DISCUSSION PAPER SERIES

DP12202
(v. 2)

STRATEGIC SAMPLE SELECTION

Alfredo Di Tillio, Marco Ottaviani and Peter Norman Sørensen

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

STRATEGIC SAMPLE SELECTION

Alfredo Di Tillio, Marco Ottaviani and Peter Norman Sørensen

Discussion Paper DP12202
First Published 05 August 2017
This Revision 20 March 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

STRATEGIC SAMPLE SELECTION

Abstract

This paper develops a notion of multivariate accuracy to characterize the welfare impact of sample selection from a larger presample. Maximal selection benefits or hurts a decision maker with interval dominance ordered preferences if the reverse hazard rate of the data distribution is log-supermodular—as in location experiments with normal noise—or log-submodular. Applying the result to auctions, we show that under non-pathological conditions the information contained in the winning bids decreases as the number of bidders increases. Exploiting a connection to extreme value theory, we quantify the limit amount of information when the presample size goes to infinity, as under perfect competition. In the context of a model of equilibrium persuasion with costly information, we also derive implications for the 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.

JEL Classification: D82, D83, C72, C90

Keywords: Strategic selection, Persuasion, Comparison of experiments, Dispersion, welfare

Alfredo Di Tillio - alfredo.ditillio@unibocconi.it
Bocconi University

Marco Ottaviani - marco.ottaviani@unibocconi.it
Bocconi University and CEPR

Peter Norman Sørensen - peter.norman.sorensen@econ.ku.dk
University of Copenhagen

Acknowledgements
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, Danielle Li, Nicola Limodio, Marco Pagnozzi, Marek Pycia, Karl Schlag, Matt Shum 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.

Strategic Sample Selection*

Alfredo Di Tillio† Marco Ottaviani‡ Peter Norman Sørensen§

February 26, 2020

Abstract

This paper develops a notion of multivariate accuracy to characterize the welfare impact of sample selection from a larger presample. Maximal selection benefits or hurts a decision maker with interval dominance ordered preferences if the reverse hazard rate of the data distribution is log-supermodular—as in location experiments with normal noise—or log-submodular. Applying the result to auctions, we show that under non-pathological conditions the information contained in the winning bids decreases as the number of bidders increases. Exploiting a connection to extreme value theory, we quantify the limit amount of information when the presample size goes to infinity, as under perfect competition. In the context of a model of equilibrium persuasion with costly information, we also derive implications for the 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, Danielle Li, Nicola Limodio, Marco Pagnozzi, Marek Pycia, Karl Schlag, Matt Shum 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). Can selection add information? For example, how revealing of market demand are the winning bids in a more competitive auction? 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 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 evaluator 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 observing the realization of an experiment, consisting of a random vector $X = (X_1, \ldots, X_n)$ 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. Our basic question is, in which of the following scenarios does the evaluator make better decisions?

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

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? Being the maximum of k i.i.d. draws, the observation now has distribution $F^k(x - \theta_L)$ in the low state and $F^k(x - \theta_H)$ 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 $\bar{x}$.

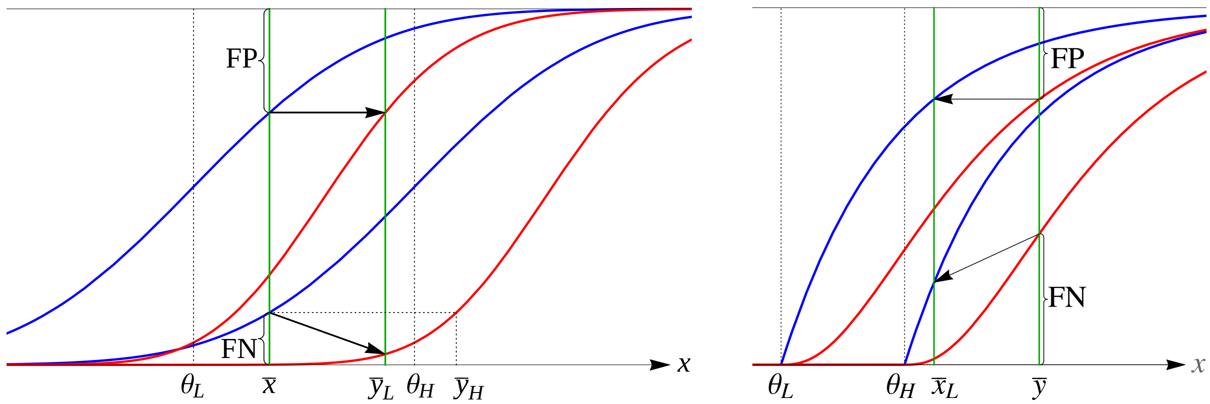

Figure 1: Maximal selection provides more accurate information with normal noise (left) but less accurate information with exponential noise (right).

high state—drawn in red. Maximal selection benefits the evaluator. By adopting the possibly sub-optimal cutoff $\bar{y}_L$ that induces as many false positives, the evaluator induces fewer false negatives: $F^k(\bar{y}_L - \theta_H) < F(\bar{x} - \theta_H)$. But, as shown in the right-hand panel of Figure 1, exactly the opposite is true with exponential noise: given any cutoff $\bar{y}$ for the selected experiment, the evaluator can match false positives and lower false negatives by adopting $\bar{x}_L$ 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.[2] As illustrated in the left-hand panel of Figure 1, an equivalent way to formulate the property that $\bar{y}_L$ induces as many false positives and fewer false negatives is to say that the selected experiment is *more accurate*. This means that the cutoff point $\bar{y}_H$ inducing as many false *negatives*, defined by $F^k(\bar{y}_H - \theta_H) = F(\bar{x} - \theta_H)$, is larger than $\bar{y}_L$. By adopting the smaller cutoff $\bar{y}_L$, the evaluator necessarily induces more acceptance—and in particular more acceptance in the high state—than by adopting the larger cutoff $\bar{y}_H$. In effect, this is what happens in the normal case (and with opposite direction in the exponential case) depicted in Figure 1.

Our first core result, Theorem 1, identifies a necessary and sufficient condition for a larger pre-sample 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 logcon-

[2] Accuracy can be defined as Blackwell's (1951,1953) sufficiency, restricted to monotone decision problems and unidimensional experiments satisfying monotone likelihood ratio. This class of problems was first studied by Karlin and Rubin (1956). The term *accuracy* was introduced by Persico (2000).

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

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

Table 1: Roadmap of results

cave 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 distribution is "more convex" (precisely, the reverse hazard function is more logconcave) than Gumbel.

To characterize the impact of selection in the general case, we develop a multidimensional version of accuracy, sharing the basic intuition with (and for $n = 1$, reducing 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 a more complicated object, an $(n-1)$-dimensional hypersurface.[3] Similarly to the unidimensional case, we say Y is more accurate than X if a suitably defined hypersurface that induces as many false positives in Y as in X lies below the one inducing as many false negatives. Theorem 0 proves that welfare increases with accuracy, extending previous results by Persico (2000), Jewitt (2007) and Quah and Strulovici (2009).

Our notion of accuracy is the key technical tool needed to tackle the new issues arising in the multidimensional case. The main difficulty lies in the fact that 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 presample size, according to whether the reverse hazard rate $f(x|\theta)/F(x|\theta)$ is log-supermodular or log-submodular. In a location experiment, log-supermodularity reduces to logconcavity of the noise distribution's reverse hazard rate f/F, strengthening the logconcavity criterion in Theorem 1.

Drawing on extreme value theory (Fisher and Tippett, 1928; Gnedenko, 1943), in Section 4 we 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

[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 $\tilde{x}$.

experiment with scale parameter (proportional to variance) equal to the inverse of the limit hazard rate—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). In Section 5 we consider 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 contained in the winning bids. In particular, in non-pathological cases in which bidders' signals have log-submodular reverse hazard rate, competition decreases information—overturning received wisdom. Moreover, we quantify the amount of information in the perfectly competitive limit.

Next, we consider 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). Taking a design perspective, what is the optimal experiment for the evaluator, when sampling and presampling are costly and endogenous? Section 7 illustrates how the evaluator can use presample size as an additional information channel, to economize on sample size when selection is beneficial. Building on Theorem 3, Proposition 2 shows that with sufficiently small presampling costs the optimal experiment must feature sample selection.

Motivated by our researcher bias application, in Section 7 we also consider a 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 constrained by presample data. The choice of presample size is akin to agent effort in Holmström's (1999) career concern model.[4] The wrinkle is that the sender's rat race to collect more presample data results in additional information that indirectly benefits or hurts the evaluator through maximal selection. Smaller presampling costs strengthen this sender incentive, again providing an information channel that the evaluator can use to save on sampling costs (Proposition 3). If, instead, the sender cannot commit not to disclose all information collected, full unraveling equilibrium results, with the sender effectively disclosing the whole presample data as in Grossman (1981) and Milgrom (1981). By a logic similar to Proposition 3, we show that the evaluator has an incentive to block unraveling by committing to a fixed sample size (Proposition 4), again to save on sampling costs. For related but different takes on the incentives for information collection and manipulation see recent work by Henry (2009), Dahm, Gonzàlez, and Porteiro (2009), Felgenhauer and Schulte (2014), Herresthal (2017), Henry and Ottaviani (2019), and Hoffmann, Inderst, and Ottaviani (forthcoming).

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

[4] See Glaeser (2008) for a discussion of the relevance of rat race effects to understand research incentives and biases in data collection and analysis.

by Goel and DeGroot (1992). As Theorem 4 shows, in important cases maximal selection and truncation lead to very different conclusions. For example, with normal or logistic noise maximal selection benefits while truncation hurts the evaluator. In Section 8 we also analyze median selection, where the evaluator observes the median observation in a presample (Theorem 5), deriving conclusions for peremptory challenge. Finally, in Section 9 we extend our results to allow for noisy observation of presample as well as sample data (Propositions 5 and 6).

Complementary Approaches to Strategic 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 an early precursor to our modeling approach, Fishman and Hagerty's (1990) analysis of selective disclosure focuses on different questions in a setting with binary signals and sample size $n = 1$; 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). Chassang, Padró i Miquel, and Snowberg (2012) characterize the design of experiments when outcomes are affected by experimental subjects' unobserved actions. 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 (forthcoming) propose a theory of an ambiguity-averse researcher facing an adversarial evaluator. Noe (2020) characterizes the preservation of stochastic dominance and likelihood ratio orders under competitive selection. Instead, we focus on the impact of selection on the welfare of an evaluator who takes an ex post optimal action.

2 Setup

An evaluator chooses an action $a \in A \subseteq \mathbb{R}$ under uncertainty about a state $\theta \in \Theta \subseteq \mathbb{R}$, where Θ is a finite set or a (possibly unbounded) interval. The prior is represented by a density (or mass) function $\pi(\theta)$ and the payoff function is $u : \Theta \times A \to \mathbb{R}$. For now we take the action set to be finite $A = \{a_1, \ldots, a_J\}$ with $a_1 < \ldots < a_J$ and in Appendix B we give an extension to continuous actions.

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

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

whenever $u(\theta, a'') \geqslant u(\theta, a)$ for all actions a such that $a' \leqslant a \leqslant a''$. Equivalently, if action a'' is the best action in the interval $[a', a''] \cap A$ when the state is θ, then the (weak or strict) preference of a'' over each action in the interval continues to hold at every higher state θ'. As pointed out by Quah and Strulovici (2009), the IDO class includes both single-crossing preferences (Milgrom and Shannon, 1994) and monotone preferences à la Karlin and Rubin (1956).[5]

Experiments and Welfare. Before deciding, the evaluator observes the realization of an experiment: a random vector X in $\mathbb{R}^n$ having state-dependent distribution $G(\cdot|\theta)$ and density $g(\cdot|\theta)$ with monotone likelihood ratio (MLR): if $x' \geqq x$ then $g(x'|\theta)/g(x|\theta)$ is increasing in θ.[6] An important consequence of IDO and MLR is that the evaluator can without loss adopt a monotone strategy, where the action increases in the realization. By Bayes' rule, MLR implies that the posterior belief on the state increases with the observed realization x in the likelihood ratio order—for all x and $x' \geqq x$ the ratio $\pi(\theta|x')/\pi(\theta|x)$ increases with θ—and hence that the evaluator cannot lose by increasing the action in response to a higher realization (Quah and Strulovici, 2009, Theorem 2). 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 $\int_\Theta \sum_j \Pr_\theta(X \in E_j) u(\theta, a_j) \pi(\theta) d\theta$ can then be rewritten, summing by parts and disregarding constants, as

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

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. Here MLR means that the directional derivative $\lim_{\Delta \to 0} \log[g(\varepsilon_1 + \Delta, \ldots, \varepsilon_n + \Delta)/g(\varepsilon_1, \ldots, \varepsilon_n)]/\Delta$ of the noise distribution's density g is decreasing in ε. With $n = 1$, this is simply logconcavity of g.

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 $r = [\pi(\theta_L)/\pi(\theta_H)][u(\theta_L, a_L) - u(\theta_L, a_H)]/[u(\theta_H, a_H) - u(\theta_H, a_L)]$. 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 E. Given this, welfare rewrites (disregarding constants) as $-r \Pr_{\theta_L}(X \in E) - \Pr_{\theta_H}(X \notin E)$, 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

[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.

[6] The property is required over the set where the denominator is nonzero. Here and later, given $x = (x_1, \ldots, x_n)$ and $x' = (x'_1, \ldots, x'_n)$ we say that x' is larger than x, and write $x' \geqq x$, to indicate that $x'_i \geqslant x_i$ for every i.

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 distributions $F(\cdot|\theta)$ only. In this paper we are interested in experiments involving selected rather than random observations. In this scenario $G(\cdot|\theta)$ 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 the indices $1, \ldots, i-1$ to save on notation, for every x we have

$$G_1(x_1|\theta) = F^k(x_1|\theta) \quad \text{and} \quad G_i(x_i|\theta, x_{<i}) = \frac{F^{k-i+1}(x_i|\theta)}{F^{k-i+1}(x_{i-1}|\theta)} \quad \text{for all } i > 1.^{7} \tag{2}$$

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 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.

Comparing Experiments by Accuracy. To assess the welfare impact of selection we develop a natural multidimensional generalization of Lehmann's (1988) notion of accuracy, which can be used to compare experiments (not necessarily selected experiments) with the same dimension. Let $G(t, \cdot|\theta)$ be a family of state-dependent distributions on $\mathbb{R}^n$ parametrized by $t \in [0,1]$, each with MLR density having convex support, such that letting $\Pr_\theta(t, \cdot)$ be the corresponding measure on $\mathbb{R}^n$, for all E the function $\Pr_\theta(\cdot, E)$ is continuously differentiable. Denote the corresponding family of experiments by $X(t)$. 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 real presample size $tk + (1-t)m$.

For each state θ and pair of indices s, t in $[0,1]$ we define a mapping $\varphi_{s,t}(\cdot|\theta) : \mathbb{R}^n \to \mathbb{R}^n$ such that in state θ the random vector $\varphi_{s,t}(X(s)|\theta)$ has the same distribution as $X(t)$, as follows: $\varphi_{s,t}(x_1, \ldots, x_n|\theta) = (z_1, \ldots, z_n)$, where $z_1, \ldots, z_n$ are defined recursively as

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

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

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

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

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, assuming that Θ and A are compact and the distribution of $X(t)$ has the same compact support for all t and in each state. Theorem 0 extends the result to multidimensional experiments and allows unbounded or non-constant supports (as is necessarily the case e.g. in location experiments). We prove Theorem 0, and discuss further the difference among IDO, single-crossing and monotone preferences, in Appendix B. The proofs of all other results in the paper are in Appendix A.

The intuition for why accuracy increases welfare is essentially the same as in Lehmann (1988). Consider a simple hypothesis testing problem, and let E be the optimal acceptance set in experiment $X(s)$. As the experiment changes to $X(t)$ the evaluator can respond by moving the boundary of the acceptance set, point by point, via the mapping $x \mapsto \varphi_{s,t}(x|\theta_L)$. This means adopting $\varphi_{s,t}(E|\theta_L)$ as acceptance set. Since in state θ_L the random vectors $\varphi_{s,t}(X(s)|\theta_L)$ and $X(t)$ have the same distribution, false positives remain the same: $\Pr_{\theta_L}(X(t) \in \varphi_{s,t}(E|\theta_L)) = \Pr_{\theta_L}(X(s) \in E)$. But false negatives decrease: $\Pr_{\theta_H}(X(t) \in \varphi_{s,t}(E|\theta_L)) \geqslant \Pr_{\theta_H}(X(t) \in \varphi_{s,t}(E|\theta_H)) = \Pr_{\theta_H}(X(s) \in E)$. The equality follows from $\varphi_{s,t}(X(s)|\theta_H)$ and $X(t)$ having the same distribution in state θ_H, the inequality from $\varphi_{s,t}(\cdot|\theta_H) \geqslant \varphi_{s,t}(\cdot|\theta_L)$ and from E being an upper set.

Unidimensional Location Experiments: Accuracy and Dispersion. When $X(t)$ is a family of unidimensional location experiments with corresponding noise distributions $G(t,\cdot)$, we can equivalently appeal to the notion of dispersion (Bickel and Lehmann, 1979). A 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: (i) 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; (ii) 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 Karlin and Rubin's (1956) class. By Theorem 0, 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.

3 Monotone Impact of Selection

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

3.1 Unidimensional Location Experiments

We begin with 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.

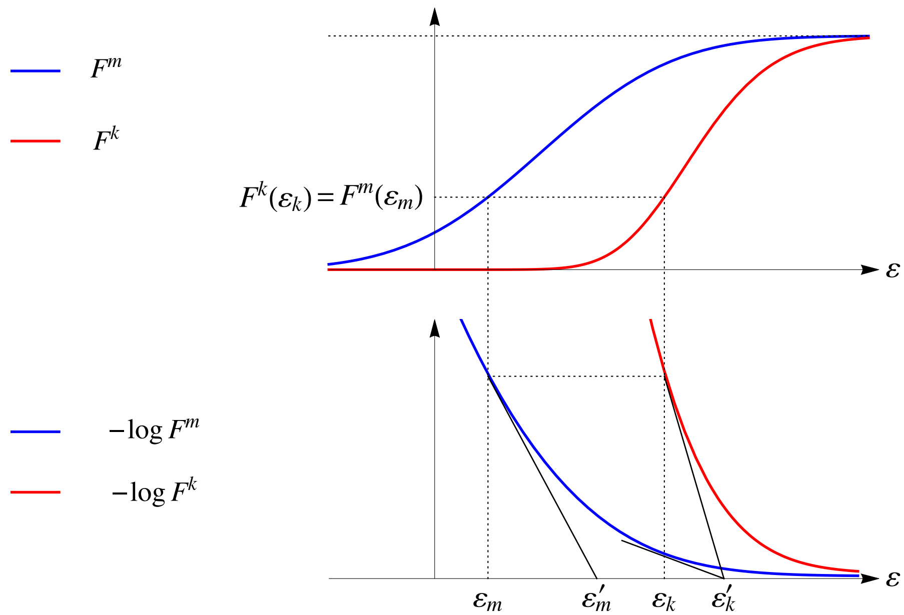

Figure 2: Normal noise: dispersion decreases with presample size (drawn for $k = 8$ and $m = 1$).

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 distribution's reverse hazard function $-\log F$ is logconcave (logconvex). Conversely, assuming that presample size can 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.*[8]

Figure 2 illustrates the proof with normal noise. Based on the equivalence between accuracy and dispersion, given two presample sizes $k > m$ we ask what makes F^k steeper at some quantile ε_k than F^m is at the corresponding quantile, namely the quantile ε_m defined by $F^m(\varepsilon_m) = F^k(\varepsilon_k)$. Since $-\log(\cdot)$ is monotone, this is equivalent to $-\log F^k$ being steeper at ε_k than $-\log F^m$ is at ε_m. Yet equivalently, the base $\varepsilon'_m - \varepsilon_m$ below the tangent to $-\log F^m$ at ε_m is larger than the base $\varepsilon'_k - \varepsilon_k$ below the tangent to $-\log F^k$ at ε_k. Since $-\log F^m$ and $-\log F^k$ only differ by a multiplicative constant factor, we can also compute $\varepsilon'_k - \varepsilon_k$ as the base below the tangent to $-\log F^k$. Thus, $\varepsilon'_k - \varepsilon_k$ is smaller than $\varepsilon'_m - \varepsilon_m$ when $-\log F$ is logconcave, because logconcavity of a decreasing function means that the function decays at an increasing rate: When k and m can be any real numbers, ε_k and ε_m can be arbitrarily close to each other, explaining necessity. The argument for $-\log F$ logconvex is analogous.

Gumbel Noise and Intuition. The only noise distribution F such that $-\log F$ is both logconcave and logconvex (loglinear) is the Gumbel extreme value distribution, $F(\varepsilon) = \exp(-\exp(-\varepsilon))$. With presample size k the noise distribution is $F^k(\varepsilon) = \exp(-k\exp(-\varepsilon)) = F(\varepsilon - \log k)$. Since

[8] 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.

selection only inflates noise by a constant ($\log k$), every maximally selected experiment gives the same welfare. Referring to Figure 2, with Gumbel noise the reverse hazard functions in the bottom panel would be both exponential—the bases $\varepsilon'_k - \varepsilon_k$ and $\varepsilon'_m - \varepsilon_m$ would be equal. Deviating from Gumbel, maximal selection changes the *shape* of the distribution. This change gives more (less) information, by making $\varepsilon'_k - \varepsilon_k$ smaller (larger) than $\varepsilon'_m - \varepsilon_m$, when the noise distribution is "more (less) convex" than Gumbel.

Logistic, Exponential and Shifted Gompertz Noise Distributions. 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 A. Our main example of the opposite case, where $-\log F$ is logconvex and hence more selection hurts, 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 distribution is the special case $a \to -1$. Finally, more selection hurts with Bemmaor's (1992) shifted Gompertz, $F(\varepsilon) = (1 - e^{-\varepsilon}\exp(-\eta e^{-\varepsilon}))$, as well as with the full support distribution $F(\varepsilon) = \exp(1 - \exp(e^{-\varepsilon}))$ introduced by Noe (2020).

Contribution to Stochastic Ordering of Order Statistics. Previous results on stochastic orderings of order statistics only covered distributions with decreasing hazard rate. Notably, Khaledi and Kochar (2000, Theorem 2.1) showed that for any distribution with decreasing hazard rate higher order statistics are more dispersed. Since logconcavity implies increasing hazard rate by Prekopa's theorem, the only distribution with logconcave density for which Khaledi and Kochar's (2000) result applies is the exponential (loglinear) distribution, which has constant hazard rate. As shown in the proof of Theorem 1, logconcavity (logconvexity) of the reverse hazard function characterizes when the distribution of the highest order statistics F^k is less (more) dispersed than F in general.

3.2 General Multidimensional Experiments

Extending our analysis to general (not necessarily location type) experiments with sample size $n \geqslant 1$ poses two related challenges. First, individual comparisons between order statistics do not provide our desired characterization. For example, in some cases the evaluator *is* better off with more selection, yet an intermediate order statistic—say, the second or third highest—is not, in isolation, more informative than a random draw; we present an instance of this fact after sketching the proof of Theorem 2. Second, order statistics are correlated, creating further ambiguity about the marginal value of information added by a single order statistic.[9] Our notion of accuracy allows us to characterize when and how this correlation adds or subtracts value to the evaluator's problem as presample size increases.

Theorem 2. *For fixed sample size $n \geqslant 1$, an increase in presample size increases (decreases) welfare if the reverse hazard rate $f(\cdot|\theta)/F(\cdot|\theta)$ is log-supermodular (log-submodular, with upper*

[9] Shaked and Tong (1990, 1993) identify conditions under which an experiment with correlated draws is less informative than an i.i.d. experiment, assuming equal marginal distributions. Their results do not apply to our context.

bound of the support of $f(\cdot|\theta)$ independent of θ), that is, if for all states θ and $\theta' > \theta$ the reverse hazard rate ratio

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

is increasing (resp. decreasing).

In location experiments, log-supermodularity of the reverse hazard rate is the same as 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 distribution's reverse hazard rate $f(\cdot)/F(\cdot)$ is logconcave (logconvex, with support of f unbounded above).*

After proving Theorem 2, in Appendix A we show 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 applies to all examples discussed earlier. The Gumbel distribution is sandwiched between the distributions for which more selection benefits and those for which more selection hurts: $f(\varepsilon)/F(\varepsilon) = \exp(-\varepsilon)$ is loglinear. More selection benefits with normal or logistic noise and hurts with exponential or shifted Gompertz noise.

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 $\theta' > \theta$, defining $z = \varphi_{s,t}(x|\theta)$ as in (3), we have

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

(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 $X_{<i}(s) = x_{<i}$, or observes $X_i(t)$ when already knowing that $X_{<i}(t) = z_{<i}$. Suppose that the acceptance cutoff is set at x_i in the first experiment. Since $X(t)$ and $\varphi_{s,t}(X(s)|\theta)$ have the same distribution in the low state θ, setting the cutoff at z_i in the second experiment gives as many false positives. By (4), cutoff z_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 greater than one in the limit as x_i grows large and (ii) monotonically increasing 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 $z = \varphi_{s,t}(x|\theta)$, we can write requirement (ii) more revealingly in terms of reverse hazard rates:

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

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:

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

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 $z \geqq x$. As we remark next, order statistics are CIS, so (5) can in fact be used.

CIS and Affiliated Random Variables. 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 the economics literature as *affiliation* (Milgrom and Weber, 1982), is stronger than CIS, as shown in Barlow and Proschan (1975).

Exponential Distribution. Corollary 1 shows that selection has a negative impact in location experiments with exponential noise. But outside the class of location experiments selection is beneficial when *observations*, rather than noise terms, are exponentially distributed, and $\theta > 0$ parametrizes scale besides location: $F(x|\theta) = 1 - e^{-x/\theta}$ (for $x \geqslant 0$). For all x and $\theta' > \theta$, we have

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

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

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.

Intermediate Order Statistics: Unidimensional vs Multidimensional Accuracy. Our notion of accuracy reveals welfare rankings not captured by unidimensional comparisons between order statistics. Consider a location experiment with positive exponential noise: $F(x|\theta) = F(x-\theta) = \exp(x-\theta)$ (for $x \leqslant \theta$). Let X be a random draw, and let Y_1 and Y_2 be the first and second highest of k draws. By Theorem 1, the evaluator is better off with Y_1 than with X, for $-\log F(\varepsilon)) = -\varepsilon$ is strictly logconcave, but Y_2 and X are incomparable.[10] Yet $f(\varepsilon)/F(\varepsilon) = 1$ is loglinear and hence satisfies the logconcavity criterion in Corollary 1 (and notably *not* the logconvexity condition, because of the bounded-above support), so the evaluator is in fact better off with (Y_1, Y_2).

Minimal Selection. Our results have symmetric counterparts for when the evaluator sees the n lowest rather than highest draws. In location experiments with $n = 1$ minimal selection increases (decreases) welfare if and only if the hazard function $-\log(1-F(\varepsilon))$ is logconcave (logconvex). In general experiments with $n \geqslant 1$, minimal selection increases (decreases) welfare if the hazard rate $f(\cdot|\theta)/[1-F(\cdot|\theta)]$ is log-supermodular (log-submodular, with support of $f(\cdot|\theta)$ independent of θ). We report these claims as Theorems 1* and 2* in Appendix A.

General Method of Proof, Alternative Version. Theorem 2* can be also verified directly, by a variation of the general method discussed in the main text. To this end, note that (4) is identical to

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

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

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

Condition (6) is another sufficient condition for a family of CIS experiments $X(t)$ to be ordered by accuracy. Remarkably, (6) is *not* the same as (5). In fact, we proved Theorem 2 using (5), but we cannot prove it using (6). Similarly, Theorem 2* can be proved via (6), but not via (5).

4 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

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

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

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 distribution's 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 $\alpha = \lim_{\varepsilon \to \bar{\varepsilon}}[1 - F(\varepsilon)]/f(\varepsilon)$,*[11] *the limit welfare is the welfare from an experiment with noise density*

$$(1/\alpha)\exp\left[-\exp(-\varepsilon_n/\alpha) - \varepsilon_1/\alpha - \cdots - \varepsilon_n/\alpha\right]. \tag{7}$$

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

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, *all* noise distributions with logconcave density having support bounded above, e.g. all beta distributions with logconcave density, including uniform. Second, in the unbounded case, beyond normal (or left-truncated normal, which has the same right tails), also every distribution in the exponential power family $f(\varepsilon) = [\gamma/\Gamma(1/\gamma)]\exp(-|\varepsilon|^{\gamma})$ with shape parameter $\gamma > 1$. Strikingly, the Laplace distribution ($\gamma = 1$), which has exponential right tails, is the *only* member of this family for which $\alpha_k \not\to 0$. In this family, the negative impact of selection with exponential noise discussed earlier is not robust to extreme selection—an arbitrarily close distribution reverses the conclusion.

[11] Logconcavity implies increasing hazard rate, so this limit exists when UHR fails.

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:

$$\sup_x \frac{f(x|\theta')}{f(x|\theta)} = \infty \qquad \text{for all } \theta' > \theta. \tag{8}$$

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, (8) 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 (8) 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 (7) 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 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, in this section we characterize when competition monotonically improves or hinders information aggregation in auctions.

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 $\pi(\theta)$ 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. We say that *competition aggregates (disaggregates) information* if an increase in k increases (decreases) the observer welfare from experiment $(B_{1,k}, \ldots, B_{n,k})$. Information aggregation is *full in the limit* if welfare converges to the full information payoff as $k \to \infty$.

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 $(B_{1,k}, \ldots, B_{n,k})$ has the same information value as $(X_{1,k}, \ldots, X_{n,k})$. The following result is therefore an immediate implication of Theorems 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 aggregates (disaggregates) information.* **(b)** *Assuming $F(x|\theta) = F(x-\theta)$,* **(i)** *information aggregation is full in the limit if and only if the noise distribution satisfies UHR;* **(ii)** *if UHR fails, the observer limit welfare is the welfare from an experiment with noisy density (7).*

The aggregation of private information by an auction mechanism as the number of bidders grows unboundedly large, first studied in Wilson's (1977) seminal paper, was investigated by Milgrom (1979) in a model similar to ours.[12] 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 the unbounded informativeness condition (8) is necessary and sufficient for information aggregation to be full in the limit. Recalling that UHR is equivalent to (8), 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 information aggregation. First, our result characterizes when information aggregation, 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 aggregation is not full in the limit. But it is notably not an instance where competition hinders information aggregation. Recall that in this case maximal selection has no impact on welfare—competition both aggregates and disaggregates information.

Second, while failure of convergence to full information is a possibility already contemplated (and characterized) in Milgrom (1979), whether competition could actually disaggregate information was not known. A natural conjecture could have been that even when information aggregation

[12] 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.

is not full in the limit, competition would still tend to improve information aggregation. 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 worsens information aggregation.

Finally, our limit result for uniform-price auctions remains true if we assume that the observer sees not the winning bids but rather the price paid by the winning bidders—the $(n+1)$-highest bid $B_{n+1,k} = b_{n,k}(X_{n+1,k})$. The limit distributions of the highest and $(n+1)$-highest order statistics are different but have the same scale. More precisely Leadbetter et al. (1983, Theorem 2.2.2) show that a pair of sequences β_k and $\alpha_k > 0$ making $(X_{1,k} - \beta_k)/\alpha_k$ converge weakly to a nondegenerate random variable guarantees the same property for $(X_{n+1,k} - \beta_k)/\alpha_k$. 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.

6 Delegated Selection

In the auction setup considered in the previous section, 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. In this selective disclosure game refinements cannot bite, given that no report can contradict maximal selection. The result follows from the fact that maximal selection is a sender best response to any monotone strategy of the evaluator, and by MLR the evaluator optimal strategy is indeed monotone. In effect, even without assuming equilibrium, the behavior described in Proposition 0 is the only behavior compatible with the following assumptions: (i) both players are rational; (ii) the sender believes the evaluator follows a monotone strategy; (iii) the sender selects the n largest draws when indifferent; (iv) the evaluator believes in (i), (ii) and (iii). To see why, note that (i), (ii) and (iii) imply maximal selection and hence, by (i) and (iv), the evaluator chooses the corresponding optimal strategy.

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 level of the 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."[13] 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 9. 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?[14] 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 9 provides results that can be used to accommodate a more general case. Potential outcomes of individual i 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. The evaluator then observes the following:

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

[14] 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.

Treatment Group	Control Group
$X_{1,1} = \theta + \varepsilon_1$	$X_{0,n+1} = \varepsilon_{n+1}$
$\vdots$	$\vdots$
$X_{1,n} = \theta + \varepsilon_n$	$X_{0,2n} = \varepsilon_{2n}$

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, so the experiment boils down to the treatment group—a random experiment with sample size n. Now 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.

7 Experiment Design

Determining the optimal sample size of a random experiment for a given sample size cost is standard in statistical decision theory—see Berger (1985) for a basic treatment. Our results offer a new angle on this issue. In many situations the evaluator has direct or indirect control over presample size—soliciting or discouraging potential bidders, giving more or less choice in exams, tolerating more or less randomness in a research study. Thus, the evaluator can strategically leverage maximal selection to increase the informativeness of the experiment, or save on sampling cost.

7.1 Optimal Presampling

Consider the problem of an evaluator who can set both sample size n and presample size k at costs $C_S(n)$ and $C_P(k)$, respectively. For example, $C_P(k)$ and $C_S(n)$ can represent an examiner's cost of preparing k questions and grading n of them. In an auction setup with a fixed number of objects for sale, we can take $C_S(n)$ to be zero up to that number and prohibitively high for larger n, while $C_S(k)$ measures the cost of soliciting k bidders. Let $U(k,n)$ denote the evaluator optimal expected

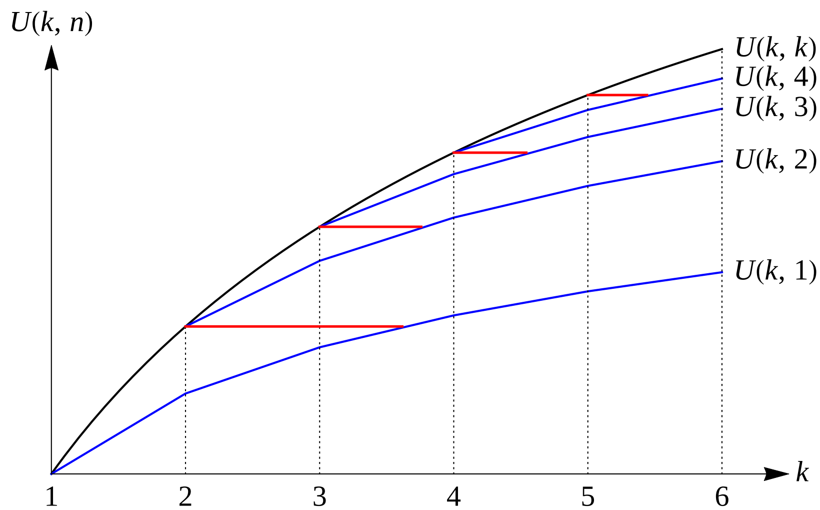

Figure 3: Welfare as a function of presample and sample size: normal location experiment.

(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, under logconvexity of the noise reverse hazard function or log-submodularity of the data reverse hazard rate, the optimal experiment format features no selection. The optimal sample size, equal to presample size, solves $\max_{n \geqslant 1}[U(n,n) - C_S(n)]$, and we are back to standard optimal sample size determination.

When instead selection benefits, the trade-off is nontrivial. The evaluator values sample size but also values sample selection: n and k are two goods. The exact trade-off depends on the specific cost functions and family $F(\cdot|\theta)$ under consideration. 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 optimal experiment format sample size is $n = 1$ and presample size solves $\max_{k \geqslant 1}[U(k,1) - C_S(1) - C_P(k)]$.

Figure 3 illustrates the less immediate trade-off arising with normal noise, in the context of simple hypothesis testing—for instance, assessing the (binary) ability of a student with a pass-fail test. Normalizing the evaluator gross payoff so that $U(1,1) = 0$, each blue function corresponds to a fixed sample size and increases with presample size (Theorem 2). Their black upper envelope represents welfare without selection ($k = n$). All red segments are shorter than two: starting from any $k = n \geqslant 2$ the evaluator has an incentive to decrease sample size by one, while increasing presample size by two: $U(n+2, n-1) > U(n,n)$ for every $n \geqslant 2$. Thus, if $C_P(n+2) + C_S(n-1) \leqslant C_P(n) + C_S(n)$ for all $n \geqslant 2$, the optimal experiment format must be such that $k > n$.

Small Presampling Costs. The examples just discussed illustrate how the evaluator can exploit selection to economize on sample size when presampling costs are sufficiently small. 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 increasing and unbounded and that $F(x|\theta) = F(x-\theta)$ satisfies UHR. There exist $c > 0$ and $\bar{k} > 1$ such that, if $C_P(k) \leqslant c\bar{k}$, then every optimal experiment format (k,n) features sample selection: $k > n$.*

The proof of the proposition is based on a simple argument. By the assumption on sampling costs, the design problem under the constraint $k = n$ has an optimal solution $(\bar{n}, \bar{n})$. By Theorem 3, we can choose k sufficiently large that $(k,1)$ gives higher (gross) payoff than $(\bar{n}, \bar{n})$. Sending c to zero gives the result. We return to Figure 3 to illustrate Proposition 2 in a simple hypothesis testing problem with normal noise. For example, with constant marginal costs c_S and c_P for sampling and presampling, whenever $c_P < c_S/2$ the optimal format must feature sample selection.

7.2 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. In this scenario, presample size would naturally be unobserved by the evaluator. 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 increasing and unbounded, $C_P(k)$ is increasing, convex and unbounded, and $F(x|\theta) = F(x-\theta)$ satisfies UHR. There exist $c > 0$ and $\bar{k} > 1$ such that, if $C_P(k) \leqslant c\bar{k}$, 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. After proving the proposition, in Appendix A we show that an equilibrium always exists when presample size can be any real number k and $C_P(k)$ is continuous. If presample size is restricted to be an integer, we show that there exist sequences $\overline{c}_k > \underline{c}_k > 0$ with $\overline{c}_k \to 0$ as $k \to \infty$ such that equilibrium exists when there is some k such that $\underline{c}_k k' \leqslant C_P(k') \leqslant \overline{c}_k k'$ for all k'.

The proof of Proposition 3 is similar to that of Proposition 2, but the argument is less immediate. First, we use Theorem 3 to find a maximally selected experiment with sample size $n = 1$ that, once sampling costs are taken into account, dominates all random experiments with any sample size. Second, given any presample size conjectured by the evaluator, we show that the sender expected payoff is strictly concave in the actual presample size chosen by the sender. This fact allows us to conclude that for sufficiently small presampling costs the sender's optimal presample size is both large enough and equal to the evaluator's conjectured presample size.

Delegated presampling triggers a rat race: given any sample size n, the sender has an incentive to increase the presample size, 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.[15] As presample costs become very small, the sender sets an arbitrarily large presample size, which by UHR benefits the evaluator even when sample size is $n = 1$. Thus, by delegating presampling the evaluator exploits the rat race and saves on sampling costs, effectively passing the cost of information to the sender.

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* perfect 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 increasing, convex and unbounded, and $F(x|\theta) = F(x - \theta)$ satisfies UHR. There exist $c > 0$ and $\bar{k} > 1$ such that, if $C_P(k) \leqslant c\bar{k}$, 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.

8 Other Forms of Selection

Maximal (or minimal) selection is but one instance of lack of randomness in a statistical sample. In this section we discuss two other forms of selection.

8.1 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

[15] With Gumbel noise presampling is a pure rat race. 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.

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

We use the two versions of our general method to prove the two sides of the claim. Goel and DeGroot (1992) proved Theorem 4 for UHR distributions and monotone preferences. Their proof strategy crucially 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 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 left-truncation (from Y_a to Y_b) hurts when the hazard rate is log-supermodular, so its impact is analogous to less minimal selection (Theorem 1*) rather than to more maximal selection. Similarly, more right-truncation (from W_b to W_a) hurts when the reverse hazard rate is log-supermodular, so its effect is analogous to less maximal selection (Theorem 1). The welfare consequences are strikingly different. With normal or logistic noise, hazard rate and reverse hazard rate are both logconcave, so more maximal *or* minimal selection benefits the evaluator (Corollary 1). However, more truncation hurts both ways: experiments Y_b and W_a are respectively worse than the less truncated Y_a and W_b.

8.2 Median Selection

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

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

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

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

The proof of this result also uses both versions of our general method. Since median and random experiment are not ranked by first-order stochastic dominance, neither version alone suffices. To grasp intuition, consider simple hypothesis testing and let $\bar{x}$ denote the optimal cutoff for the random experiment. Similarly to maximal selection, when $\bar{x}$ is below the median of $F(\cdot|\theta_L)$ the cutoff $\bar{z}$ defined by $\hat{F}(\bar{z}|\theta_L) = F(\bar{x}|\theta_L)$ is larger than $\bar{x}$. This fact, together with log-supermodularity of the reverse hazard rate, guarantees that our sufficient condition (5) holds and hence that (4) holds: $\hat{F}(\bar{z}|\theta_H) \leqslant F(\bar{x}|\theta_H)$. If instead $\bar{x}$ is above the median, then $\bar{z}$ is smaller than $\bar{x}$. In this case, log-supermodularity of the hazard rate guarantees the alternative sufficient condition (6).

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 $F(\cdot|\theta)$.[16] 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 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.

[16] 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.

9 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 *added noise* terms γ_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 $\delta_1, \ldots, \delta_k$ 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.[17] 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 hurts 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 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.

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

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 distribution's 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 within the examinee choice setting, suppose that if the 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. Then, 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 even though we call γ_i an *added* noise term, the size of c also determines the precision of the evaluator observations. In our examinee choice example, suppose that X_i and Y_i are both 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.

10 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:

$$-\pi(\theta_L)\underbrace{\left[F(\bar{x}-\theta_L) - F^2(\bar{x}-\theta_L)\right]}_{\text{increase in false positives}}\left[u(\theta_L,a_1) - u(\theta_L,a_2)\right] \\ + \pi(\theta_H)\underbrace{\left[F(\bar{x}-\theta_H) - F^2(\bar{x}-\theta_H)\right)}_{\text{reduction in false negatives}}\left[u(\theta_H,a_2) - u(\theta_H,a_1)\right]. \quad (9)$$

In the important benchmark case of *equipoise*, the evaluator is a priori indifferent between accepting and rejecting, that is, $\pi(\theta_L)[u(\theta_L,a_1) - u(\theta_L,a_2)] = \pi(\theta_L)[u(\theta_H,a_2) - u(\theta_H,a_1)]$, 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 (9) 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 (9) 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 $\pi(\theta_L)u(\theta_L, a_2) + \pi(\theta_H)u(\theta_H, a_2)$, no larger than the no-information payoff.

A Proofs

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 $F^k(\varepsilon_k) = F^m(\varepsilon_m)$,

$$mF^{m-1}(\varepsilon_m)f(\varepsilon_m) \leqslant (\geqslant)\, kF^{k-1}(\varepsilon_k)f(\varepsilon_k). \tag{10}$$

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, (10) 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 in the above inequality are negative and equal to each other, 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 numbers $k > m$ if and only if $\lambda(F(\cdot))$ is concave (convex), because $k > m$ implies $\varepsilon_k \geqslant \varepsilon_m$. □

Proof of Theorem 2. Fix two presample sizes k and m, and for every $t \in [0,1]$ denote by $X(t)$ the selected experiment with presample size $k_t := tk + (1-t)m$. Fix $s < t$ and $\theta < \theta'$, and write $\varphi_{s,t}(x|\theta) = z$ and $\varphi_{s,t}(x|\theta') = z'$ for brevity. As a preliminary observation, note that the support of $X_1(t)$ is the support of $f(\cdot|\theta)$ and hence it does not depend on t. Thus, as x_1 converges to the upper bound of this support, so does z_1. Similarly, for every $i = 2, \ldots, n$ and every x_{i-1}, as x_i converges to x_{i-1} (its largest possible value), z_i converges to z_{i-1}. We must prove that under either condition in the theorem ($m \geqslant k$ and $f(\cdot|\theta)/F(\cdot|\theta)$ log-supermodular, or $m \leqslant k$ and $f(\cdot|\theta)/F(\cdot|\theta)$ log-submodular with support of $f(\cdot|\theta)$ independent of θ) for every x we have $z' \geqq z$, or equivalently

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

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

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

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

$$\frac{F^{k_s}(z_1|\theta')}{F^{k_t}(x_1|\theta')} \leqslant 1 \quad \text{and} \quad \frac{F^{k_s-i+1}(z_i|\theta')}{F^{k_t-i+1}(x_i|\theta')} \leqslant \frac{F^{k_s-i+1}(z_{i-1}|\theta')}{F^{k_t-i+1}(x_{i-1}|\theta')} \quad \text{for } i = 2, \ldots, n. \tag{11}$$

By the preliminary observation, as x_1 tends to the upper bound of the support of $f(\cdot|\theta)$ so does z_1. Thus, under either condition in the theorem the left-hand side of the first inequality in (11) tends to a number no greater than one. This implies that the first inequality in (11) holds if the left-hand side of the inequality increases with x_1. Thus, differentiating with respect to x_1 and dropping the positive denominator in the derivative, we need

$$k_s F^{k_s-1}(z_1|\theta') f(z_1|\theta') \frac{dz_1}{dx_1} F^{k_t}(x_1|\theta') \geqslant k_t F^{k_t-1}(x_1|\theta') f(x_1|\theta') F^{k_s}(z_1|\theta'). \tag{12}$$

By definition of z,

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

Plugging the latter in (12) and simplifying, we conclude that the first inequality in (11) holds if

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

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

Again by the preliminary observation, for every $i = 2, \ldots, n$ and every x_{i-1}, as x_i converges to x_{i-1}, z_i converges to z_{i-1}. Thus, as before, under either condition in the theorem the left-hand side of the second inequality in (11) tends to a number no greater than the right-hand side. The second inequality in (11) then holds if its left-hand side increases with x_i. Differentiating with respect to x_i and simplifying, as before, we obtain

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

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

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.[18] Still, the examples discussed after the theorem—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 $e^{3\varepsilon} + \eta(e^{2\varepsilon} - 1)$.

Minimal Selection. Call *minimally selected* an experiment $X = (X_1, \ldots, X_n)$ where X_1 is the smallest of $k \geqslant n$ random draws from $F(\cdot|\theta)$, X_2 the second smallest, and so on.[19] Consider the dual problem with state space $\tilde{\Theta} = \{-\theta : \theta \in \Theta\}$, action space $\tilde{A} = \{-a : a \in A\}$, payoff function $\tilde{u}(\tilde{\theta}, \tilde{a}) = u(-\tilde{\theta}, -\tilde{a})$, and signal distribution $\tilde{F}(\tilde{x}|\tilde{\theta}) = 1 - F(-\tilde{x}| - \tilde{\theta})$, with density $\tilde{f}(\tilde{x}|\tilde{\theta}) = f(-\tilde{x}| - \tilde{\theta})$. For location experiments, the cumulative noise distribution is therefore $\tilde{F}(\tilde{\varepsilon}) = 1 - F(-\tilde{\varepsilon})$. Having changed sign to both states and actions, the dual problem is also monotone, and having changed sign to both states and signals, the MLR property holds. Finally, action a is optimal given a realization x in the original problem if and only if $-a$ is optimal given realization $-x$ in the dual problem. Since the linear transformation $\varepsilon \mapsto -\varepsilon$ does not change the logconcavity (or logconvexity) of $-\log(\tilde{F}(\cdot))$, this property is equivalent to logconcavity (or logconvexity) of $-\log(1 - F(\cdot))$. Thus, Theorem 1 has the following counterpart:

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

For general experiments, note that $\tilde{f}(\tilde{x}|\tilde{\theta})/\tilde{F}(\tilde{x}|\tilde{\theta}) = f(-\tilde{x}| - \tilde{\theta})/[1 - F(-\tilde{x}| - \tilde{\theta})]$, and the switch of sign in both arguments does not affect log-supermodularity (or log-submodularity). Thus, Theorem 2 has the following counterpart:

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

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

is increasing (resp. decreasing).

Proof of Theorem 3. We start by showing that under 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

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

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

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.[20] 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, $\overline{U} := \int_\Theta \max_a u(\theta,a)\pi(\theta)d\theta$, 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, one consequence of this observation is that if $\theta_j < \theta_{j-1}$ then action a_j can be removed from A without affecting the IDO property. Another consequence is that action a_j is never optimal at any state, and hence it is never used under full information. Thus, we may assume without loss of generality that $\theta_j \geqslant \theta_{j-1}$ for all $j > 1$. The full information payoff can then be written, summing by parts, as

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

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

$$\int_\Theta \sum_{j<J} \mathbf{1}_{\{\theta_j-\eta\leqslant\theta<\theta_j\}} \big[u(\theta,a_{j+1}) - u(\theta,a_j)\big]\pi(\theta)d\theta \leqslant \frac{\delta}{2} \tag{13}$$

and, furthermore,

$$(1-\eta)\int_\Theta \sum_{j<J} \mathbf{1}_{\{\theta\geqslant\theta_j\}} \big[u(\theta,a_{j+1}) - u(\theta,a_j)\big]\pi(\theta)d\theta \geqslant \overline{U} - \frac{\delta}{2}. \tag{14}$$

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,

$$\begin{aligned}
\Pr_\theta(\theta-\eta+\beta_k \leqslant X \leqslant \theta+\eta+\beta_k) &\geqslant \Pr_\theta(\theta-\alpha_k\bar{\varepsilon}+\beta_k \leqslant X \leqslant \theta+\alpha_k\bar{\varepsilon}+\beta_k) \\
&= F^k(\alpha_k\bar{\varepsilon}+\beta_k) - F^k(\alpha_k\bar{\varepsilon}+\beta_k) \\
&\geqslant \hat{F}(\bar{\varepsilon}) - \frac{\eta}{4} - \hat{F}(-\bar{\varepsilon}) - \frac{\eta}{4} \geqslant 1-\eta,
\end{aligned} \tag{15}$$

[20] 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 $\bar{F}$.

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

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

By (13), (14) and (15), this payoff is at least as large as $\overline{U}-\delta$.

Finally, we show that when UHR fails, the limit welfare is the welfare from an experiment with noise density (7). Violation of UHR implies that $\bar{\varepsilon}=\infty$ and hence (as already established earlier) that $\bar{F}$ is Gumbel, with $\lim_{k\to\infty}\alpha_k=\lim_{\varepsilon\to\bar{\varepsilon}}[1-F(\varepsilon)]/f(\varepsilon)=:\alpha>0$. 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 the location of each draw by β_k, converges weakly to the noise distribution with density (7). □

Proof of Proposition 2. Let $\bar{n}=\arg\max_{n\geqslant1}U(n,n)-C_S(n)$. 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, $U(\bar{n},\bar{n})-C_S(\bar{n})<\overline{U}-C_S(1)$, 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 $\delta>0$ such that

$$U(\bar{k},1)-C_S(1)-\delta>U(\bar{n},\bar{n})-C_S(\bar{n}). \tag{16}$$

Let $c=\delta/\bar{k}$ and assume that $C_P(\bar{k})\leqslant c\bar{k}=\delta$. Then

$$\begin{aligned}U(\bar{k},1)-C_S(1)-C_P(\bar{k})>U(\bar{n},\bar{n})-C_S(\bar{n})&\\ \geqslant\max_{n\geqslant1}U(n,n)-C_S(n)\geqslant\max_{n\geqslant1}U(n,n)-C_S(n)-C_P(n).&\end{aligned}$$

□

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 $\bar{k}\geqslant1$ such that

$$U(k,1)-C_S(1)>U(\bar{n},\bar{n})-C_S(\bar{n})\qquad\text{for all } k\geqslant\bar{k}. \tag{17}$$

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 gross payoff the sender obtains by choosing presample size k is

$$V(k,k'):=\int_\Theta\Big[v(\theta,a_1)+\sum_{j<J}[1-F^k(\bar{x}_{j+1}(k'))][v(\theta,a_{j+1})-v(\theta,a_j)]\Big]\pi(\theta)d\theta,$$

where we used summation by parts to rewrite the payoff. Thus, by increasing presample size from k to $k+1$ the sender incurs marginal cost $C_P(k+1)-C_P(k)$ and obtains marginal gain

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

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 $C_P(k) \leqslant ck$ for every k in that set. Then it is clear that there is no Bayes Nash equilibrium in which the sender chooses a presample size $k \leqslant \bar{k}$. □

Equilibrium Existence in the Delegated Presampling Game. Assume first that presample size can be any real $k \geqslant 1$ and $C_P(k)$ is continuous. 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 $\int_\Theta v(\theta,a_1)\pi(\theta)d\theta - C_P(1) \geqslant \int_\Theta v(\theta,a_J)\pi(\theta)d\theta - C_P(\hat{k})$. Thus, we can also restrict the evaluator strategy set to be the set of all optimal monotone strategies for selected experiments with presample size $k \in [1,\hat{k}]$. Moreover, identifying the optimal strategy for presample size 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

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

has a fixed point. (The sender best response is a function because the sender objective function is strictly concave: $C_P(k)$ 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.

Next, assume that k must be an integer. For each $k \geqslant 1$ consider the interval $[\underline{c}_k, \bar{c}_k]$ such that

$$\underline{c}_k := V(k+1,k) - V(k,k) \quad \text{and} \quad \bar{c}_k := V(k,k) - V(k-1,k).$$

By concavity of the sender objective function, this interval is nonempty. Moreover, in order to check that the sender has no profitable deviations it suffices to check that the sender has no incentive to increase or decrease sample size by one. It follows that an equilibrium exists if there exists k such that $\underline{c}_k k' \leqslant C_P(k') \leqslant \bar{c}_k k'$ for every k'.

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 $n \geqslant 1$. As shown in the proof of Proposition 3, there exists c such that if $C_P(k) \leqslant ck$ for every $k \leqslant \bar{k}$ then in every Bayes Nash equilibrium of the delegated presampling game the sender must choose a presample size greater than $\bar{k}$. 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 $U(\bar{n},\bar{n}) - C_S(\bar{n}) \leqslant \overline{U} - C_S(\bar{n}) < U(\bar{k},1) - C_S(1)$ for every $k \geqslant \bar{k}$, so we are done. □

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Proof of Theorem 5. Since payoffs satisfy Karlin and Rubin's (1956) monotonicity, Theorem 0 holds for families of experiments $X(t)$ where t is an index from an arbitrary ordered set T, as shown in Appendix B. Thus, to prove the theorem 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 $\varphi_{0,1}(x|\theta)$ is defined by the equality

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

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

$$\frac{dz}{dx} = \frac{f(x|\theta)}{cF^{r-1}(z|\theta)[1-F(z|\theta)]^{r-1}f(z|\theta)}. \tag{19}$$

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

$$\frac{F^r(z|\theta')\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}[1-F(z|\theta')]^j}{F(x|\theta')} \leqslant 1, \tag{20}$$

which is the same as

$$\frac{[1-F(z|\theta')]^r\sum_{j=0}^{r-1}\binom{r+j-1}{r-1}F^j(z|\theta')}{1-F(x|\theta')} \geqslant 1. \tag{21}$$

Suppose first that $x \leqslant x_m$, so that $z \geqslant x$. Since $F(\cdot|\theta')$ first-order stochastically dominates $F(\cdot|\theta)$, condition (20) holds at $x = x_m = z$. Thus, it suffices to show that the left-hand side of (20) increases in x when $x \leqslant z$. The derivative of the left-hand side is nonnegative if and only if

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

Plugging in (19) and using (18), the latter inequality is the same as

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

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

Suppose now that $x \geqslant x_m$, so that $z \leqslant x$. Since (21) is the same as (20), it holds at $x = x_m = z$, so it suffices to show that its left-hand side increases in x when $x \geqslant z$. The derivative of the left-hand side of (21) is nonnegative if and only if

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

Plugging in (19) and using (18), the latter inequality is the same as

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

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

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 $c\delta_{i_k} + \gamma_1$. But the latter has the same distribution as $c\delta_{i_k} + \gamma_{i_k}$, hence the result follows. □

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 $\pi(\theta|\hat{\theta})$ increases with $\hat{\theta}$ in the likelihood ratio order. Thus, defining $\hat{u}(\hat{\theta},\cdot) := \int_\Theta u(\theta,\cdot)\pi(\theta|\hat{\theta})d\theta$ 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. To see this, given any $E \subseteq \mathbb{R}^n$ and $a \in \mathbb{R}$ write $E(a)$ to denote the set $\{(x_1-a,\ldots,x_n-a) : x \in E\}$. Then, given any partition $(E_1,\ldots,E_J)$ of $\mathbb{R}^n$, denoting the density of $\hat{\theta}$ by $\hat{\pi}$, we have

$$\begin{aligned}
&\int_{\mathbb{R}} \sum_j \Pr((\theta+c\delta_1+\gamma_1,\ldots,\theta+c\delta_n+\gamma_1) \in E_j)u(\theta,a_j)\pi(\theta)d\theta \\
&= \int_{\mathbb{R}} \sum_j \Pr((c\delta_1,\ldots,c\delta_n) \in E_j(\theta+\gamma_1))u(\theta,a_j)\pi(\theta)d\theta \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} \sum_j \Pr((c\delta_1,\ldots,c\delta_n) \in E_j(\hat{\theta}))u(\theta,a_j)\hat{\pi}(\hat{\theta})\pi(\theta|\hat{\theta})d\theta d\hat{\theta} \\
&= \int_{\mathbb{R}} \sum_j \Pr((c\delta_1,\ldots,c\delta_n) \in E_j(\hat{\theta}))\hat{u}(\hat{\theta},a_j)\hat{\pi}(\hat{\theta})d\hat{\theta}.
\end{aligned}$$

□

B Accuracy and Welfare

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

B.1 Monotone Preferences

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

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

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

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

$$\Pr_\theta(X(t) \in \bar{E}'_{j+1}) - \Pr_\theta(X(s) \in \bar{E}_{j+1}(s)) \tag{22}$$

is nonpositive for $\theta \leqslant \theta_j$ and nonnegative for $\theta > \theta_j$. Indeed, this implies that the evaluator can achieve a higher expected payoff in experiment $X(t)$ by adopting the following strategy: choose a_1 when $X(t) \notin \bar{E}'_2$, choose a_2 when $X(t) \in \bar{E}'_2 \setminus \bar{E}'_3$, and so on. Define $\bar{E}'_{j+1} = \varphi_{s,t}(\bar{E}_{j+1}|\theta_j)$ for every $j < J$. Then we can rewrite the difference in (22) as

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

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

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

B.2 IDO Preferences

The above proof does not extend immediately to IDO preferences or even only single-crossing preferences. The IDO property does imply that the difference $u(\theta, a_{j+1}) - u(\theta, a_j)$ exhibits single crossing, but does not require the crossing points θ_j to be increasing in j. This makes the upper sets $\bar{E}'_2, \ldots, \bar{E}'_J$ 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.

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

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

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

$$U(X(u)) - U(X(t)) = \int\limits_t^u \int\limits_\Theta \sum_{j<J} \frac{\partial \Pr_\theta(s, E_j(s))}{\partial t} \left[u(\theta, a_j) - u(\theta, a_{j+1})\right] \pi(\theta) d\theta ds, \tag{23}$$

and we have to show that the expression in (23) 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 (23), as follows. Recall that, by IDO, for every $1 \leqslant j < J$ there exists a state θ_j such that the difference $u(\theta, a_j) - u(\theta, a_{j+1})$ is nonnegative for $\theta \leqslant \theta_j$ and nonpositive for $\theta \geqslant \theta_j$. An immediate consequence of this observation is that if $\theta_j < \theta_{j-1}$ then action a_j can be removed from A without affecting the IDO property—letting $\tilde{u} : \Theta \times A \setminus \{a_j\} \to \mathbb{R}$ denote the restriction of u to $\Theta \times A \setminus \{a_j\}$, the family $\{\tilde{u}(\theta, \cdot)\}_{\theta \in \Theta}$ is again an IDO family. By using this fact (repeatedly, if necessary) together with the fact that $E_j(s) = E_{j-1}(s)$ whenever $D_j(s) = \varnothing$, we conclude that for every s there exists a list of indices $1 \leqslant j(s,1) < \cdots < j(s, I_s) \leqslant J$ of some length $I_s \leqslant J$, and a list of states $\theta_1(s), \ldots, \theta_{I_s}(s)$, with the following properties. First,

$$U(X(u)) - U(X(t)) = \int\limits_t^u \int\limits_\Theta \sum_{i<I_s} \frac{\partial \Pr_\theta\left(s, E_{j(s,i)}(s)\right)}{\partial t} \left[u(\theta, a_{j(s,i)}) - u(\theta, a_{j(s,i+1)})\right] \pi(\theta) d\theta ds. \tag{24}$$

Second,

$$u(\theta, a_{j(s,i)}) - u(\theta, a_{j(s,i+1)}) \gtrless 0 \quad \text{for } \theta \lessgtr \ \theta_i(s). \tag{25}$$

Third,

$$\theta_i(s) \geqslant \theta_{i-1}(s) \qquad \text{for all } i \in \{2, \ldots, I_s\} \text{ such that } D_{j(s,i)}(s) = \varnothing. \tag{26}$$

Step 2—Use accuracy to set a lower bound on the payoff difference. Take any θ, s and $i < I_s$, and consider the corresponding derivative appearing inside the summation in (24). We have

$$\begin{aligned}
\frac{\partial \Pr_\theta\left(s, E_{j(s,i)}(s)\right)}{\partial t} &= \lim_{\delta\to 0} \frac{\Pr_\theta\left(s+\delta, E_{j(s,i)}(s)\right) - \Pr_\theta\left(s, E_{j(s,i)}(s)\right)}{\delta} \\
&= \lim_{\delta\to 0} \frac{\Pr_\theta\left(s, \varphi_{s+\delta,s}(E_{j(s,i)}(s)|\theta)\right) - \Pr_\theta\left(s, E_{j(s,i)}(s)\right)}{\delta} \\
&\gtrless \lim_{\delta\to 0} \frac{\Pr_\theta\left(s, \varphi_{s+\delta,s}(E_{j(s,i)}(s)|\theta_i(s))\right) - \Pr_\theta\left(s, E_{j(s,i)}(s)\right)}{\delta} \quad \text{for } \theta \lesseqgtr \theta_i(s).
\end{aligned} \tag{27}$$

The second equality follows from the definition of the function $\varphi_{s+\delta,s}(\cdot|\cdot)$. The inequality, from $X(s+\delta)$ being more accurate than $X(s)$ for every $\delta > 0$ (as this means that $\varphi_{s+\delta,s}(x|\theta)$ is decreasing in θ for every x and $\delta > 0$) and $E_{j(s,i)}(s)$ being a lower set (the complement of an upper set). Letting $L(\theta, s, i)$ denote the right-hand side of (27), from (24), (25) and (27) we obtain

$$U(X(u)) - U(X(t)) \geqslant \int_t^u \int_\Theta \sum_{i<I_s} L(\theta,s,i)\left[u(\theta,a_{j(s,i)}) - u(\theta,a_{j(s,i+1)})\right]\pi(\theta)d\theta ds. \tag{28}$$

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

$$\int_\Theta \sum_{i<I_s} L(\theta,s,i)\left[u(\theta,a_{j(s,i)}) - u(\theta,a_{j(s,i+1)})\right]\pi(\theta)d\theta \geqslant 0. \tag{29}$$

The result will then follow from (28) and (29). First note that, since $E_{j(s,i)}(s)$ is a lower set, for some function $\bar{x}_n : \mathbb{R}^{n-1} \to \mathbb{R}\cup\{-\infty,+\infty\}$ we have

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

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

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

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

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

where $\partial E_{j(s,i)}(s)$ denotes the boundary of $E_{j(s,i)}(s)$. Thus, the expression in (29) can be written as

$$\sum_{i<I_s} \int_{\partial E_{j(s,i)}(s)} K(i,x_{<n}) \int_\Theta g(s,x_{<n},x_n|\theta)\left[u(\theta,a_{j(s,i)}) - u(\theta,a_{j(s,i+1)})\right]\pi(\theta)d\theta. \tag{30}$$

Step 4—Show that the lower bound is nonnegative. Since $\varphi_{s+\delta,s}(\theta,x)$ is decreasing in θ, for each $i < I_s$ such that $\theta_i(s) \leqslant \theta_{i+1}(s)$ we have $\bar{x}_{n,i}(\delta,x_{<n}) \geqslant \bar{x}_{n,i+1}(\delta,x_{<n})$, and hence $K(i,x_{<n}) \geqslant K(i+1,x_{<n})$ for all $x_{<n}$. Let $i_1 < \ldots < i_{H(s)}$ denote the set of indices $i < I_s$ such that $D_{j(s,i)}(s) \neq \varnothing$. Then, for every $h \in \{2,\ldots,H(s)\}$ and $i \in \{i_h+1,\ldots,i_{h+1}-1\}$, at each point in the boundary of $E_{j(s,i_h)}(s)$ the evaluator prefers $a_{j(s,i_{h+1})}$ to $a_{j(s,i)}$, and hence, using (26),

$$\sum_{i=i_h}^{i_{h+1}-1} \int\limits_{\partial E_{j(s,i_h)}(s)} K(i,x_{<n}) \int\limits_{\Theta} g(s,x_{<n},x_n|\theta)\left[u(\theta,a_{j(s,i)}) - u(\theta,a_{j(s,i+1)})\right]\pi(\theta)d\theta$$
$$\geqslant \int\limits_{\partial E_{j(s,i_h)}(s)} K(i_h,x_{<n}) \int\limits_{\Theta} g(s,x_{<n},x_n|\theta)\left[u(\theta,a_{j(s,i_h)}) - u(\theta,a_{j(s,i_{h+1})})\right]\pi(\theta)d\theta.$$

It follows that the expression in (30) is at least as large as

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

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

B.3 Continuous Actions

To deal with a continuous action set A we make two assumptions. First, we assume payoffs are continuous and bounded below (e.g. nonnegative). Second, we impose regularity on the family of functions $\{u(\theta,\cdot)\}_{\theta\in\Theta}$ by assuming that the family of their restrictions to every sufficiently large but finite subset of actions is also an IDO family. Note that the latter assumption is automatically satisfied with single-crossing or monotone preferences.[21] 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

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

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

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

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

and

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

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

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

and

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

Since Θ' is arbitrary and we can replace c with any a in A (because the union of the B's is dense in A), we conclude that the probability measure on states and actions induced by $a_B(\cdot)$ converges weakly to that induced by $a(\cdot)$. Thus, $\liminf_B \mathrm{E}_B(u) \geqslant \mathrm{E}(u)$, 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.

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 (forthcoming): "A Theory of Experimenters: Robustness, Randomization, and Balance," *American Economic Review*.

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.

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

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

Blackwell, D. (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.

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

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

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

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

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

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

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

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

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

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

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

Hoffmann, F., R. Inderst, and M. Ottaviani (forthcoming): "Persuasion through Selective Disclosure: Implications for Marketing, Campaigning, and Privacy Regulation," *Management Science*.

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.

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.

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.

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

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

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

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

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.