_i(t,x_i|\theta,x_{<i})$ decreases in $x_{<i}$ for all x_i. Our method builds on two immediate observations. The first observation is that a CIS family $X(t)$ is ordered by accuracy if for all $s < t$ and $\theta' > \theta$, defining $z = \varphi_\theta(x)$ as in (3), we have

$$G_i(s,z_i|\theta', z_{<i})/G_i(t,x_i|\theta', x_{<i}) \geqslant 1 \quad \text{for all } i \geqslant 1 \tag{4}$$

(the conditioning on $x_{<i}$ and $z_{<i}$ is vacuous when $i = 1$). To interpret, consider a simple hypothesis testing problem where the evaluator either observes $X_i(s)$ when already knowing that $X_{<i}(s) = z_{<i}$, or observes $X_i(t)$ when already knowing that $X_{<i}(t) = x_{<i}$. Suppose that the acceptance cutoff is set at z_i in the first experiment. Since $X(s)$ and $\varphi_\theta(X(t))$ have the same distribution in the low state

[13] Since all distributions with logconcave densities have increasing hazard rate by Prekopas theorem, the only distribution with logconcave density for which Khaledi and Kochar's (2000) result applies is the exponential.

[14] According to Noe's (2020) Theorem 2, monotonicity of the reverse hazard ratio also characterizes when likelihood ratio dominance (between X drawn from $F(\cdot|\theta')$ and Y drawn from $F(\cdot|\theta)$ with two states $\theta' > \theta$) is preserved by competitive selection (and thus holds also between $X|X > Y$ and $Y|Y > X$). Remarkably, we find that the same condition drives the sign of the impact of maximal selection on the welfare of an evaluator with IDO preferences.

θ, setting the cutoff at x_i in the second experiment gives as many false positives. By (4), cutoff x_i also gives fewer false negatives—rejection in the high state θ' is less likely.

Second, a sufficient condition for (4) is that for each i the ratio is (i) no smaller than one in the limit as x_i grows large and (ii) monotonically decreasing in x_i. Continuing with the interpretation, the cutoff z_i reduces false negatives in the limit, and the reduction becomes relatively smaller as the cutoff x_i becomes larger. Applying the implicit function theorem to $z = \varphi_\theta(x)$, we can write requirement (ii) more revealingly in terms of reverse hazard rates:

$$\frac{g_i(s,z_i|\theta',z_{<i})/G_i(s,z_i|\theta',z_{<i})}{g_i(s,z_i|\theta,z_{<i})/G_i(s,z_i|\theta,z_{<i})} \leqslant \frac{g_i(t,x_i|\theta',x_{<i})/G_i(t,x_i|\theta',x_{<i})}{g_i(t,x_i|\theta,x_{<i})/G_i(t,x_i|\theta,x_{<i})} \quad \text{for all } i \geqslant 1. \tag{5}$$

Thus, (5) is a general sufficient condition for any CIS family 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. Recall from (2) that with presample size k 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}. Powers and right-truncations only change the reverse hazard rate by a multiplicative constant. In particular, the reverse hazard rate of the ith observation is $k-i+1$ times that of a random draw:

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

Taking the ratio between reverse hazard rates at different states, the constant disappears. In other words, when $X(t)$ and $X(s)$ are selected experiments with presample sizes $k_t \geqslant k_s$ condition (5) depends on t and s only through z, and (5) is simply log-supermodularity of $f(\cdot|\theta)/F(\cdot|\theta)$, because $k_t \geqslant k_s$ implies $z \leqq x$. Order statistics are CIS, so (5) can in fact be used.[15]

Positive Exponential Distribution: Neutrality of Maximal Selection. Positive-exponentially distributed observations, $F(x|\theta) = e^{\theta x}$ for $x \leqslant 0$ (and $\theta > 0$), are neutral to selection, analogous to Gumbel noise for location experiments. The reverse hazard rate is log-modular in this case, because $f(x|\theta)/F(x|\theta) = \theta$ is independent of x. Moreover, the upper bound of the support of $f(\cdot|\theta)$ is $x = 0$ independently of θ. Thus, by Theorem 2, selection has no impact on welfare.

Connection to Theorem 1. In location experiments, log-supermodularity of the reverse hazard rate is equivalent to logconcavity of the noise reverse hazard rate. Moreover, the upper bound of the support of $f(\cdot|\theta)$ is independent of θ if and only if it is infinite. Thus, we have:

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

Appendix B.2 shows that the hypotheses in Corollary 1, logconcavity or logconvexity of the reverse hazard rate, are stronger than the corresponding conditions in Theorem 1, yet the corollary

[15] Karlin and Rinott (1980) prove that order statistics are *multivariate totally positive of order 2*. This notion of positive dependence among random variables, known in economics as *affiliation* (Milgrom and Weber, 1982), is stronger than CIS, as shown in Barlow and Proschan (1975).

applies to all examples discussed earlier. As before, the Gumbel distribution is sandwiched between the noise distributions for which more selection benefits and those for which more selection harms: $f(\varepsilon)/F(\varepsilon) = \exp(-\varepsilon)$ is loglinear. More selection benefits with normal or logistic noise and harms with exponential or shifted Gompertz noise.

4.3 Extreme Selection

Complementing the monotone comparative statics results derived so far, we now analyze extreme selection, where presample size grows unbounded. The main result here, Theorem 3, characterizes the corresponding limit welfare. For simplicity, 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

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

The fundamental result of extreme value theory says that $\bar{F}$ must be Gumbel, Extreme Weibull or Frechet—see e.g. Leadbetter, Lindgren, and Rootzén (1983) for a primer. Our maintained assumption that F has a logconcave density f implies that $\bar{F}$ is, in fact, either Gumbel or Extreme Weibull, and always Gumbel if the support of f is unbounded above (Müller and Rufibach, 2008).

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. If instead α_k converges to a number $\alpha > 0$, then an extremely selected experiment is welfare-equivalent to a random experiment based on $\bar{F}(\cdot/\alpha)$. The limit behavior of the sequence α_k is in turn governed by the limit behavior of the noise hazard rate.

Theorem 3. **(a)** *For fixed sample size $n \geqslant 1$, as presample size grows without bound welfare converges to the full information payoff if and only if the noise distribution has unbounded hazard rate, that is, letting $\bar{\varepsilon} \in (-\infty, \infty]$ denote the upper bound of the support of f,*

$$\lim_{\varepsilon \to \bar{\varepsilon}} \frac{f(\varepsilon)}{1 - F(\varepsilon)} = \infty. \tag{UHR}$$

(b) *If UHR does not hold (so that, in particular, $\bar{\varepsilon} = \infty$) then, letting $\alpha = \lim_{\varepsilon \to \bar{\varepsilon}}[1 - F(\varepsilon)]/f(\varepsilon)$, the limit welfare is the welfare from an experiment with noise density*

$$(1/\alpha) \exp\left[-\exp(-\varepsilon_n/\alpha) - \varepsilon_1/\alpha - \cdots - \varepsilon_n/\alpha\right]. \tag{6}$$

In particular, with $n = 1$, the limit welfare is the welfare from an experiment with Gumbel noise $F_{\text{Gum}}(\cdot/\alpha) = \exp(-\exp(-\varepsilon/\alpha))$.

Pairing this result with Theorem 1, there is monotonic convergence to full information when noise has logconcave reverse hazard function and satisfies UHR. The hypotheses in Theorems 1 and 3 are overlapping but distinct. For example, UHR holds for normal but not for logistic noise. Moreover, UHR covers many distributions without logconcave reverse hazard function. First, UHR holds for all distributions with logconcave density and bounded-above support, e.g. all beta distributions with logconcave density, including uniform. Second, with unbounded-above support, beyond normal (or left-truncated normal, which has the same right tail), UHR holds also for all distributions in the exponential power family $f(\varepsilon) = [\gamma/\Gamma(1/\gamma)]\exp(-|\varepsilon|^\gamma)$ with shape parameter $\gamma > 1$. Strikingly, the Laplace ($\gamma = 1$), with exponential right tail, is the *only* member of this family with logconcave density for which $\alpha_k \not\to 0$. In this family, the negative impact of selection with exponential noise is not robust to extreme selection—an arbitrarily close distribution reverses the conclusion.

UHR and Unbounded Informativeness. The contribution of Theorem 3 is twofold. First, at a qualitative level, the result identifies UHR as the necessary and sufficient condition to have full information under extreme selection. A related notion of *unbounded informativeness* of a real-valued signal, familiar in the economics literature since Milgrom (1979), is the following:

$$\sup_x \frac{f(x|\theta')}{f(x|\theta)} = \infty \qquad \text{for all } \theta' > \theta. \tag{7}$$

In a location experiment satisfying MLR, which is our setup here, the two notions coincide. To see this, note that, when $f(x|\theta) = f(x-\theta)$ and MLR holds, (7) can be also written as follows: $\lim_{\varepsilon\to\bar{\varepsilon}} f(\varepsilon)/f(\varepsilon+c) = \infty$ for all $c > 0$. Clearly, every f with support bounded above satisfies this condition and, as we have already remarked, UHR. If instead the support is unbounded above, then the condition $\lim_{\varepsilon\to\varepsilon} f(\varepsilon)/f(\varepsilon+c) = \infty$ for all $c > 0$ is in turn equivalent to $-\lim_{\varepsilon\to\bar{\varepsilon}} f'(\varepsilon)/f(\varepsilon) = \lim_{\varepsilon\to\bar{\varepsilon}} f(\varepsilon)/[1-F(\varepsilon)] = \infty$, which is again UHR. The equivalence between UHR and (7) sheds light on the role of UHR in Theorem 3. Extreme selection pushes the signal toward its upper bound, and by MLR this is precisely where the signal is arbitrarily precise in identifying the state.

Second, Theorem 3 precisely quantifies the information contained in an extremely selected sample, also when that information is not full. In a unidimensional experiment the evaluator welfare approaches the level corresponding to a Gumbel experiment with scale parameter (proportional to variance) equal to α, the inverse of the limit hazard rate. More generally, the distribution in (6) with the same scale parameter α quantifies the limit welfare for any $n \geqslant 1$. This novel result showcases the power of extreme value theory.

5 Application to Strategic Settings

5.1 Information Aggregation in Auctions

Our characterization of the welfare impact of maximal selection abstracted away from the mechanism generating the data. Market competition for scarce resources naturally results in maximally selected outcomes—winning bids in an auction are the highest. Through a direct application of our results, this section characterizes when competition monotonically increases or decreases the amount of information contained in the winning bids.

Consider k symmetric unit-demand bidders competing in an auction for $n < k$ identical objects. Bidder i values an object $v_{k,i}(\theta, X_1, \ldots, X_k) = v_k(\theta, X_i, \{X_j\}_{j \neq i})$, where v_k is a nonnegative, continuous and increasing function, θ a common taste shifter, and X_i bidder i's private signal. Bidders have a common prior $q(\theta)$ and conditional on θ their signals $X_1, \ldots, X_k$ are i.i.d. draws from a distribution $F(\cdot|\theta)$ with MLR density $f(\cdot|\theta)$. The auction is either *discriminatory*, with each of the n highest bidders receiving an object at a price equal to the submitted bid, or *uniform-price*, with each of the n highest bidders receiving an object for a price equal to the highest rejected bid. MLR implies that $\theta, X_1, \ldots, X_k$ are affiliated random variables (Milgrom and Weber, 1982). Thus, as shown by Milgrom and Weber (2000), in a symmetric equilibrium of either auction each bidder i bids according to a continuously differentiable, strictly increasing function $b_{n,k}(\cdot)$.

Taking the point of view of an outside observer who has preferences in the IDO class and observes only the winning bids, we ask whether the extent of competition, namely k, has a beneficial or harmful impact on the observer welfare. Let $B_{1,k} \geqslant \cdots \geqslant B_{n,k}$ denote the winning bids. Since $b_{n,k}(\cdot)$ is strictly increasing, the bidder with the ith highest signal submits the ith highest bid—letting $X_{i,k}$ denote the ith highest signal, we have $B_{i,k} = b_{n,k}(X_{i,k})$ for all $1 \leqslant i \leqslant k$. This implies that the experiments $(B_{1,k}, \ldots, B_{n,k})$ and $(X_{1,k}, \ldots, X_{n,k})$ are informationally equivalent. The following result is therefore an immediate implication of Theorems 0, 2 and 3.

Proposition 1. **(a)** *If the reverse hazard rate $f(\cdot|\theta)/F(\cdot|\theta)$ is log-supermodular (log-submodular, with support of $f(\cdot|\theta)$ independent of θ) then competition increases (decreases) the accuracy of the winning bids and hence increases (decreases) the observer welfare.* **(b)** *Assuming $F(x|\theta) = F(x-\theta)$,* **(i)** *information is full in the limit $k \to \infty$ if and only if UHR holds;* **(ii)** *if UHR fails, the observer limit welfare is the welfare from an experiment with noise density* (6).

The aggregation of private information by an auction mechanism as the number of bidders grows unbounded, first studied in Wilson's (1977) seminal paper, was analyzed by Milgrom (1979) in a model similar to ours.[16] Assuming a single object ($n = 1$) and a pure common value $v_k(\theta, \cdot, \cdot) = \theta$ drawn from an ordered and nowhere-dense set, Milgrom (1979, Theorem 2) shows that unbounded informativeness (7) is necessary and sufficient to have full information in the perfectly competitive limit $k \to \infty$. Recalling that UHR is equivalent to (7), Proposition 1.b.i restates Milgrom's result, modulo the slightly different setup. The novel Proposition 1.b.ii quantifies the information value of perfect competition when UHR fails.

Beyond the limit case, in the auction literature nothing was known about the impact of soliciting an additional bidder on the information contained in the winning bids. Proposition 1.a gives broadly applicable conditions under which competition *monotonically* improves or worsens the

[16] Milgrom's (1979) model is identical to our unidimensional ($n = 1$) model, apart from the fact that he assumes a nowhere-dense Θ and does not restrict signals to obey MLR or even be real-valued. The two models become identical if (in his model) we impose the MLR property on $f(\cdot|\theta)$ and (in our model) we assume that Θ is finite.

information contained in the winning bids. First, our result characterizes when information, while not full in the limit, does improve with competition. Besides location signals with logconcave reverse hazard rate and bounded hazard rate (e.g. logistic noise), an example is with signals drawn from a positive exponential distribution $F(x|\theta) = e^{\theta x}$ (for $\theta > 0$ and $x \leqslant 0$). This is precisely Wilson's (1977) example of a case where information is not full in the limit. But it is notably not an instance where competition worsens information—recall that in this case maximal selection has no impact on welfare.

Second, while failure of convergence to full information is a possibility already contemplated (and characterized) in Milgrom (1979), whether competition could actually worsen information was not known. A natural conjecture could have been that even when information is not full in the limit, competition would still tend to improve information. Proposition 1 shows that in non-pathological cases the opposite is true: under logconvexity of the reverse hazard rate (e.g. location type signals with exponential or shifted Gompertz noise) more competition monotonically decreases the amount of information contained in the winning bids.

Finally, we show that our limit result for uniform-price auctions remains true if instead of assuming that the observer sees the winning bids we assume that the observer sees the price paid by the winning bidders, namely the $(n+1)$-highest bid $B_{n+1,k} = b_{n,k}(X_{n+1,k})$. Leadbetter et al. (1983, Theorem 2.2.2) show that if a pair of sequences β_k and $\alpha_k > 0$ makes $(X_{1,k} - \beta_k)/\alpha_k$ converge weakly to a nondegenerate random variable, then the same pair of sequences makes $(X_{n+1,k} - \beta_k)/\alpha_k$ converge weakly to a nondegenerate random variable. In particular, the limit distributions of the highest and $(n+1)$-highest order statistics, while different, have the same scale. Thus, by the arguments we used to prove Theorem 3 and the limit result in Proposition 1, experiment $B_{n+1,k}$ becomes arbitrarily informative as $k \to \infty$ if and only if UHR holds.

5.2 Delegated Selection

In the auction setup maximal selection arises from competition among strategic bidders who have no interest in the decision made by the outside observer. Now we turn to strategic situations in which the sample data provider cares about the action taken by the evaluator. Maximal selection results when the sender has an incentive to steer higher actions.

Delegated Selection Game. A strategic *sender*, privately informed about presample data, is tasked with sample selection. Taking sample size n and presample size k as given for now, consider the following *delegated selection game*. First, the sender privately observes k random draws from $F(\cdot|\theta)$ and chooses a subset of n draws. Second, the evaluator observes the selected draws and acts. The evaluator has IDO preferences and in every state the payoff of the sender is strictly increasing in the action of the evaluator. The following observation is immediate.

Proposition 0. *For all n and $k \geqslant n$ there is a Bayes Nash equilibrium in which for every realization of the k draws the sender selects the n largest draws, and the evaluator follows the optimal strategy for the maximally selected experiment with sample size n and presample size k.*

We use Bayes Nash equilibrium because the sender has private information. The result follows, as maximal selection is a best response for the sender to any monotone strategy of the evaluator, and likewise an evaluator's best response is monotone by MLR. Note that the equilibrium in the proposition is not eliminated by refinements, since no report can contradict maximal selection.

Our game constrains the sender to report precisely n data points. If instead the sender is unconstrained, unraveling occurs in equilibrium, as it is well known in the literature on strategic disclosure at least since Grossman (1981) and Milgrom (1981). When presample realizations are such that the action induced by submitting the full presample is larger than every action the sender can induce by submitting less than k data points, the sender has a strict incentive to disclose the whole presample. Thus, in the unconstrained game the evaluator behaves as if the entire presample were disclosed. However, once sample and presample size are costly and endogenous, Proposition 4 below shows that even when the sender cannot commit to disclose less than n data points, the *evaluator* can value commitment not to look at more than a set sample size.

Examinee Choice. The procedure of *agrégation* used in France to screen candidates for high school and university professor positions works as follows: "Candidates draw randomly a couple of subjects. The candidate is free to choose the subject which pleases him among these two, the one in which he feels best able to show and highlight his knowledge. He does not have to justify or comment on his choice."[17] This "give me your best shot" type procedure is commonly adopted in many other contexts. For example, first-year microeconomics exams at Bocconi require students to pick and answer only four out of five or two out of three questions presented in the exam.

Consider an examiner who must assign a grade $a \in A$ to a candidate of unknown ability $\theta \in \Theta$ after testing the candidate with a number of questions. From the ex ante perspective of the examiner, the performance in any given question is a random variable with distribution $F(\cdot|\theta)$, independent across questions. Once presented with any question, the candidate perfectly anticipates the performance in that question, an assumption relaxed in Section 6.3. Assuming that time allows a test with n questions, should the examiner ask n questions at random or require the candidate to select n questions from a larger set of $k > n$ questions?[18] Combining Proposition 0 with Theorem 2, we can immediately conclude that examinee choice improves or worsens the quality of testing depending on log-supermodularity or log-submodularity of $f(\cdot|\theta)/F(\cdot|\theta)$.

Researcher Bias in Potential Outcomes Framework. Consider a population of individuals and two alternative treatments—a default, known treatment 0 and a new treatment 1 whose benefit beyond the default is unknown. Following Neyman (1923) and Rubin (1974, 1978), let $X_{t,i}$ denote the potential outcome of individual i when receiving treatment $t \in \{0,1\}$. For simplicity, assume for now that the treatment effect $X_{1,i} - X_{0,i}$ is homogeneous across the population—Section 6 provides results that can be used to accommodate a more general case. Potential outcomes of individual i

[17] Source: Ministère de l'Éducation nationale et de la Jeunesse, https://agreg.org/data/uploads/rapports/rapport2019.pdf

[18] Wainer and Thissen (1994) emphasize that it is challenging for examiners to formulate questions of similar ex ante difficulty. Our assumption that performance is i.i.d. across questions assumes away this effect.

are $X_{0,i} = \varepsilon_i$ and $X_{1,i} = \theta + \varepsilon_i$, with ε_i drawn from a known F with logconcave density f.

Enter a researcher, who runs a controlled trial with n treated and n untreated individuals, denoted $1, \ldots, n$ and $n+1, \ldots, 2n$, respectively. Thus, the evaluator observes $X_{1,i} = \theta + \varepsilon_i$ for $i = 1, \ldots, n$ and $X_{0,i} = \varepsilon_i$ for $i = n+1, \ldots, 2n$. If sample selection and treatment assignment are random (subject to the equal group size constraint) then $\varepsilon_1, \ldots, \varepsilon_{2n}$ are random draws from F. Since F is known, the control group adds no information, and the experiment boils down to the treatment group—a random experiment with sample size n. Suppose instead that the researcher knows the outcome of treatment 0 for $k \geqslant 2n$ individuals, and on this basis (i) selects $2n$ individuals for the experiment and (ii) assigns n individuals to each treatment. An immediate extension of Proposition 0 gives an equilibrium in which the researcher assigns the individuals with the highest value of X_0 to treatment 1 and those with the lowest value of X_0 to treatment 0. Thus, $\varepsilon_1, \ldots, \varepsilon_n$ and $\varepsilon_{n+1}, \ldots, \varepsilon_{2n}$ are respectively the n largest and the n smallest of k random draws from F.

How do the two scenarios compare? By Corollary 1, under logconcavity of f/F selection benefits the evaluator directly—the treatment group alone is already more informative, and the control group can only add information. But it also benefits indirectly. The control group *does* add information under selection, because the untreated outcomes of the untreated individuals, $\varepsilon_{n+1}, \ldots, \varepsilon_{2n}$, are correlated with and hence informative about the counterfactual untreated outcomes of the treated, $\varepsilon_1, \ldots, \varepsilon_n$. When f/F is logconvex, the impact of selection is instead ambiguous: the treatment group alone is less informative than the random experiment, but this negative effect is partly balanced by the fact that the evaluator observes the control group.

5.3 Experiment Design

In many situations the evaluator has direct or indirect control over both sample and presample size. Realistically, presampling and sampling each involve a cost in our three leading applications. First, besides paying for soliciting bids the outside observer of an auction might have to compensate winning bidders unwilling to disclose their bids. Second, for an examiner both preparing and grading questions consume time and resources. Third, in the application to persuasion through data, the researcher bears cost for presample collection, and the evaluator spends time and cognitive effort to receive, verify, and analyze the sample selected by the researcher. Our results characterize when the evaluator can strategically leverage selection to increase the informativeness of an experiment and save on sampling cost.

Optimal Presampling. Suppose the evaluator can set both sample size n and presample size k at increasing costs $C_S(n)$ and $C_P(k)$, respectively. Let $U(k,n)$ denote the evaluator optimal expected (gross) payoff in the experiment with presample size k and sample size n. An *optimal experiment format* is a pair (k,n) maximizing $U(k,n) - C_P(k) - C_S(n)$.

Clearly, $U(k,n)$ increases in n and increases or decreases in k according to Theorems 1 and 2 (paired with Proposition 0 if selection is delegated). Thus, if selection harms then the optimal format features no selection. If selection benefits, the evaluator values sample size but also values sample selection: n and k are two goods. For a stark example, consider a location experiment with positive exponential noise. In this case the posterior belief about the state only depends on the largest observed realization, which for fixed k does not depend on n. As a consequence, $U(k,n) = U(k,1)$ for every k and $n \leqslant k$, hence at the optimum $n = 1$ and k solves $\max_{k \geqslant 1}[U(k,1) - C_S(1) - C_P(k)]$.

In general, the exact trade-off depends on the specific cost functions and family $F(\cdot|\theta)$ under consideration. But when selection is beneficial and presampling costs are sufficiently small the evaluator can always exploit selection to economize on sample size. Theorem 3 allows us to push this intuition further when presampling costs are very small and UHR holds.

Proposition 2. *Assume that $C_S(n)$ is unbounded and $F(x|\theta) = F(x-\theta)$ satisfies UHR. There exist $c > 0$ and $\bar{k} > 1$ such that if $C_P(\bar{k}) \leqslant c$ then every optimal format (k,n) is such that $k > n$.*

The proof of the proposition is based on a simple continuity argument. By the assumption on sampling costs, the design problem under constraint $k = n$ has an optimal solution $(\bar{n}, \bar{n})$, and by Theorem 3 we can choose k sufficiently large that $(k,1)$ gives higher (gross) payoff than $(\bar{n}, \bar{n})$.

Delegated Presampling. In some strategic settings the evaluator has limited control over presample size. For example, data collection could be delegated to a biased researcher who can costly increase the presample size. Presample size would then naturally be unobserved by the evaluator. As we shall see, much of the intuition from optimal presampling carries over to this scenario, although the evaluator now faces a more delicate trade-off.

Suppose that the evaluator sets the sample size n at cost $C_S(n)$ but can only decide *whether* to allow selection. By not allowing selection the evaluator obtains $U(n,n) - C_S(n)$. Otherwise, the following *delegated presampling game* is played. First, the sender *privately* chooses a presample size $k \geqslant n$ at cost $C_P(k)$. Second, the evaluator observes a maximally selected experiment with sample size n and presample size k and acts. As before, the evaluator has IDO preferences, while the sender's (gross) payoff is strictly increasing in the action of the evaluator.

Proposition 3. *Assume $C_S(n)$ is unbounded and $F(x|\theta) = F(x-\theta)$ satisfies UHR. There exist $c > 0$ and $\bar{k} > 1$ such that if $C_P(\bar{k}) \leqslant c$ then in every Bayes Nash equilibrium of the delegated presampling game with $n = 1$ the evaluator obtains a higher payoff than by setting any sample size $n \geqslant 1$ and not allowing selection.*

We use Bayes Nash equilibrium here, too, because the choice of k is unobserved. The proof of Proposition 3 is similar to that of Proposition 2. Given any sample size n, the sender has an incentive to increase k, because this choice is unobserved. But, in equilibrium, the evaluator correctly anticipates this incentive. Thus, from the perspective of the sender the additional presampling costs are partly wasted.[19] As presample costs become very small, the sender sets an arbitrarily large presample size, which by UHR benefits the evaluator even when $n = 1$. Thus, by delegating

[19]With Gumbel noise presampling is pure waste. Selection only shifts the noise distribution by a constant, so the equilibrium distribution over states and actions and hence both parties' gross payoff are the same as in the random experiment with equal sample size. Since inference is neutral to selection, presampling costs are completely wasted.

presampling the evaluator saves on sampling costs, effectively passing the cost of information to the sender. Besides proving the proposition, Appendix B.3 shows that an equilibrium always exists if we allow presample size to be any real number k and $C_P(k)$ is convex.

Blocking Unraveling. Finally, can the evaluator benefit by fully delegating the experiment to the sender? In a *full delegation game*, the sender either privately or publicly chooses $k \geqslant 1$ at cost $C_P(k)$, observes a presample of k draws, and selects from the presample a sample of any size $n \in \{1, \ldots, k\}$. The evaluator pays $C_S(n)$, observes the n selected draws, and acts. In an *unraveling* Bayesian equilibrium of a full delegation game, the evaluator behaves as if the sender discloses the whole presample.

Proposition 4. *Assume that $C_S(n)$ is strictly increasing, $C_P(k)$ is convex and unbounded, and $F(x|\theta) = F(x-\theta)$ satisfies UHR. There exist $c > 0$ and $\bar{k} > 1$ such that if $C_P(\bar{k}) \leqslant c$ then in every Bayes Nash equilibrium of the delegated presampling game with $n = 1$ the evaluator obtains a higher payoff than in every unraveling equilibrium of the full delegation game.*

The logic of the result is similar to Proposition 3. In an unraveling equilibrium the evaluator gets to see the entire presample—a random experiment, but this time a random experiment with sample size chosen by the sender. Due to the sender's lack of commitment not to disclose all data, in the full delegation game the evaluator ends up bearing a possibly large sample cost. Thus, the evaluator values commitment to receive only a limited amount of data. When presampling costs are sufficiently small, committing to a fixed sample size strengthens the sender incentive to collect enough presample data. It is therefore optimal for the evaluator to block unraveling. The evaluator does not get to see the entire presample, but this is better than seeing a whole but more expensive and possibly smaller presample.

Contribution to Literature on Strategic Data Selection. Blackwell and Hodges (1957) analyzed 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. Without modeling the information available to the researcher at the assignment stage, they posited that selection harms the evaluator—our analysis challenges this presumption.

In a complementary approach, Chassang, Padró i Miquel, and Snowberg (2012) characterize the design of experiments when outcomes are affected by unobserved actions by experimental subjects, rather than researchers. Kasy (2016) shows how deterministic assignment rules improve inference over randomization conditional on covariates. Tetenov (2016) analyzes an evaluator's optimal commitment to a decision rule when privately informed researchers select into costly testing. Banerjee, Chassang, Monteiro, and Snowberg (2020) analyze experiment design by an ambiguity-averse researcher facing an adversarial evaluator.

More closely, our work relates to the literature on voluntary disclosure and the welfare comparison to mandatory disclosure; see Matthews and Postlewaite (1985), Dahm, Gonzàlez, and Porteiro (2009), Henry (2009), Polinsky and Shavell (2012), Felgenhauer and Schulte (2014), Henry and

Ottaviani (2019), and Herresthal (2017). In the earliest precursor to our modeling approach, Fishman and Hagerty (1990) analyze selective disclosure in a setting with two states, binary signals, and sample size $n = 1$.[20] The implications we derive for the design of selected experiments hinge on our novel characterization of the impact of hidden data on the informativeness of the disclosed evidence.

6 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 other forms of selection.

6.1 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: $U(W_a) \leqslant U(W_b)$.*

Goel and DeGroot (1992) proved Theorem 4 for UHR distributions and monotone preferences. Their proof relies on UHR, which our different proof strategy shows to be inessential. Our result also applies more generally to IDO preferences. Meyer (2017) proves a related result and applies it to the comparison of the performance of sequential and simultaneous assignment protocols.

Theorem 4 compares unidimensional experiments. The extension 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 results in a more accurate experiment: if two families $X(t)$ and $X'(t)$ are both ordered by accuracy and $X(t)$ is independent of $X'(t)$ for every t, then $(X(t), X'(t))$ is also ordered by accuracy.

Truncation vs Maximal and Minimal Selection. Like maximal selection, left-truncation moves probability mass toward the upper tail of the distribution. Similarly, minimal selection as well as

[20] Di Tillio, Ottaviani, and Sørensen (2017) compare different types of selection in the potential outcome framework for an illustrative model with binary noise (violating the logconcavity assumption maintained in this paper). Hoffmann, Inderst, and Ottaviani (2020) analyze selective disclosure of a single realization of a continuous variable when the evaluator payoff is equal to the sum of two i.i.d. realizations from the same variable (and thus the state is two-dimensional, rather than one-dimensional as in the model considered here).

right truncation move mass toward the lower tail. However, Theorem 4 shows that in terms of welfare the right analogy to make is different. More right-truncation (from W_b to W_a) harms when the reverse hazard rate is log-supermodular, so its effect is analogous to less maximal selection (Theorem 1). Similarly, more left-truncation (from Y_a to Y_b) harms when the hazard rate is log-supermodular, so its impact is analogous to less minimal selection. The welfare consequences are strikingly different. With normal or logistic noise, hazard rate and reverse hazard rate are both logconcave, so both more maximal and more minimal selection are beneficial (Corollary 1). However, more truncation harms both ways: Y_b and W_a are respectively worse than Y_a and W_b.

6.2 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

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

Theorem 5. **(a)** *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 in every monotone problem.* **(b)** *A random experiment cannot increase welfare over a median selected experiment in every simple hypothesis testing problem.*

Thus, median selection is beneficial when both maximal and minimal selection are beneficial. But without any assumptions on the hazard rates (other than MLR, assumed throughout) a random experiment cannot be more accurate than a median selected experiment—another manifestation of the fundamental asymmetry between beneficial and harmful selection highlighted earlier.

Peremptory Challenge. The common-law right of peremptory challenge allows the attorneys on each side of a trial to reject a certain number of jurors. Consider a judge who must order a sentence based on the opinion of one juror. The judge knows that conditional on the defendant's level of guilt θ the jurors' estimates are independently distributed according to $F(\cdot|\theta)$.[21] The prosecuting attorney—desiring the judge to take higher actions—and the defense attorney—desiring the judge to take smaller actions—have the right to strike down $(k-1)/2$ jurors each from an initial set of

[21] 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.

k jurors. Both attorneys anticipate each juror'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. 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. Thus, peremptory challenge leads the judge to decide based on the opinion of the median juror. Theorem 5 provides a prior-free criterion to assess whether peremptory challenge provides the judge with more accurate information, relative to a randomly chosen juror.

6.3 Noisy Selection

The results obtained so far assume that selection occurs directly on the variable observed by the evaluator. In some applications it is natural to consider also a more general setup, where selection operates indirectly through a concomitant variable unobserved by the evaluator. Focusing on location experiments, here we assume the evaluator observes $X_i = \theta + \varepsilon_i$ but selection occurs on the concomitant variable $Y_i = \theta + \delta_i$. Thus, the evaluator observes X_i when Y_i ranks among the n highest in a presample of k units—for concreteness, imagine the evaluator observing the weight of the n tallest in a group of $k > n$ individuals.

In order to investigate the impact of selection in this more general setup, we must specify a model relating the *concomitant noise* terms δ_i and the noise terms ε_i. We posit a linear model $\varepsilon_i = c\delta_i + \gamma_i$ with $c > 0$ and we assume that $\delta_1, \ldots, \delta_k$ are i.i.d. draws from F_δ, while the *added noise* terms $\gamma_1, \ldots, \gamma_k$ are independent of $\delta_1, \ldots, \delta_k$ and identically (not necessarily independently) drawn from F_γ. Note that if $c = 0$ then X_i depends on Y_i only through θ and hence selection is clearly irrelevant—the further dependence added through noise ($c \neq 0$) is necessary for selection to have impact.[22] With $c = 1$ and F_γ degenerate at $\gamma = 0$ we recover our baseline setup.

Unidimensional Experiments. The conclusion of Theorem 1 extends seamlessly. If the added noise density is logconcave, maximal selection benefits or harms under the same conditions, applied to the concomitant noise distribution.

Proposition 5. *Assume that the added noise distribution F_γ has a logconcave density. Fixing sample size to $n = 1$, an increase in presample size increases (decreases) welfare if the concomitant noise reverse hazard function $-\log F_\delta(\cdot)$ is logconcave (logconvex).*

The proof of the proposition exploits the equivalence between accuracy and dispersion in unidimensional location experiments (Lehmann, 1988). As shown in the proof of Theorem 1, when $-\log F_\delta(\cdot)$ is logconcave (logconvex) a larger presample size reduces (increases) dispersion in the

[22] The assumption $c > 0$ is without loss; the case $c < 0$ is covered by redefining δ_i to be $-\delta_i$, with distribution $1 - F_\delta(-\delta_i)$.

location experiment with noise distribution F_δ. The dispersive order is closed under multiplication by a constant and convolution with an independent variable having a logconcave density (Lewis and Thompson, 1981). Since the added noise is independent of concomitant noise and identically distributed across the presample, a larger presample size will also reduce (increase) dispersion in the experiment actually observed by the evaluator.

Multidimensional Experiments. Extending the conclusion of Corollary 1 is considerably more complicated. Even though the noise terms $\gamma_1, \ldots, \gamma_k$ are identically distributed, their variation can make the sample units' ranking in terms of δ values different from their ranking in terms of ε values—an issue that cannot arise in a unidimensional experiment. Assuming a presample-wide common value $\gamma_1 = \cdots = \gamma_k$ eliminates this variation, allowing us to extend Corollary 1.

Proposition 6. *Assume the added noise distribution F_γ has a logconcave density and $\gamma_1 = \cdots = \gamma_k$. For fixed sample size $n \geqslant 1$, an increase in presample size increases (decreases) welfare if the concomitant noise reverse hazard rate $f_\delta(\cdot)/F_\delta(\cdot)$ is logconcave (logconvex, with support of f_δ unbounded above).*

The proof of this result is based on a simple intuition. No matter how large n or k are, or how little the dispersion of F_δ is, all the evaluator can hope to learn about is $\theta + \gamma_1$ rather than θ. Considering the auxiliary problem where the state is $\theta + \gamma_1$, the conditions in the theorem characterize when a more selected experiment is more or less informative in that problem (Corollary 1, plus the fact that the reverse hazard rate of a linear transformation of a random variable with logconcave reverse hazard rate is again logconcave). Since F_γ has a logconcave density, the state $\theta + \gamma_1$ in the auxiliary problem is in turn informative about the state θ in the original problem, so the same conditions characterize the impact of selection in the original model, too.

Noisy Delegated Selection. Proposition 0 also extends to the setup of this section. Consider the *noisy delegated selection game*: the sender privately observes $Y_i = \theta + \delta_i$ for each unit i in a set of k units and on this basis selects n units; the evaluator observes $X_i = \theta + \varepsilon_i = \theta + c\delta + \gamma_i$ for each selected unit and takes an action. Assuming that γ_i has a logconcave density, in this game there is a Bayes Nash equilibrium in which the sender selects the n units with the largest Y values, and the evaluator's strategy is monotone in the observed X values. Maximal selection on Y is in turn a best response for the sender—compared to any other strategy, it induces a first-order stochastic dominance increase in the X values observed by the evaluator.

Examinee Choice and Researcher Bias with Noise. Pushing our strategic applications in a natural direction, the results in this section cover realistic cases in which sender and evaluator have asymmetric information: the sender observes (and selects sample units based on) a variable Y, the evaluator observes X for the selected units. To illustrate this extension in the examinee choice setting, suppose that if candidate ability is θ then, ex ante, examiner and candidate view the performance in any given question i as normal random variables $X_i = \theta + \varepsilon_{e,i}$ and $Y_i = \theta + \delta_i$, respectively. Let σ_ε^2 and σ_δ^2 denote the variances of ε_i and δ_i, and ρ their correlation coefficient. Letting $c = \rho\sigma_\varepsilon/\sigma_\delta$ denote the coefficient of the regression of ε on δ, and $\gamma_i = \varepsilon_i - c\delta_i$ the corresponding error term (which is orthogonal to δ_i), we have $\varepsilon_i = c\delta_i + \gamma_i$, as posited in our model. Note that while we call γ_i an *added* noise term, the precision of the evaluator observations also depends on c. In the examinee choice example, suppose X_i and Y_i are imperfect signals of the candidate true performance $Z_i = \theta + \eta_i$ so that $\varepsilon_i = \eta_i + \tilde{\varepsilon}_i$ and $\delta_i = \eta_i + \tilde{\delta}_i$. The candidate observes the true performance more or less well than the evaluator according to whether $c < \rho$ or $c > \rho$.

Noisy Median Selection. Suppose that the evaluator observes X_i for the unit i with the rth highest value of Y in a presample of $k = 2r+1$ units. By the arguments in the proof of Theorem 5 and again by Lewis and Thompson's (1981) result, this experiment improves welfare over a random experiment when the concomitant noise has both a logconcave hazard rate $f_\delta(\cdot)/[1-F_\delta(\cdot)]$ and a logconcave reverse hazard rate $f_\delta(\cdot)/F_\delta(\cdot)$. Thus, Theorem 5 also generalizes (for location type signals) to the model of this section. This result covers, for example, peremptory challenge scenarios in which lawyers have private information about juror bias but are less or more informed than the judge about how the juror will vote.

6.4 Uncertain Selection and Beyond

Our analysis assumes that the evaluator knows the presample size. In more realistic scenarios the evaluator may be uncertain about presample size or even fail 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. Our results on beneficial selection are robust to small amounts of uncertainty—the evaluator can behave as if k is known, and expected payoffs are continuous in k. But more sizeable uncertainty tends to harm the evaluator, an important caveat. This is particularly evident in a location experiment with Gumbel noise; anticipated selection leaves the evaluator indifferent, so any uncertainty on k makes the evaluator strictly worse off. A characterization of the impact of uncertain selection remains an open problem.

Unanticipated Selection. Consider an unwary evaluator who wrongly anticipates a smaller presample size than true. This evaluator is clearly worse off than a rational evaluator. More interestingly, if a rational evaluator benefits from selection then it is ambiguous whether the unwary evaluator gains or loses when the true presample size is larger than expected. Consider simple hypothesis testing with noise distribution F symmetric around zero, so that $F(\varepsilon) = 1 - F(-\varepsilon)$. Let $\bar{x}$ be the optimal acceptance cutoff when $k = 1$. Selection with $k = 2$ changes the welfare of an unwary evaluator who maintains the cutoff at $\bar{x}$ as follows:

$$-q(\theta_L)\underbrace{\left[F(\bar{x}-\theta_L) - F^2(\bar{x}-\theta_L)\right]}_{\text{increase in false positives}}\left[u(\theta_L,a_1) - u(\theta_L,a_2)\right]$$
$$+q(\theta_H)\underbrace{\left[F(\bar{x}-\theta_H) - F^2(\bar{x}-\theta_H)\right)}_{\text{reduction in false negatives}}\left[u(\theta_H,a_2) - u(\theta_H,a_1)\right]. \quad (8)$$

In the important benchmark case of *equipoise*, the evaluator is a priori indifferent between accepting and rejecting, that is, $q(\theta_L)[u(\theta_L,a_1)-u(\theta_L,a_2)] = q(\theta_L)[u(\theta_H,a_2)-u(\theta_H,a_1)]$, and hence $\bar{x} = (\theta_L+\theta_H)/2$. By symmetry, $F(\bar{x}-\theta_L)+F(\bar{x}-\theta_H) = 1$, so the loss from the increase in false positives exactly offsets the gain from the reduction in false negatives—the expression in (8) is zero, so the unwary evaluator is indifferent between no selection and selection with $k=2$. In fact, the first order derivative of the expression in (8) with respect to $\bar{x}$ is positive at $\bar{x} = (\theta_L+\theta_H)/2$, so the unwary evaluator strictly benefits from selection when $\bar{x}$ is slightly above $(\theta_L+\theta_H)/2$, that is, when the evaluator would reject at the prior. Going beyond this example, it is easy to see that the impact of selection on an unwary evaluator can be also negative. For example, as $k\to\infty$ the unwary evaluator accepts with probability converging to one—with payoff converging to $q(\theta_L)u(\theta_L,a_2)+q(\theta_H)u(\theta_H,a_2)$, at most equal to the no-information payoff.

A 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. Both proofs rely on Lemma 3 below.

Before proceeding to Lemma 3 we need to establish two preliminary results.

Lemma 1. *Fix an experiment X and for every state θ let $\bar{S}(X|\theta)$ denote the smallest upper set containing $S(X|\theta)$. Then $\bar{S}(X|\theta)$ is decreasing in θ.*[23]

Proof. Suppose by contradiction that for some $\theta' > \theta$ and $x\in\bar{S}(X|\theta')$ we have $x\notin\bar{S}(X|\theta)$. By definition of $\bar{S}(X|\theta')$ there exists $x'\in S(X|\theta')$ such that $x \geqq x'$. Note that since $\bar{S}(X|\theta)$ is an upper set and $x\notin\bar{S}(X|\theta)$, we also have $x'\notin\bar{S}(X|\theta)$. Now pick any $x''\in S(X|\theta)$. Since $x'\wedge x'' \leqq x'$ and $x'\notin\bar{S}(X|\theta)$, it follows that $x'\wedge x''\notin S(X|\theta)$, that is, $g(x'\wedge x''|\theta)=0$.[24] The LR ordering of X with respect to the state now gives the contradiction:

$$\underbrace{g(x'\wedge x''|\theta)}_{=0}\,g(x'\vee x''|\theta') \geqslant \underbrace{g(x''|\theta)}_{>0}\underbrace{g(x'|\theta')}_{>0}.$$

□

In what follows we write $S(X|>\theta)$ as an abbreviation for $\cup_{\theta'>\theta}S(X|\theta')$.

Lemma 2. *Given any experiment X, the set $S(X|>\theta)\setminus S(X|\theta)$ is decreasing in θ.*

Proof. Suppose by contradiction that for some $\theta' > \theta$ and $x\in S(X|>\theta')\setminus S(X|\theta')$ we have $x\notin S(X|>\theta)\setminus S(X|\theta)$. Then $x\notin S(X|\theta')$, and there exists $\theta''>\theta'$ such that $x\in S(X|\theta)\cap S(X|\theta'')$.

[23] In the statements of Lemmas 1, 2 and 3, the decrease is in the sense of set inclusion.

[24] Recall the notation $x\vee x' = (\max\{x_1,x_1'\},\dots,\max\{x_n,x_n'\})$ and $x\wedge x' = (\min\{x_1,x_1'\},\dots,\min\{x_n,x_n'\})$.

Observe that $x \in S(X|\theta)$ and $x \notin S(X|\theta')$ together imply $x \notin \bar{S}(X|\theta')$. To see why, suppose that $x \in \bar{S}(X|\theta')$. Then there exists $x' \in S(X|\theta')$ such that $x \geqq x'$. Thus, by the LR ordering of X,

$$\underbrace{g(x \vee x'|\theta')}_{=g(x|\theta')=0} g(x \wedge x'|\theta) \geqslant \underbrace{g(x|\theta)}_{>0} \underbrace{g(x'|\theta')}_{>0},$$

a contradiction. Since $x \notin \bar{S}(X|\theta')$ and $\theta'' > \theta'$, we conclude by Lemma 1 that $x \notin \bar{S}(X|\theta'')$, contradicting our earlier assumption that $x \in S(X|\theta'')$. $\square$

Lemma 3. *If a family of experiments $X(t)$ is ordered by accuracy and E is an upper set, then for all $t > s$ the set $\tilde{E}(\theta) = \varphi_\theta^{-1}\big(E \cap S(X(s)|\theta)\big) \cup \big(S(X(t)| > \theta) \setminus S(X(t)|\theta)\big)$ is decreasing in θ.*

Proof. Fix two states $\theta' > \theta$. For brevity, write X and Y instead of $X(s)$ and $X(t)$, respectively. We must show that

$$\varphi_{\theta'}^{-1}(E \cap S(X|\theta')) \cup (S(Y| > \theta') \setminus S(Y|\theta')) \subseteq \varphi_\theta^{-1}(E \cap S(X|\theta)) \cup (S(Y| > \theta) \setminus S(Y|\theta)).$$

By Lemma 2, $(S(Y| > \theta') \setminus S(Y|\theta')) \subseteq (S(Y| > \theta) \setminus S(Y|\theta))$. Thus, it suffices to prove

$$\varphi_{\theta'}^{-1}(E \cap S(X|\theta')) \subseteq \varphi_\theta^{-1}(E \cap S(X|\theta)) \cup (S(Y| > \theta) \setminus S(Y|\theta)).$$

But the range of $\varphi_{\theta'}^{-1}(\cdot)$ is $S(Y|\theta')$, and $S(Y|\theta') \subseteq S(Y| > \theta)$. The above inclusion is therefore equivalent to the following:

$$\varphi_{\theta'}^{-1}(E \cap S(X|\theta')) \cap S(Y|\theta) \subseteq \varphi_\theta^{-1}(E \cap S(X|\theta)).$$

Fix any $x' \in E \cap S(X|\theta')$, let $y = \varphi_{\theta'}^{-1}(x')$, and suppose that $y \in S(Y|\theta)$. Note that $y \in S(Y|\theta')$ by definition. We have to prove that there exists $x \in E \cap S(X|\theta)$ such that $\varphi_\theta^{-1}(x) = y$. Let $x = \varphi_\theta(y)$, which is a well defined element of $S(X|\theta)$ because $y \in S(Y|\theta)$. Then $x = \varphi_\theta(y) \geqq \varphi_{\theta'}(y) = x'$, where the inequality follows from the order by accuracy, given that $y \in S(Y|\theta) \cap S(Y|\theta')$. Since E is an upper set and $x' \in E$, we are done. $\square$

A.1 Monotone Preferences

Recall that preferences are monotone 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 $(E_1, \ldots, E_J)$ be the evaluator's optimal partition of $\mathbb{R}^n$ for experiment $X(s)$, and $\bar{E}_j := E_j \cup \cdots E_J$. The evaluator welfare is

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

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

$$\Pr_\theta(X(t) \in E'_{j+1}) - \Pr_\theta(X(s) \in \bar{E}_{j+1}) \tag{9}$$

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 $X(t) \notin E'_2$, choose a_2 when $X(t) \in E'_2 \setminus E'_3$, and so on. For every $j < J$ and state θ define

$$\tilde{E}_{j+1}(\theta) = \varphi_\theta^{-1}(\bar{E}_{j+1} \cap S(X(s)|\theta)) \cup (S(X(t)| > \theta) \setminus S(X(t)|\theta)).$$

Let $E'_{j+1} = \tilde{E}_{j+1}(\theta_j)$ for every $j < J$. Since $\bar{E}_{j+1}$ is decreasing in j, $\tilde{E}_{j+1}(\theta)$ is decreasing in j and, by Lemma 3, also decreasing in θ. Thus, $E'_2 \supseteq \cdots \supseteq E'_J$. Moreover,

$$\begin{aligned}\Pr_\theta(X(s) \in \bar{E}_{j+1}) &= \Pr_\theta\left(X(t) \in \varphi_\theta^{-1}(\bar{E}_{j+1} \cap S(X(s)|\theta))\right) \\ &= \Pr_\theta(X(t) \in \tilde{E}_{j+1}(\theta)),\end{aligned}$$

where the first equality follows from the fact that $\varphi_\theta(X(t))$ and $X(s)$ have the same distribution in state θ, and the second from the definition of $\tilde{E}_{j+1}(\theta)$ and the fact that $S(X(t)|\theta)$ is the interior of the support of $X(t)$ in state θ, so that $\Pr_\theta(X(t) \in S(X(t)| > \theta) \setminus S(X(t)|\theta)) = 0$. Thus, we can rewrite (9) as

$$\Pr_\theta(X(t) \in \tilde{E}_{j+1}(\theta_j)) - \Pr_\theta(X(t) \in \tilde{E}_{j+1}(\theta)).$$

For $\theta \leqslant \theta_j$ the difference is nonpositive, because in this case $\tilde{E}_{j+1}(\theta_j) \subseteq \tilde{E}_{j+1}(\theta)$. For $\theta > \theta_j$ it is nonnegative, because then $\tilde{E}_{j+1}(\theta_j) \supseteq \tilde{E}_{j+1}(\theta_j)$. □

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 in Appendix B.5.

A.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 sets $\bar{E}'_2, \ldots, \bar{E}'_J$ 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.

Besides Lemma 3, the proof of the theorem uses the following:

Lemma 4. *Fix a family of experiments $X(t)$, a state θ, two indices s,t, and an upper set E. Define $\tilde{E}(\theta)$ as in Lemma 3. Fix any $x_{<n} \in \mathbb{R}^{n-1}$. Then the section of $\tilde{E}(\theta)$ corresponding to $x_{<n}$, namely the set $\{x' \in \tilde{E}(\theta) : x'_{<n} = x_{<n}\}$, is an upper set relative to the section of $\cup_{\theta' \in \Theta} S(X(t)|\theta')$ corresponding to $x_{<n}$, namely the set $\{x' \in \mathbb{R}^n : x'_{<n} = x_{<n}$ and $x' \in S(X(t)|\theta')$ for some $\theta' \in \Theta\}$.*

Proof. Fix any $x_n \in \mathbb{R}$ such that $x = (x_{<n}, x_n) \in \tilde{E}(\theta)$ and any $x'_n > x_n$ such that $x' = (x_{<n}, x'_n) \in S(X(t)|\theta')$ for some state θ'. We have to prove that $x' \in \tilde{E}(\theta)$. We distinguish three cases:

Case 1: $x' \in S(X(t)|\theta)$ and $x \in S(X(t)|\theta)$. In this case $\varphi_\theta(x)$ and $\varphi_\theta(x')$ are both well defined, and moreover $\varphi_\theta(x) \in E \cap S(X(s)|\theta)$ and $\varphi_\theta(x') \in S(X(s)|\theta)$. Since $x' \geqq x$ and $x'_{<n} = x_{<n}$, and given $x_{<n}$ the last coordinate of φ_θ is a strictly increasing function on $S(X_n(t)|\theta, x_{<n})$, we conclude that $\varphi_\theta(x') \geqq \varphi_\theta(x)$. But E is an upper set and $\varphi_\theta(x) \in E$, so we must also have $\varphi_\theta(x') \in E$. We conclude that $\varphi_\theta(x') \in E \cap S(X(s)|\theta)$, that is, $x' \in \tilde{E}(\theta)$.

Case 2: $x' \in S(X(t)|\theta)$ and $x \in S(X(t)| > \theta) \setminus S(X(t)|\theta)$. This case cannot arise, because $x' \geqq x$, the LR ordering of $X(t)$, and $x \in S(X(t)|\theta')$ for some $\theta' > \theta$ together give the contradiction

$$\underbrace{g(x|\theta)}_{=0} g(x'|\theta') \geqslant \underbrace{g(x|\theta')}_{>0} \underbrace{g(x'|\theta)}_{>0}.$$

Case 3: $x' \notin S(X(t)|\theta)$. In this case we must have $x' \in S(X(t)| > \theta)$. Indeed, if $x' \notin S(X(t)|\theta) \cup S(X(t)| > \theta)$ then $x' \in S(X(t)|\theta')$ for some $\theta' < \theta$. But $x \in S(X(t)|\theta) \cup S(X(t)| > \theta)$, so for some $\theta'' \geqslant \theta$ the LR ordering of $X(t)$ gives the contradiction

$$\underbrace{g(x'|\theta'')}_{=0} g(x|\theta') \geqslant \underbrace{g(x|\theta'')}_{>0} \underbrace{g(x'|\theta')}_{>0}.$$

□

Proof of Theorem 0—IDO preferences. Let $\{E_1^t, \ldots, E_J^t\}$ be the optimal partition of $\mathbb{R}^n$ for experiment $X(t)$, with $\bar{E}_j^t = E_j^t \cup \cdots \cup E_J^t$ an upper set for all $1 < j \leqslant J$, and action a_j chosen when $X(t) \in E_j^t$. Let $\Pr_\theta(t, \cdot)$ denote the measure on $\mathbb{R}^n$ induced by $X(t)$. The evaluator welfare is

$$U(X(t)) = \int_\Theta \sum_{j<J} \Pr_\theta(t, \bar{E}_{j+1}^t) \left[u(\theta, a_{j+1}) - u(\theta, a_j)\right] dQ(\theta).$$

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

$$U(X(u)) - U(X(t)) = \int_t^u \int_\Theta \sum_{j<J} \frac{\partial \Pr_\theta(s, \bar{E}_{j+1}^s)}{\partial t} \left[u(\theta, a_{j+1}) - u(\theta, a_j)\right] dQ(\theta) ds, \tag{10}$$

and we have to show that the expression in (10) 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 (10), as follows. Recall that, by IDO, for every $1 \leqslant j < J$ there exists a state θ_j such that the difference $u(\theta,a_{j+1})-u(\theta,a_j)$ is nonpositive for $\theta \leqslant \theta_j$ and nonnegative for $\theta \geqslant \theta_j$. An immediate consequence of this observation is that for any $1<j<J$ such that $\theta_j < \theta_{j-1}$, action a_j can be removed from A without affecting the IDO property: letting $\tilde{u}:\Theta\times A\setminus\{a_j\}\to\mathbb{R}$ be the restriction of u to $\Theta\times A\setminus\{a_j\}$, the family $\{\tilde{u}(\theta,\cdot)\}_{\theta\in\Theta}$ is again IDO. By using this fact (repeatedly, if necessary) together with the fact that $\bar{E}^s_{j+1}=\bar{E}^s_j$ when $E^s_j=\varnothing$, 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 $\theta^s_1,\dots,\theta^s_{I_s-1}$, with the following properties. First,

$$U(X(u))-U(X(t))=\int_t^u\int_\Theta\sum_{i<I_s}\frac{\partial \mathrm{Pr}_\theta\big(s,\bar{E}^s_{j(s,i+1)}\big)}{\partial t}\big[u(\theta,a_{j(s,i+1)})-u(\theta,a_{j(s,i)})\big]dQ(\theta)ds. \quad (11)$$

Second, for every state θ and $1\leqslant i<I_s$,

$$u(\theta,a_{j(s,i+1)})-u(\theta,a_{j(s,i)})\lesseqgtr 0 \quad \text{if} \quad \theta\lesseqgtr\theta^s_i. \quad (12)$$

Third, for every $1\leqslant i<I_s-1$,

$$\theta^s_{i+1}\geqslant\theta^s_i \quad \text{if} \quad E^s_{j(s,i+1)}=\varnothing. \quad (13)$$

Step 2—Use order by accuracy to set a lower bound on the payoff difference. For every θ, s, $\delta>0$, and $i<I_s$, define

$$\tilde{E}^s_{i+1}(\theta,\delta)=\varphi_\theta^{-1}\big(\bar{E}^s_{j(s,i+1)}\cap S(X(s-\delta)|\theta)\big)\cup\big(S(X(s)|>\theta)\setminus S(X(s)|\theta)\big).$$

By construction, $\mathrm{Pr}_\theta(s-\delta,\bar{E}^s_{j(s,i+1)})=\mathrm{Pr}_\theta(s,\tilde{E}^s_{i+1}(\theta,\delta))$, and by Lemma 3, $\tilde{E}^s_{i+1}(\theta,\delta)$ is decreasing in θ. It follows that

$$\begin{aligned}\frac{\partial \mathrm{Pr}_\theta\big(s,\bar{E}^s_{j(s,i+1)}\big)}{\partial t}&=\lim_{\delta\to 0}\frac{\mathrm{Pr}_\theta\big(s,\bar{E}^s_{j(s,i+1)}\big)-\mathrm{Pr}_\theta\big(s-\delta,\bar{E}^s_{j(s,i+1)}\big)}{\delta}\\&=\lim_{\delta\to 0}\frac{\mathrm{Pr}_\theta\big(s,\bar{E}^s_{j(s,i+1)}\big)-\mathrm{Pr}_\theta\big(s,\tilde{E}^s_{i+1}(\theta,\delta)\big)}{\delta}\\&\lesseqgtr\lim_{\delta\to 0}\frac{\mathrm{Pr}_\theta\big(s,\bar{E}^s_{j(s,i+1)}\big)-\mathrm{Pr}_\theta\big(s,\tilde{E}^s_{i+1}(\theta^s_i,\delta)\big)}{\delta}\quad \text{for } \theta\lesseqgtr\theta^s_i.\end{aligned} \quad (14)$$

Let $L^s_i(\theta)$ denote the right-hand side of (14). From (11), (12) and (14) we obtain

$$U(X(u))-U(X(t))\geqslant\int_t^u\sum_{i<I_s}\int_\Theta L^s_i(\theta)\big[u(\theta,a_{j(s,i+1)})-u(\theta,a_{j(s,i)})\big]dQ(\theta)ds. \quad (15)$$

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

$$\sum_{i<I_s}\int_\Theta L^s_i(\theta)\big[u(\theta,a_{j(s,i+1)})-u(\theta,a_{j(s,i)})\big]dQ(\theta)\geqslant 0. \quad (16)$$

The result will then follow from (15) and (16). First note that, since $\bar{E}^s_{j(s,i+1)}$ is an upper set, for some function $\bar{x}^s_n : \mathbb{R}^{n-1} \to \mathbb{R} \cup \{-\infty, +\infty\}$ we have

$$\Pr_\theta\left(s, \bar{E}^s_{j(s,i+1)}\right) = \int_{\mathbb{R}^{n-1}} g_{<n}(s, x_{<n}|\theta)\left[1 - G_n(s, \bar{x}^s_n(x_{<n})|\theta, x_{<n})\right] dx_{<n},$$

where $g_{<n}(s,\cdot|\theta)$ is the density of $(X_1(s),\ldots,X_{n-1}(s))$ in state θ. Similarly, by Lemma 4, for every $i < I_s$, $\delta > 0$ and $x_{<n} \in \mathbb{R}^{n-1}$ there is a function $\bar{x}^s_{n,i}(\delta,\cdot) : \mathbb{R}^{n-1} \to \mathbb{R} \cup \{-\infty, +\infty\}$ such that

$$\Pr_\theta\left(s, \bar{E}^s_{i+1}(\theta,\delta)\right) = \int_{\mathbb{R}^{n-1}} g_{<n}(s, x_{<n}|\theta)\left[1 - G_n(s, \bar{x}^s_{n,i}(\delta, x_{<n})|\theta, x_{<n})\right] dx_{<n}.$$ [25]

Thus,

$$\begin{aligned}
\lim_{\delta\to 0} \frac{\Pr_\theta\left(s, \bar{E}^s_{j(s,i+1)}\right) - \Pr_\theta\left(s, \bar{E}^s_{i+1}(\theta,\delta)\right)}{\delta} &= \lim_{\delta\to 0} \frac{G_n(s, \bar{x}^s_n(x_{<n})|\theta, x_{<n}) - G_n(s, \bar{x}^s_{n,i}(x_{<n})|\theta, x_{<n})}{\delta} \\
&= \underbrace{\lim_{\delta\to 0} \frac{G_n(s, \bar{x}^s_n(x_{<n})|\theta, x_{<n}) - G_n(s, \bar{x}^s_{n,i}(\delta, x_{<n})|\theta, x_{<n})}{\bar{x}^s_n(x_{<n}) - \bar{x}^s_{n,i}(\delta, x_{<n})}}_{= g(s, x_{<n}, \bar{x}^s_n(x_{<n})|\theta)} \times \underbrace{\lim_{\delta\to 0} \frac{\bar{x}_n(x_{<n}) - \bar{x}^s_{n,i}(\delta, x_{<n})}{\delta}}_{=: K^s_i(x_{<n})}.
\end{aligned}$$

It follows that

$$L^s_i(\theta) = \int_{\partial \bar{E}^s_{j(s,i+1)}} K^s_i(x_{<n}) g(s, x_{<n}, x_n|\theta) dx_n dx_{<n},$$

where $\partial \bar{E}^s_{j(s,i+1)}$ denotes the boundary of $\bar{E}^s_{j(s,i+1)}$, and the expression in (16) can be written as

$$\sum_{i<I_s} \int_{\partial \bar{E}_{j(s,i+1)}} K^s_i(x_{<n}) \int_\Theta g(s, x_{<n}, x_n|\theta)\left[u(\theta, a_{j(s,i+1)}) - u(\theta, a_{j(s,i)})\right] dQ(\theta). \tag{17}$$

Step 4—Show that the lower bound is nonnegative. Since for every $i < I_s$ and $\delta > 0$ the set $\bar{E}^s_{i+1}(\theta,\delta)$ is decreasing in θ by Lemma 3, for each $1 \leqslant i < I_s - 1$ such that $\theta^s_i \leqslant \theta^s_{i+1}$ we have $\bar{x}^s_{n,i}(\delta, x_{<n}) \leqslant \bar{x}^s_{n,i+1}(\delta, x_{<n})$, and hence $K^s_i(x_{<n}) \geqslant K^s_{i+1}(x_{<n})$, for all $x_{<n}$. Let $i_1 < \ldots < i_{H_s}$ denote the set of indices $1 \leqslant i < I_s$ such that $E_{j(s,i+1)}(s) \neq \varnothing$. Then for every $1 \leqslant h < H_s$ and $i \in \{i_h + 1, \ldots, i_{h+1} - 1\}$ we have $\bar{E}^s_{j(s,i)} = \bar{E}^s_{j(s,i_{h+1})}$. Furthermore, at each point on the boundary $\partial \bar{E}_{j(s,i_{h+1})}(s)$ the expected payoff from action $a_{j(s,i_{h+1})}$ is at least as large as the expected payoff from $a_{j(s,i)}$. Thus, using (13), for every $1 \leqslant h < H(s)$ we have

$$\begin{aligned}
&\sum_{i_h \leqslant i < i_{h+1}} \int_{\partial \bar{E}_{j(s,i+1)}} K^s_i(x_{<n}) \int_\Theta g(s, x_{<n}, x_n|\theta)\left[u(\theta, a_{j(s,i+1)}) - u(\theta, a_{j(s,i)})\right] dQ(\theta) \\
&\qquad \geqslant \int_{\partial \bar{E}_{j(s,i_{h+1})}} K^s_{i_h}(x_{<n}) \int_\Theta g(s, x_{<n}, x_n|\theta)\left[u(\theta, a_{j(s,i_{h+1})}) - u(\theta, a_{j(s,i_h)})\right] dQ(\theta).
\end{aligned}$$

[25] The function $\bar{x}^s_n$ is such that $x_{<n} \mapsto \inf\{x_n : (x_{<n}, x_n) \in \bar{E}^s_{j(s,i+1)}\}$, while $\bar{x}^s_{n,i}(\delta,\cdot)$ is such that

$$x_{<n} \mapsto \min\left\{\inf\{x_n \in \mathbb{R} : (x_{<n}, x_n) \in \varphi^{-1}_{\theta^s_i, s-\delta, s}(E \cap S(X(s-\delta)|\theta))\}, \inf\{x_n \in \mathbb{R} : (x_{<n}, x_n) \in S(Y(s)| > \theta) \setminus S(Y(s)|\theta)\}\right\}.$$

In both cases, we adopt the usual convention that the infimum of the empty set is assumed to be $+\infty$. Our expressions use the obvious notations $g_n(-\infty|\theta, x_{<n}) = g_n(+\infty|\theta, x_{<n}) = G_n(-\infty|\theta, x_{<n}) = 0$ and $G_n(+\infty|\theta, x_{<n}) = 1$.

It follows that the expression in (17) is at least as large as

$$\sum_{h<H_s} \int_{\partial \bar{E}_{j(s,i_{h+1})}} K^s_{i_h}(x_{<n}) \int_{\Theta} g(s,x_{<n},x_n|\theta) \left[u(\theta,a_{j(s,i_{h+1})}) - u(\theta,a_{j(s,i_h)})\right] dQ(\theta).$$

Since $E_{j(s,i_h)} \neq \varnothing$ and $E_{j(s,i_{h+1})} \neq \varnothing$ for every $h < H_s$, at each point on the boundary $\partial \bar{E}_{j(s,i_{h+1})}$ the evaluator is indifferent between the actions $a_{j(s,i_h)}$ and $a_{j(s,i_{h+1})}$. Thus, all inner integrals, and hence all terms in the summation, are zero, and the proof is complete. $\square$

A.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. This assumption is automatically satisfied with single-crossing or monotone preferences.[26] 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 $\Theta' \subseteq \Theta$,

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

and

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

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

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

and

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

[26]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 as a moves away from $a(\theta)$. Quah and Strulovici (2009) refer to these preferences as *quasi-concave with increasing peaks*.

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, $\liminf_B \mathrm{E}_B(u) \geqslant \mathrm{E}(u)$, where E and 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 $\mathrm{E}_B(u) \leqslant \mathrm{E}(u)$ for every B, we are done.

B Proofs

B.1 Maximal Selection

Proof of Theorem 1. By Theorem 0 and Theorems 5.1 and 5.2 in Lehmann (1988) it suffices to show that $\log(-\log F)$ is concave (convex) if and only if for all real numbers $k > m$ the distribution F^k is less (more) dispersed than F^m, that is, for all ε_m and ε_k such that $F^k(\varepsilon_k) = F^m(\varepsilon_m)$,

$$mF^{m-1}(\varepsilon_m)f(\varepsilon_m) \leqslant (\geqslant)\, kF^{k-1}(\varepsilon_k)f(\varepsilon_k). \tag{18}$$

Let $\lambda(\cdot) := \log(-\log(\cdot))$ and note that $\lambda(F^k(\cdot))$ and $\lambda(F^m(\cdot))$ only differ from $\lambda(F(\cdot))$ by a constant: $\lambda(F^k(\cdot)) - \log k = \lambda(F^m(\cdot)) - \log m = \lambda(F(\cdot))$. Differentiating, we obtain

$$\lambda'(F^m(\cdot))mF^{m-1}(\cdot)f(\cdot) = \lambda'(F^k(\cdot))kF^{k-1}(\cdot)f(\cdot) = \lambda'(F(\cdot))f(\cdot).$$

Thus, (18) is equivalent to

$$\frac{\lambda'(F(\varepsilon_m))f(\varepsilon_m)}{\lambda'(F^m(\varepsilon_m))} \leqslant (\geqslant)\, \frac{\lambda'(F(\varepsilon_k))f(\varepsilon_k)}{\lambda'(F^k(\varepsilon_k))}.$$

Since $F^k(\varepsilon_k) = F^m(\varepsilon_m)$ and $\lambda(\cdot)$ is strictly decreasing, the denominators are negative and coincide, so the inequality is equivalent to $\lambda'(F(\varepsilon_m))f(\varepsilon_m) \geqslant (\leqslant)\, \lambda'(F(\varepsilon_k))f(\varepsilon_k)$. This holds for all ε_m and all real $k > m$ if and only if $\lambda(F(\cdot))$ is concave (convex), because $k > m$ implies $\varepsilon_k \geqslant \varepsilon_m$. $\square$

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 $k_t := tk + (1-t)m$. For every state θ, let $\bar{x}_\theta$ and $\underline{x}_\theta$ respectively denote the upper bound and the lower bound of the support of $f(\cdot|\theta)$. Note that the support of $X(t)$ in state θ does not depend on t. In particular, $S(X(t)|\theta) = \{x \in \mathbb{R}^n : \bar{x}_\theta > x_1 > \cdots > x_n > \underline{x}_\theta\}$, and moreover $S(X_i(t)|\theta, x_{<i}) = (\underline{x}_\theta, x_{i-1})$ for every $i > 1$ and every $x_{<i} \in \{(x_1, \ldots, x_{i-1}) \in \mathbb{R}^{i-1} : \bar{x}_\theta > x_1 > \cdots > x_{i-1} > \underline{x}_\theta\}$. Fix two indices $s < t$ and two states $\theta < \theta'$, and write $\varphi_\theta(x) = z$ and $\varphi_{\theta'}(x) = z'$ for brevity. Since the support of $X_1(t)$ in state θ does not depend on t, as x_1 converges to $\bar{x}_\theta$, so does z_1. Similarly, for every $i = 2, \ldots, n$ and every x_{i-1}, as x_i converges to x_{i-1}, z_i converges to z_{i-1}.

We must prove that under either condition in the theorem ($k \geqslant m$ and $f(\cdot|\theta)/F(\cdot|\theta)$ is log-supermodular, or $m \geqslant k$ and $f(\cdot|\theta)/F(\cdot|\theta)$ is log-submodular with support of $f(\cdot|\theta)$ independent of θ) for every $x \in S(X(t)|\theta) \cap S(X(t)|\theta')$ we have the accuracy ordering $z' \leqq z$, or equivalently

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

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

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

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

$$\frac{F^{k_s}(z_1|\theta')}{F^{k_t}(x_1|\theta')} \geqslant 1 \quad \text{and} \quad \frac{F^{k_s-i+1}(z_i|\theta')}{F^{k_t-i+1}(x_i|\theta')} \geqslant \frac{F^{k_s-i+1}(z_{i-1}|\theta')}{F^{k_t-i+1}(x_{i-1}|\theta')} \quad \text{for } i=2,\dots,n. \tag{19}$$

Since $x_1 \to \bar{x}_\theta$ implies $z_1 \to \bar{x}_\theta$, under either condition in the theorem as $x_1 \to \bar{x}_\theta$ the left-hand side of the first inequality in (19) tends to a number no smaller than one. This implies that the first inequality in (19) holds if the left-hand side of the inequality decreases with x_1. Differentiating with respect to x_1 and dropping the positive denominator in the derivative, we need

$$k_s F^{k_s-1}(z_1|\theta') f(z_1|\theta') \frac{dz_1}{dx_1} F^{k_t}(x_1|\theta') \leqslant k_t F^{k_t-1}(x_1|\theta') f(x_1|\theta') F^{k_s}(z_1|\theta'). \tag{20}$$

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 this expression in (20) and simplifying, the first inequality in (19) holds if

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

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

Take any $i > 1$ now. Recall that for 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 (19) tends to a number no smaller than the right-hand side. The second inequality in (19) then holds if its left-hand side decreases with x_i. Differentiating with respect to x_i and simplifying,

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

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

B.2 Logconcavity of Reverse Hazard Rate and Reverse Hazard Function

The reverse hazard function is the right-sided integral of the reverse hazard rate: $-\log F(\varepsilon) = \int_\varepsilon^\infty (f(\varepsilon)/F(\varepsilon))d\varepsilon$. The reverse hazard function therefore inherits logconcavity (and logconvexity, if the support of f is unbounded above) of the reverse hazard rate (An, 1998, Lemma 3). Thus, the hypotheses in Corollary 1 are stronger than the corresponding conditions in Theorem 1; for example, the reverse hazard function of distribution $F(\varepsilon) = \varepsilon - 1/(1+e^{-\varepsilon}) + \log(1+e^{-\varepsilon})$ is logconcave, but the reverse hazard rate is not.[27] Still, the examples discussed after Theorem 1—normal, logistic, generalized exponential, shifted Gompertz—satisfy the hypotheses in the corollary. 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 $f(\varepsilon)/F(\varepsilon) = 1/(e^{\varepsilon}+1)$ is logconcave. With generalized exponential noise, $f(\varepsilon)/F(\varepsilon) = (1-e^{-\varepsilon})^a e^{-\varepsilon}$ 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 $e^{3\varepsilon} + \eta(e^{2\varepsilon} - 1)$.

B.3 Extreme Selection

Proof of Theorem 3. We start by showing that under UHR we have $\alpha_k \to 0$ as $k \to \infty$. 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, by Proposition 1.13 in Resnick (2008) we can set $\alpha_k = \bar{\varepsilon} - F^{-1}(1-1/k)$, whence $\alpha_k \to 0$ follows. In the second case, as well as (Müller and Rufibach, 2008, Lemma 3.5) in the case $\bar{\varepsilon} = \infty$, the limiting distribution is Gumbel. Thus, by Proposition 1.9 in Resnick (2008), we can set α_k to be the mean residual life evaluated at $\bar{\varepsilon}_k := F^{-1}(1-1/k)$, that is, $\alpha_k = k\int_{\bar{\varepsilon}_k}^{\bar{\varepsilon}} \varepsilon f(\varepsilon) d\varepsilon$. As shown in Calabria and Pulcini (1987), the limiting behavior of the mean residual life is the same as the limiting behavior of the inverse of the hazard rate.[28] Thus, using the fact that $\bar{\varepsilon}_k \to \bar{\varepsilon}$ 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, $\overline{U} := \int_{\Theta} \max_a u(\theta,a) dQ(\theta)$, as $k \to \infty$. Clearly, if the conclusion holds for $n = 1$ then a fortiori it holds for $n > 1$. Thus, we can assume $n = 1$. 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, this observation implies that if $\theta_j < \theta_{j-1}$ then a_j can be removed from A without affecting the IDO property. Moreover, a_j is never optimal at any state, so it is never used under full information. Thus, we may assume without loss of generality that $\theta_j \geqslant \theta_{j-1}$ for all $j > 1$, and the full information payoff can then be written, summing by parts, as $\overline{U} = \int_{\Theta} \sum_{j<J} \mathbf{1}_{\{\theta \geqslant \theta_j\}} \left[u(\theta,a_{j+1}) - u(\theta,a_j)\right] dQ(\theta)$. Fix $\delta > 0$, and let $\eta > 0$ be such that

$$\int_{\Theta} \sum_{j<J} \mathbf{1}_{\{\theta_j - \eta \leqslant \theta < \theta_j\}} \left[u(\theta,a_{j+1}) - u(\theta,a_j)\right] dQ(\theta) \leqslant \frac{\delta}{2} \tag{21}$$

[27] The support of the density $f(\varepsilon) = (1+e^{-\varepsilon})^{-2}$ is the interval $(-\infty, \bar{\varepsilon}]$, where $\bar{\varepsilon}$ solves $F(\bar{\varepsilon}) = 1$.

[28] Calabria and Pulcini (1987) assume that the support of f is bounded below. But for every ε such that $0 < F(\varepsilon) < 1$ the hazard rate of distribution F is the same as the hazard rate of the left-truncated distribution $F(\cdot)/[1-F(\varepsilon)]$. Furthermore, the two distributions have the same right tails, and hence the same limiting distribution $\bar{F}$.

and, furthermore,

$$(1-\eta)\int_{\Theta}\sum_{j<J}\mathbf{1}_{\{\theta\geqslant\theta_j\}}\big[u(\theta,a_{j+1})-u(\theta,a_j)\big]\,dQ(\theta)\geqslant\overline{U}-\frac{\delta}{2}. \tag{22}$$

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,

$$\begin{aligned}\Pr_\theta(\theta-\eta+\beta_k\leqslant X\leqslant\theta+\eta+\beta_k)&\geqslant\Pr_\theta(\theta-\alpha_k\bar{\varepsilon}+\beta_k\leqslant X\leqslant\theta+\alpha_k\bar{\varepsilon}+\beta_k)\\&=F^k(\alpha_k\bar{\varepsilon}+\beta_k)-F^k(\alpha_k\bar{\varepsilon}+\beta_k)\\&\geqslant\hat{F}(\bar{\varepsilon})-\frac{\eta}{4}-\hat{F}(-\bar{\varepsilon})-\frac{\eta}{4}\geqslant 1-\eta,\end{aligned} \tag{23}$$

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

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

By (21), (22) and (23), this payoff is at least as large as $\overline{U}-\delta$.

Finally, we show that if UHR fails, limit welfare is the welfare from an experiment with noise density (6). Failure of UHR implies that $\bar{\varepsilon}=\infty$ and hence (as established earlier) that $\bar{F}$ is Gumbel, with $\lim_{k\to\infty}\alpha_k=\lim_{\varepsilon\to\bar{\varepsilon}}[1-F(\varepsilon)]/f(\varepsilon)=:\alpha>0$.[29] Thus, as shown by Weissman (1978) and Leadbetter et al. (1983, Theorem 2.3.1), the distribution of the n largest of k draws from F, after normalizing each draw's location by β_k, converges weakly to a distribution with density (6). □

B.4 Applications

Proof of Proposition 2. Let $\bar{n}=\arg\max_{n\geqslant 1}U(n,n)-C_S(n)-C_P(n)$. Note that $\bar{n}$ exists because $U(n,n)$ is bounded above by the evaluator's full information payoff, $\overline{U}$, while $C_S(n)$ is unbounded. Furthermore, $U(\bar{n},\bar{n})-C_S(\bar{n})-C_P(\bar{n})<\overline{U}-C_S(1)$, because $C_S(n)$ is increasing. By Theorem 3, $U(k,1)\to\overline{U}$ as $k\to\infty$. Thus, there exist $\bar{k}>1$ and $c>0$ such that

$$U(\bar{k},1)-C_S(1)-c>U(\bar{n},\bar{n})-C_S(\bar{n})-C_P(\bar{n}). \tag{24}$$

If $C_P(\bar{k})\leqslant c$ then $U(\bar{k},1)-C_S(1)-C_P(\bar{k})>U(\bar{n},\bar{n})-C_S(\bar{n})-C_P(\bar{n})$, hence no experiment format with $k=n$ is optimal. □

[29] Logconcavity of f implies increasing hazard rate $f/(1-F)$, so this limit exists when UHR fails.

Proof of Proposition 3. Define $\bar{n}$ as in the proof of Proposition 2. By Theorem 3, $U(k,1) \to \overline{U}$ as $k \to \infty$. Since $C_S(n)$ is increasing, there exists $\bar{k} \geqslant 1$ such that

$$U(k,1) - C_S(1) > U(\bar{n},\bar{n}) - C_S(\bar{n}) \qquad \text{for all } k \geqslant \bar{k}. \tag{25}$$

Let $\bar{x}_2(k'), \ldots, \bar{x}_J(k')$ denote the cutoffs the evaluator sets when conjecturing presample size k'. Given these cutoffs, the expected gross payoff the sender obtains by choosing presample size k is $V(k,k') := \int_\Theta \sum_{j<J} [1 - F^k(\bar{x}_{j+1}(k'))][v(\theta,a_{j+1}) - v(\theta,a_j)]dQ(\theta)$, where $v(\theta,a)$ denotes the sender's payoff function and we used summation by parts and disregarded constants. Thus, going from k to $k+1$ the sender incurs marginal cost $C_P(k+1) - C_P(k)$ and obtains marginal gain

$$V(k+1,k') - V(k,k') = \int_\Theta \sum_{j<J} F^k(\bar{x}_{j+1}(k'))[1 - F(\bar{x}_{j+1}(k'))][v(\theta,a_{j+1}) - v(\theta,a_j)]dQ(\theta) > 0.$$

Pick any $c > 0$ such that $c < V(k+1,k) - V(k,k)$ for all k in the set $\{1, \ldots, \bar{k}\}$, and suppose that $C_P(\bar{k}) \leqslant c$ for all k. Then, since $C_P(k)$ is increasing, $C_P(k+1) - C_P(k) < c$ for every $k < \bar{k}$. Thus, in every Bayes Nash equilibrium the sender chooses $k > \bar{k}$, and the result follows from (25). □

Equilibrium Existence in the Delegated Presampling Game. Assume that presample size can be any real $k \geqslant 1$ and $C_P(k)$ is convex and hence, in particular, continuous and unbounded. Since $C_P(k)$ is unbounded, we can without loss restrict the sender's strategy set to some interval $[1,\hat{k}]$, e.g. choosing $\hat{k}$ large enough that $\int_\Theta v(\theta,a_1)dQ(\theta) - C_P(1) \geqslant \int_\Theta v(\theta,a_J)dQ(\theta) - C_P(\hat{k})$. Thus, we can also assume the evaluator strategy set to be the set of optimal monotone strategies for selected experiments with $k \in [1,\hat{k}]$. Moreover, identifying the optimal strategy for k with k itself, we obtain a game where the strategy set of each player is the interval $[1,\hat{k}]$ and the evaluator best response is the identity function. Thus, an equilibrium exists if the sender best response function

$$k \mapsto \arg\max_{k' \in [1,\hat{k}]} \int_\Theta \sum_{j<J} \left[1 - F^k(\bar{x}_{j+1}(k') - \theta)\right]\left[v(\theta,a_{j+1}) - v(\theta,a_j)\right]dQ(\theta) - C_P(k')$$

has a fixed point. (This is a function by the objective function's strict concavity: $C_P(k)$ is convex and, moreover, $v(\theta,a_{j+1}) > v(\theta,a_j)$ 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.) This follows from Brouwer's fixed point theorem.

Proof of Proposition 4. By Theorem 3, $U(k,1) \to \overline{U}$ and hence, since $C_S(n)$ is strictly increasing, there exists $\bar{k}$ such that $\overline{U} - U(k,1) < C_S(n) - C_S(1)$ for every $k \geqslant \bar{k}$ and $n > 1$. As shown in the proof of Proposition 3, there exists c such that if $C_P(\bar{k}) \leqslant c$ then in every Bayes Nash equilibrium of the delegated presampling game the sender must choose $k > \bar{k}$. Fix such a presampling cost function, and let $\bar{n}$ be the corresponding presample size (equal to sample size) chosen by the sender in an unraveling equilibrium of the full delegation game. The payoff of the evaluator in this equilibrium is $U(\bar{n},\bar{n}) - C_S(\bar{n}) \leqslant \overline{U} - C_S(\bar{n}) < U(k,1) - C_S(1)$ for every $k \geqslant \bar{k}$, so we are done. □

B.5 Other Forms of Selection

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

$y \geqslant a_t$. Fix $s < t$ and consider the function $\varphi_\theta(\cdot)$, which is defined as follows:

$$[F(\varphi_\theta(y)|\theta) - F(a_s|\theta)]/[1 - F(a_s|\theta)] = [F(y|\theta) - F(a_t|\theta)]/[1 - F(a_t|\theta)],$$

or equivalently

$$[1 - F(\varphi_\theta(y)|\theta)]/[1 - F(y|\theta)] = [1 - F(a_s|\theta)]/[1 - F(a_t|\theta)], \tag{26}$$

for $y \geqslant a_t$. Now fix two states $\theta' > \theta$. We must show that $\varphi_{\theta'}(y) \leqslant \varphi_\theta(y)$ for every $y \geqslant a_t$. Using the definition of $\varphi_{\theta'}(\cdot)$, it suffices to show that

$$[F(y|\theta') - F(a_t|\theta')]/[1 - F(a_t|\theta')] \leqslant [F(\varphi_\theta(y)|\theta') - F(a_s|\theta')]/[1 - F(a_s|\theta')],$$

that is,

$$[1 - F(y|\theta')]/[1 - F(a_t|\theta')] \geqslant [1 - F(\varphi_\theta(y)|\theta')]/[1 - F(a_s|\theta')].$$

Since $\varphi_\theta(y)$ converges to a_s as y decreases to the lower bound a_t, it suffices to prove that the ratio between right-hand and left-hand side decreases with y. Taking derivatives and using (26), we need

$$\frac{f(\varphi_\theta(y)|\theta')/[1 - F(\varphi_\theta(y)|\theta')]}{f(\varphi_\theta(y)|\theta)/[1 - F(\varphi_\theta(y)|\theta)]} \geqslant \frac{f(y|\theta')/[1 - F(y|\theta')]}{f(y|\theta)/[1 - F(y|\theta)]}.$$

This inequality holds when the hazard rate is log-supermodular because $Y(s)$ first-order stochastically dominates $Y(t)$ and hence $\varphi_\theta(y) \geqslant y$.

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 $\varphi_\theta(\cdot|)$, which is defined as follows: for every $w \leqslant b_s$,

$$F(\varphi_\theta(w)|\theta)/F(b_s|\theta) = F(w|\theta)/F(b_t|\theta).$$

We must show that if $\theta' > \theta$ then $\varphi_{\theta',t,s}(w) \leqslant \varphi_\theta(w)$ for all $w \leqslant b_t$. Using the definition of $\varphi_{\theta',t,s}(\cdot)$, it suffices to show that

$$F(w|\theta')/F(b_t|\theta') \leqslant F(\varphi_\theta(w)|\theta')/F(b_s|\theta').$$

Given that $\varphi_\theta(w)$ converges to b_s as w increases to the upper bound b_t, it is enough to prove that the ratio between the right-hand side and the left-hand side of the inequality decreases with w. Taking derivatives, this condition says that

$$\frac{f(\varphi_\theta(w)|\theta')/F(\varphi_\theta(w)|\theta')}{f(\varphi_\theta(w)|\theta)/F(\varphi_\theta(w)|\theta)} \leqslant \frac{f(w|\theta')/F(w|\theta')}{f(w|\theta)/F(w|\theta)}.$$

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

Proof of Theorem 5. When payoffs satisfy Karlin and Rubin's (1956) monotonicity, Theorem 0 holds for families $X(t)$ with t in an arbitrary ordered set T, as shown in Appendix A. Thus, to prove part (a) it suffices to take $T = \{0,1\}$, with $X(0)$ the random experiment and $X(1)$ the median selected 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 $\varphi_\theta(x)$ is defined by

$$F(\varphi_\theta(x)|\theta) = F^r(x|\theta)\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}[1-F(x|\theta)]^j. \tag{27}$$

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

$$dz/dx = cF^{r-1}(x|\theta)[1-F(x|\theta)]^{r-1}f(x|\theta)/f(z|\theta). \tag{28}$$

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

$$F^r(x|\theta')\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}[1-F(x|\theta')]^j/F(z|\theta') \leqslant 1, \tag{29}$$

which is the same as

$$[1-F(x|\theta')]^r\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}F^j(x|\theta')/1-F(z|\theta') \geqslant 1. \tag{30}$$

Suppose first that $x \leqslant x_m$, so that $z \leqslant x$. Since $F(\cdot|\theta')$ first-order stochastically dominates $F(\cdot|\theta)$, condition (29) holds at $x = x_m = z$. Thus, it suffices to show that the left-hand side of (29) increases in x when $x \geqslant z$. The derivative of the left-hand side is nonnegative if and only if

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

Plugging in (28) and using (27), the inequality is the same as

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

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

Suppose now that $x \geqslant x_m$, so that $z \geqslant x$. Since (30) is the same as (29), it holds at $x = x_m = z$, so it suffices to show that its left-hand side increases in x when $x \leqslant z$. The derivative of the left-hand side of (30) is nonnegative if and only if

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

Plugging in (28) and using (27), the inequality is the same as

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

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

Turning to the proof of part (b), we exhibit a simple hypothesis testing problem with states $\theta' > \theta$ where $X(1)$ gives higher welfare than $X(0)$. Let $\bar{x}$ be such that $F(\bar{x}|\theta') < 1/2 < F(\bar{x}|\theta)$ and suppose that under $X(0)$ the evaluator sets the cutoff at $\bar{x}$.[30] Then keeping the cutoff at $\bar{x}$ the payoff is higher when the experiment changes to $X(1)$. Indeed, $F(\bar{x}|\theta) > 1/2$ implies

$$F^r(\bar{x}|\theta)\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}[1-F(\bar{x}|\theta)]^j > F(\bar{x}|\theta),$$

while $F(\bar{x}|\theta') < 1/2$ implies

$$F^r(\bar{x}|\theta')\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}[1-F(\bar{x}|\theta')]^j < F(\bar{x}|\theta').$$

□

Proof of Proposition 5. Let $i_k = \arg\max_{1 \leqslant i \leqslant k} \delta_i$. By Theorem 0 and Theorems 5.1 and 5.2 in Lehmann (1988) it suffices to prove that if $-\log F_\delta(\cdot)$ is logconcave (logconvex) then the distribution of $c\delta_{i_k} + \gamma_{i_k}$ becomes less (more) dispersed as k increases. As shown in the proof of Theorem 1, this is true for the distribution of $c\delta_{i_k}$. By Theorem 8 in Lewis and Thompson (1981), it is also true for the distribution of $c\delta_{i_k} + \gamma_1$, which is the same as the distribution of $c\delta_{i_k} + \gamma_{i_k}$. □

Proof of Proposition 6. Define $\hat{\theta} = \theta + \gamma_1$ and note that, since γ_1 has a logconcave density, the conditional density of $\hat{\theta}$ given θ satisfies the MLR property. This implies that the posterior belief

[30] To see why $\bar{x}$ exists, write $S(\theta)$ and $S(\theta')$ for the supports of $X(0)$ (and $X(1)$) in states θ and θ'. By MLR, $F(x|\theta') < F(x|\theta)$ for all $x \in S(\theta') \cap S(\theta)$. We can pick $\bar{x}$ right above the median of $F(\cdot|\theta)$ if this median is in the closure of $S(\theta')$, or right below the median of $F(\cdot|\theta')$ if this is in the closure of $S(\theta)$. If both of these tries fail, any $x \in S(\theta') \cap S(\theta)$ must have $F(x|\theta') < 1/2 < F(x|\theta)$, and we can let $\bar{x}$ be any of these points.

$q(\theta|\hat{\theta})$ increases with $\hat{\theta}$ in the likelihood ratio order. Thus, defining $\hat{u}(\hat{\theta},\cdot) := \int_{\Theta} u(\theta,\cdot)dQ(\theta|\hat{\theta})$ for every $\hat{\theta}$, it follows from Theorem 2 in Quah and Strulovici (2009) that the family of functions $\hat{u}(\hat{\theta},\cdot)$ is an IDO family. Consider the auxiliary decision problem where the state is $\hat{\theta} = \theta + \gamma_1$, the payoff function is $\hat{u}(\hat{\theta},\cdot)$, and the selected experiment is $(\hat{\theta}+c\delta_1, \hat{\theta}+c\delta_n)$, where $c\delta_1 \geqslant \cdots \geqslant c\delta_n$ are the n highest of k random draws from distribution $F_\delta(\cdot/c)$ with density $(1/c)f_\delta(\cdot/c)$. Clearly, $(1/c)f_\delta(\cdot/c)/F_\delta(\cdot/c)$ is logconcave (logconvex, with support of $(1/c)f_\delta(\cdot/c)$ unbounded above) if and only if $f_\delta(\cdot)/F_\delta(\cdot)$ is logconcave (logconvex, with support of $f_\delta(\cdot)$ unbounded above), in which case, by Corollary 1, welfare increases (decreases) in k. The result now follows from the fact that the payoff from experiment $(\hat{\theta}+c\delta_1, \hat{\theta}+c\delta_n)$ in the auxiliary problem is the same as the payoff from experiment $(\theta+c\delta_1+\gamma_1, \cdots, \theta+c\delta_n+\gamma_1)$ in the original problem. $\square$

References

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 (2020): "A Theory of Experimenters: Robustness, Randomization, and Balance," *American Economic Review*, 110(4), 1206–1230.

Barlow, R. E. and F. Proschan (1975): *Statistical Theory of Reliability and Life Testing: Probability Models*, New York, Holt, Rinehart and Winston.

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.

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. (1951): "Comparison of Experiments," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 93–102.

——— (1953): "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, 24, 265–272.

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

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.

Chassang, S., G. Padró i Miquel, and E. Snowberg (2012): "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments," *American Economic Review*, 102(4), 1279–1309.

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.

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.

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

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.

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 (2020): “Persuasion Through Selective Disclosure: Implications for Marketing, Campaigning, and Privacy Regulation,” *Management Science*, 66(1), 4958–4979.

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 Y. Rinott (1980): “Classes of Orderings of Measures and Related Correlation Inequalities. I. Multivariate Totally Positive Distributions,” *Journal of Multivariate Analysis*, 10(4), 467–498.

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.

Lewis, T. and J. W. Thompson (1981): “Dispersive Distributions, and the Connection between Dispersivity and Strong Unimodality,” *Journal of Applied Probability*, 18, 76–90.

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

Matthews, S. and A. Postlewaite (1985): “Quality Testing and Disclosure,” *Rand Journal of Economics*, 16(3), 328–340.

Meyer, M. (2017): “Assignment Protocols with Informational Frictions,” ESSET Presentation.

Milgrom, P. (1979): “A Convergence Theorem for Competitive Bidding with Differential Information,” *Econometrica*, 47(3), 679–688.

——— (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.

Milgrom, P. and R. Weber (1982): “A Theory of Auctions and Competitive Bidding,” *Econometrica*, 50(5), 1089–1122.

——— (2000): “A Theory of Auctions and Competitive Bidding, II,” in *The Economic Theory of Auctions*, ed. by P. Klemperer, Edward Elgar Publishing, Ltd, vol. II, 179–194.

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.

Noe, T. (2020): “Comparing the Chosen: Selection Bias When Selection Is Competitive,” *Journal of Political Economy*, 128(1), 342–390.

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

Polinsky, A. M. and S. Shavell (2012): “Mandatory Versus Voluntary Disclosure of Product Risks,” *Journal of Law, Economics, and Organization*, 28(2), 360–379.

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. (1995): “Subverting Randomization in Controlled Trials,” *Journal of the American Medical Association*, 274(18), 1456–1458.

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 J. G. Shanthikumar (2007): *Stochastic Orders*, Springer.

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

van Zwet, W. R. (1964): *Convex Transformations of Random Variables*, Amsterdam: Mathematisch Centrum.

Veinott, A. F. (1965): "Optimal Policy in a Dynamic, Single Product, Nonstationary Inventory Model with Several Demand Classes," *Operations Research*, 13(5), 761–778.

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

Weissman, I. (1978): "Estimation of Parameters and Large Quantiles Based on the k Largest Observations," *Journal of the American Statistical Association*, 73, 812–815.

Wilson, R. (1977): "A Bidding Model of Perfect Competition," *Review of Economic Studies*, 44(3), 511–518.