Strategic Sample Selection*

Alfredo Di Tillio† Marco Ottaviani‡ Peter Norman Sørensen§

April 23, 2017

Abstract

What is the impact of sample selection on the inference payoff of an evaluator testing a simple hypothesis based on the outcome of a location experiment? Compared to a random data point, data selected as the highest of several observations is less dispersed and thus always increases the evaluator's welfare if and only if quantile density of the noise distribution is less elastic than for the Gumbel distribution, as with logistic or normal noise. More generally, we characterize the welfare impact of sample selection depending on its effect on local dispersion. Also, we show that extreme selection benefits the evaluator. The results are applied to the analysis of strategic sample selection by a biased researcher who strategically selects the most favorable of several observations.

*Research funded by the European Research Council through Advanced Grant 29583 (EVALIDEA). We thank Matteo Camboni for outstanding research assistance. We also thank Pierpaolo Battigalli and participants at seminars at Bocconi, Bruxelles, Carlos III, East Anglia, Helsinki, Leicester, Milan, Nottingham, Southampton, Surrey, and Toulouse 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

As analysts have long recognized, observational data are often nonrandomly selected, because of two main selection mechanisms: either self selection is induced by choices made by the subjects under investigation, or selection originates from sample inclusion decisions made by data analysts themselves when carrying out the study.[1] Experimental data can also suffer from selection problems: either because the study population is not representative of the population of interest (challenging external validity),[2] or because the randomized allocation to treatment rather than control is subverted (challenging internal validity).[3] Whatever the source of selection, it is natural to wonder about the impact of sample selection on the quality of statistical inference.

This paper provides a framework to characterize the information value of sample selection. To set the stage, Section 2 considers an evaluator assessing a binary hypothesis regarding the treatment effect θ. Observing evidence $\theta + \varepsilon$, the evaluator suspects some degree of sample selection which biases upward the distribution of the noise term ε. The model for ε influences the evaluation of the evidence and the ensuing conclusion reached by the evaluator. To study the impact of selection, we initially focus attention on the case where this model is correct given the actual extent of sample selection. Does greater selection benefit or harm the evaluator? The answer depends on the distribution of the un-selected ε.

We assume that sample selection manifests itself through observation of the highest out of k realizations. The noise term follows a given log-concave distribution F.[4] Log-concavity implies

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

[2] In this regard, Alcott (2015) documents the presence of hard-to-control-for selection of the experimental sites in the context of experiments conducted by utility companies to evaluate the effectiveness of energy-savings policies.

[3] See Schulz (1995), Schulz et al. (1995), and Berger (2005) for extensive accounts and examples of subversion of randomization in clinical trials. As explained by Berger (2005), the practice of blocking to ensure an equal number of patients in the control and in the treatment group tends to make allocation to control/treatment more predictable toward the end of the block, allowing researchers to subvert the assignment of individual patients depending on the outcomes they expect for individual patients.

[4] Di Tillio, Ottaviani and Sørensen (2015) compare different types of selection in the context of a illustrative

that the evaluator takes higher actions when the observation is higher, for given model. This property is inherited by the distribution of the maximum of k i.i.d. realizations, F^k, where k measures the extent of selection.

The evaluator makes a binary choice to accept or reject. In this hypothesis testing problem, a central role is played by the type I error rate (false acceptance) α and the type II error rate (false rejection) β. The relative decision weight attached to α represents the evaluator's preference against false acceptance. Efficient use of the evidence yields a convex relationship of β as a function of α, which we call the information constraint of the experiment. Sample selection affects the evaluator through the change in this information constraint.

Section 3 characterizes a necessary and sufficient conditions under which the highest order statistics (among k observations) is less dispersed than a random observation, and further dispersion is monotonically decreasing in the number of observations. Our first main result (Theorem 1) shows that the distribution F^{k+1} of the maximum of $k+1$ observations is less dispersed than the distribution F^k of the maximum of k observations *if and only if* $-\log(-\log F)$ is convex, or equivalently whenever the quantile density of F is less elastic than for Gumbel's extreme value distribution. Thus, under this condition, increase in selection shifts up the entire information constraint for every α. Given that less dispersed noise results in an unambiguous increase in the payoff of the evaluator, this result characterizes the class of distributions F for which the evaluator's payoff must be increasing (or decreasing) in the extent k of selection. Equivalently, this criterion says that the right tail of the distribution F is thinner than for the Gumbel distribution. For example, the condition implies that the evaluator always gains from sample selection when F is normal or logistic, but always loses when F is exponential.[5]

The proof of Theorem 1 applies Lehmann's (1988) dispersion order to compare distribu-

model with binary errors, which violate log-concavity.

[5]Previous results in the literature on stochastic ordering of order statistics only covered distributions with decreasing hazard rate—the only logconcave distribution to which they apply is the exponential distribution for which selection monotonically hurts the evaluator—see Section 3's discussion of Khaledi and Kochar (2000). Our characterization, instead, also covers the relevant case of distribution with increasing hazard rate, which is implied by logconcavity of the distribution.

tions.[6] It is easier to estimate θ through observation of $\theta + \varepsilon$ when the noise term ε is less dispersed. It may seem intuitive that the maximum of k variables tends to be more concentrated toward the top and less concentrated toward the bottom. However, with the correct model of selection, the evaluator can adjust for the upward bias. The question is whether the distribution becomes less dispersed after this adjustment. Gumbel's extreme value distribution is precisely such that its shape is unaffected by selection. For distribution functions with quantile density less elastic than for the Gumbel distribution, their upper tails are brought leftward by the evaluator's adjustment for selection, and hence the evaluator benefits.

The criterion in Theorem 1 asks for a monotone change in the evaluator's payoff, uniformly across possible preference parameters and prior beliefs. It therefore only permits a partial comparison of experiments affected by selection. For many distributions of interest, the preference of the evaluator for more or less selection depends on the parameters. Section 4 turns to a local comparison of the value of information based on the notion of dispersion. Our second main contribution (Theorem 2) provides a tool for comparing experiments whose information constraints cross. To compare two experiments for a neighborhood of parameter values, it suffices to check for dispersion in a corresponding neighborhood of the distributions underlying the two experiments (Proposition 2). This local dispersion criterion thus provides a generalization of Lehmann's results for dichotomies; Appendix B develops a more generally applicable result that uses local dispersion to compare experiments that have quantile difference with an arbitrary non-monotonic pattern.

If, for instance, the quantile density is inelastic (relative to the Gumbel) at low quantiles but elastic at high quantiles, then selection harms evaluators with a strong preference against false acceptance. Intuitively, those evaluators set an acceptance threshold at high observations of $\theta + \varepsilon$, corresponding to the top part of the distribution of ε, and thus are harmed by the increase in mass for high quantiles created by selection. Overall, the evaluator is harmed when the data distribution has sufficiently thick right tail (as for the Laplace distribution) and the hypothesis

[6]For dichotomies, Lehmann's order corresponds to the well-known Blackwell ordering of experiments. However, Lehmann's notion is more relevant for our study. Theorem 1 remains valid when the state variable is real—however, the equivalence to Blackwell fails for general composite hypothesis testing problems.

would be rejected at the prior. Symmetrically, the evaluator is harmed by selection when the left tail is sufficiently thin (as for the uniform distribution) and the prior is to accept the hypothesis.

Section 5 turns to analyze the case with large selection, when k tends to infinity. Drawing on extreme value theory, Theorem 3 characterizes the limit impact of selection. For illustration, when F is normal, we show that the evaluator in the limit is able to identify the true state on the basis of one, extremely selected observation. In this case, the evaluator thus obtains the highest possible payoff where both error rates are zero. By contrast, $-\log(-\log F)$ is concave for the exponential distribution, so selection increasingly harms the evaluator (Theorem 1) and information in the limit is less than full. We show that for all strictly logconcave distributions in the exponential power family achieves full information in the limit with extreme selection. Thus, unless the noise distribution is exactly exponential in the upper tail, extreme selection benefits the evaluator for all parameter values—however, the rate of convergence is extremely slow.

Section 6 provides a strategic foundation for sample selection. A researcher aims at demonstrating that a treatment is effective. The researcher's incentives to bias upward the evaluator's inference through sample selection are anticipated in equilibrium by the evaluator. Under the global condition of Theorem 1, we show that equilibrium selective sampling benefits also the researcher in the empirically relevant case where the prior strongly favors rejection.[7] Instead, when the errors have thick right tail and selection is mild (small k), equilibrium selective sampling harms not only the evaluator but also the researcher when the prior strongly favors rejection—generating a credibility crisis.[8]

We also endogenize the amount of selection in terms of costly investment by the researcher in obtaining pre-sample k realizations from which selection is made. The evaluator's anticipation of the extent of selection k and the resulting adjustment for selection bias at least partly frustrates the researcher's attempt to manipulate. When selection is fully anticipated in equilib-

[7] In this case, we also show that equilibrium selection harms the researcher when the prior strongly favors acceptance.

[8] In this case, we also show that equilibrium selection benefits the researcher when the prior strongly favors acceptance. The researcher is harmed by selection for intermediate priors.

rium (for example because the researcher's cost of pre-sample collection is known), pre-sample collection and selection is a pure rat race whenever the noise follows a Gumbel distribution. In that case, correctly anticipated presample collection and selection has no impact on the acceptance probability and decision payoffs of evaluator and researcher, and so results in a loss by the researcher exactly equal to positive cost of presample collection. Thus, in Gumbel's pure rat race the researcher unambiguously benefits from commitment not to allow (or, equivalently to disclose) presample collection.[9]

Section 7 turns to the welfare impact of unanticipated selection. Clearly, if the noise distribution is Gumbel, unanticipated selection always hurts the evaluator. We characterize conditions under which unanticipated selection benefits the evaluator. We showcase a notable situation in which unanticipated selection leaves the evaluator exactly indifferent: whenever the noise distribution is symmetric (as with normal, logistic, or uniform noise) and equipoise holds at the prior (an ethical condition requiring indifference between acceptance and rejection), an evaluator who observes the maximum of $k = 2$ realizations while expecting random realization obtains the same payoff as when observing (and correctly expecting) a random realization. Intuitively, unexpected selection results in an increased probability of acceptance—however, by symmetry, the loss associated to the increase in false positives (higher α) is exactly offset by the benefit associated to the reduction in false negatives (lower β). In addition, Section 7 shows that uncertainty in the extent of selection tends to damage the evaluator. Section 8 concludes by contrasting our contribution to related work in the literature.

2 Statistical Setting

An evaluator is interested in the true value of an unknown binary state $\theta \in \{\theta_L, \theta_H\}$, where θ_L and $\theta_H > \theta_L$ are real numbers, and assigns prior probability p to state θ_H. The evaluator faces

[9] More generally, selection is not completely self defeating, even when the evaluator correctly anticipates the extent of selection k. Our results characterize the net impact of properly anticipated selection on acceptance probability (the researcher's decision payoff) and the evaluator's decision payoff.

a binary choice, to accept or to reject. Acceptance results in payoff θ, while rejection gives a safety payoff R, where $\theta_L < R < \theta_H$.

Information and Optimal Decision. Before making a choice, the evaluator observes an informative signal about the true state, $x = \theta + \varepsilon$. The noise term, ε, is independent of θ and drawn from a known cumulative distribution function F admitting a log-concave density f, that is, $\log(f)$ is a concave function. This assumption implies the monotone likelihood ratio property—the ratio $\ell_F(x) \equiv f(x-\theta_H)/f(x-\theta_L)$ is increasing—and hence monotonicity of the optimal decision.[10] The evaluator accepts if and only if x is at least as large as some threshold $\bar{x}$. The optimal threshold is $\bar{x}_F^* \equiv \min\left\{x : \ell_F(x) \geq \bar{\ell}\right\}$, where

$$\bar{\ell} \equiv \frac{1-p}{p}\frac{R-\theta_L}{\theta_H - R}.$$

Reformulation. The evaluator's problem can be viewed as a simple trade-off between correct choices — rejection in state θ_L and acceptance in state θ_H. Any given threshold $\bar{x}$ induces a probability of incorrect acceptance $\alpha = 1 - F(\bar{x} - \theta_L)$ in state θ_L and a probability of incorrect rejection $\beta = F(\bar{x} - \theta_H)$ in state θ_H. The corresponding payoff is, disregarding constants, a linear and strictly decreasing function of α and β,

$$-(1-p)(R-\theta_L)\alpha - p(\theta_H - R)\beta.$$

Thus, we can view $(1-p)(R-\theta_L)$ as the marginal cost of an incorrect acceptance (or the marginal benefit of a correct rejection) and $p(\theta_H - R)$ as the marginal cost of an incorrect rejection (or the marginal benefit of a correct acceptance). The ratio $\bar{\ell}$ then measures the relative marginal cost of an incorrect acceptance (i.e. relative to an incorrect rejection). The evaluator can achieve any pair (α, β) such that $\alpha = 1 - F(\bar{x} - \theta_L)$ and $\beta = F(\bar{x} - \theta_H)$ for some $\bar{x}$, or

$$\beta = \beta_F(\alpha) \equiv F\left(F^{-1}(1-\alpha) + \theta_L - \theta_H\right), \tag{1}$$

which we call the *information constraint* of experiment F. Note that $\beta_F(\alpha) \leq 1 - \alpha$, because $\theta_H > \theta_L$. Moreover, $\beta_F(\alpha)$ is decreasing and convex. Indeed, by decreasing the threshold $\bar{x}$

[10] Since we consider location experiments with arbitrary $\theta_H > \theta_L$, the monotone likelihood ratio property is not only necessary but also sufficient for log-concavity of the error distribution

the evaluator accepts more often in both states, increasing α but decreasing β. Furthermore, the relative rate of change, the slope of $\beta_F(\alpha)$ at $\alpha = 1 - F(\bar{x} - \theta_L)$, is the negative of the likelihood ratio at the threshold $\bar{x}$, that is,

$$\frac{d\beta_F(\alpha)}{d\alpha} = -\frac{f\left(F^{-1}(1-\alpha)+\theta_L-\theta_H\right)}{f\left(F^{-1}(1-\alpha)\right)} = -\frac{f(\bar{x}-\theta_H)}{f(\bar{x}-\theta_L)}.$$

We conclude that the problem of the evaluator is equivalent to choosing α and β to minimize the linear objective function $\bar{\ell}\alpha + \beta$ subject to the decreasing, convex constraint $\beta = \beta_F(\alpha)$.

Random vs Selected Experiment. Our goal is to compare the following two scenarios in terms of evaluator's welfare. In the first scenario, the evaluator observes a random sample: the noise term ε is drawn from a distribution F. In the second scenario, the evaluator observes the result of a selected experiment, where ε is the maximum of k independent draws from F. In this case, the cumulative distribution function of the noise term is F^k, with density $kF^{k-1}f$. The latter inherits log-concavity from f, hence in the selected experiment the evaluator also uses a cutoff rule, but the optimal threshold is higher, and increasing in k. The evaluator accepts if and only if $\ell_{F^k}(x) = [F(x-\theta_H)/F(x-\theta_L)]^{k-1}\ell_F(x) \geq \bar{\ell}$, and the term in square brackets is increasing in k for any x. By our reformulation, the evaluator's optimal decision minimizes $\bar{\ell}\alpha + \beta$ subject to $\beta = \beta_{F^k}(\alpha)$. Compared to the first scenario ($k = 1$), selection has two opposite effects on the probability of acceptance in the two states, and hence contrasting effects on the value of the evaluator's problem. Raising the threshold tends to decrease acceptance in both states, that is, decreases α and increases β. But F^k first-order stochastically dominates F, which tends to increase acceptance in both states, increasing α and decreasing β.

3 Uniform Impact of Selection

How is the welfare impact of selection related to the noise distribution and to the parameters of the evaluator's problem? Toward a full answer to this question, we first provide a characterization of noise distributions F for which the extent of selection, k, has a *uniform* and *monotone* effect on the evaluator's payoff—uniformly over all parameter values, an increase in k makes the evaluator better off, or worse off.

Dispersion Ordering of Selected Experiments. Our characterization is based on the following notion of stochastic ordering, first proposed by Bickel and Lehmann (1979).[11] An experiment G is *less dispersed* than an experiment F if the quantile difference $G^{-1}(u) - F^{-1}(u)$ is a weakly decreasing function of u, that is,

$$f\big(F^{-1}(u)\big) \leq g\big(G^{-1}(u)\big) \qquad \text{for all } 0 < u < 1. \tag{2}$$

Our first theorem compares the dispersion of F^k as k varies. An increase in k results in lower dispersion if the slope at any quantile becomes steeper, as in (2). This comparison of slopes at quantiles is simplified if we first make a transformation of the distribution, such that changing k determines a parallel shift of the transformed function. The suitable transformation of $u = F^k$ is given by the increasing function $-\log(-\log(u))$, for $-\log(-\log(F^k)) = -\log(-\log(F)) - \log(k)$. As k increases, the requisite slope is steeper precisely if $-\log(-\log(F))$ is convex.

Theorem 1. *F^k is less dispersed the greater is $k \geq 1$ if and only if $-\log(-\log(F))$ is convex. Likewise, F^k is more dispersed the greater is $k \geq 1$ if and only if $-\log(-\log(F))$ is concave.*

Proof. Observe that $-\log(-\log(u))$ is a strictly increasing, differentiable function from $(0,1)$ to the real line. Convexity of $-\log(-\log(F))$ is equivalent to a non-decreasing derivative of this function. Equivalently, for any $\varepsilon_1 < \varepsilon_2$ where $0 < F(\varepsilon_1) < F(\varepsilon_2) < 1$, the slope of $-\log(-\log(F(\varepsilon)))$ is no smaller at ε_2 than at ε_1. That $0 < F(\varepsilon_1) < F(\varepsilon_2) < 1$ is equivalent to the existence of $k > 1$ such that $F(\varepsilon_1) = F^k(\varepsilon_2)$. The slope of $-\log\left(-\log\left(F^k(\varepsilon)\right)\right)$ at ε_1 equals the slope of $-\log(-\log(F(\varepsilon)))$ at ε_2. When $F(\varepsilon_1) = F^k(\varepsilon_2)$, a steeper slope of $-\log\left(-\log\left(F^k(\varepsilon)\right)\right)$ at ε_2 than of $-\log(-\log(F(\varepsilon)))$ at ε_1 is equivalent to a steeper slope of $F^k(\varepsilon)$ at ε_2 than of $F(\varepsilon)$ at ε_1. By (2) this steeper slope of F^k is equivalent to less dispersion. $\square$

There is only one distribution F such that $-\log(-\log(F))$ is linear, namely Gumbel's extreme value distribution, $F(\varepsilon) = \exp(-\exp(-\varepsilon))$. This distribution, which plays a special role in our ensuing analysis, is such that for every k the experiment F^k is neither less nor more dispersed than F.

[11] They, in turn, credit Brown and Tukey (1946) for the essence of this definition.

Next, note that the derivative of the transformation $-\log(-\log(u))$ is given by $-1/(u\log(u))$. Taking the next derivative, when f is differentiable, it is not hard to verify that convexity of $-\log(-\log(F))$ can be restated as

$$\frac{f'(\varepsilon)/f(\varepsilon)}{f(\varepsilon)/F(\varepsilon)} > \frac{1+\log F(\varepsilon)}{\log F(\varepsilon)}. \tag{3}$$

This condition has another interpretation. Denote the quantile function by $Q(u) = F^{-1}(u)$. Its derivative is the quantile density $q(u) = 1/f\left(F^{-1}(u)\right)$, also known as Tukey's sparsity function. The elasticity of the quantile density function is $uq'(u)/q(u)$. It can be directly computed that, when q is differentiable,

$$\frac{uq'(u)}{q(u)} = -\frac{uf'\left[F^{-1}(u)\right]}{f^2\left[F^{-1}(u)\right]} = -\frac{f'\left(F^{-1}(u)\right)/f\left(F^{-1}(u)\right)}{f\left(F^{-1}(u)\right)/F\left(F^{-1}(u)\right)}.$$

For Gumbel's extreme value distribution, the quantile function is $Q(u) = -\log(-\log u)$ and hence $q(u) = -1/(u\log u)$. Thus, the elasticity of its quantile density function is

$$\frac{uq'(u)}{q(u)} = -\frac{u(1+\log u)/(u\log u)^2}{1/(u\log u)} = -\frac{1+\log u}{\log u}.$$

Summing up, our key condition (3) can also be rephrased as requiring that distribution F has a quantile density function that is less elastic than that of Gumbel's extreme value distribution.

Finally, while we described selection with a natural number k, the interpretation—as well as the statement in the theorem—for real numbers $k > 1$ is equally valid. If an experiment F is such that selection monotonically benefits (or hurts) the evaluator, then F^k has the same property for every real number $k > 1$. Compared to the distribution with selection k, the distribution with more selection $k' > k$ is $F^{k'} = (F^k)^{k'/k}$. The change is akin to starting from F^k and applying selection of extent $k'/k > 1$.

Uniform Impact of Selection on Welfare. We now use Theorem 1 to characterize the distributions F such that, for all values of the parameters of the problem, more selection benefits the evaluator, and those such that more selection always hurts. The characterization exploits Lehmann's (1988) criterion for comparing location experiments. Consider two experiments, F and G. An immediate separation argument shows that the evaluator is better off with the second experiment, uniformly over all values of θ_L, θ_H, p and R, if and only if, for all values of θ_L

and θ_H, the information constraint of G lies entirely below that of F. Lehmann (1988) proved that this property holds if and only if G is less dispersed than F. Given this result, the next proposition is an immediate consequence of Theorem 1.

Proposition 1. *The evaluator is better off (worse off) with experiment $F^{k'}$ than with experiment F^k, for all $k' \geq k \geq 1$ and all values of θ_L, θ_H, p and R, if and only if $-\log(-\log(F))$ is convex (resp. concave).*

The logic of Lehmann's criterion is simple. Let $\bar{x}$ be any cutoff that the evaluator may choose in experiment F, with $\alpha = 1 - F(\bar{x} - \theta_L)$ and $\beta = F(\bar{x} - \theta_H)$ the corresponding incorrect acceptance and rejection rates. Since $\beta_F(\alpha) \leq 1 - \alpha$ and the quantile difference is decreasing, $G^{-1}(1-\alpha) - F^{-1}(1-\alpha) \leq G^{-1}(\beta) - F^{-1}(\beta)$. This inequality says that the cutoff adjustment needed to induce the same α under G, is smaller than the one needed to induce the same β. That is, defining $\bar{x}'$ and $\bar{x}''$ by $\alpha = 1 - G(\bar{x}' - \theta_L)$ and $\beta = G(\bar{x}'' - \theta_H)$, we have $\bar{x}'' \geq \bar{x}'$. But this means that in experiment G the evaluator can, by choosing $\bar{x}'$, achieve the same probability of incorrect acceptance as in F, namely α, but a lower probability of incorrect rejection, namely $G(\bar{x}' - \theta_H) \leq G(\bar{x}'' - \theta_H) = \beta$. In other words, the information constraint of G is always lower: $\beta_G(\alpha) \leq \beta_F(\alpha)$ for every α. When $G = F^k$ and $-\log(-\log(F))$ is convex, this means that the evaluator always prefers experiment F^k to experiment F.

If the noise follows Gumbel's extreme value distribution, the evaluator is indifferent to selection. An intuitive argument also shows this. With selection at extent k, the noise distribution is $F^k(\varepsilon) = \exp(-k\exp(-\varepsilon)) = F(\varepsilon - \log(k))$. Compared to a random sample, in the selected experiment the noise is simply raised by the constant $\log(k)$. The evaluator adjusts for this translation of the noise distribution, and is back to square one.[12]

Examples. For the normal distribution, it can be verified that condition (3) is met, and hence more selection benefits the evaluator. A plot of the information constraints corresponding to F and F^2 illustrates this in the right panel of Figure 1. The left panel of Figure 1 allows us to

[12]It is well known that a location experiment is unaffected by a change of the location of the noise ε, but not to a change in its scale. However, it can be seen by inspection of the elasticity of the quantile density function, that the uniform criterion in Theorem 1 is unaffected by a change in scale.

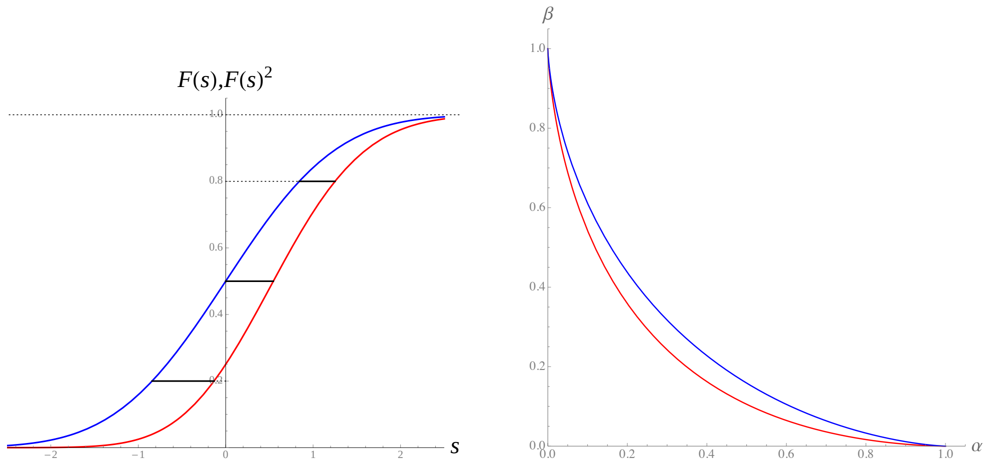

Figure 1: Comparison between F (blue) and F^2 (red) for the normal distribution.

revisit the intuitive explanation of Theorem 1. Going from F^k to F^{k_1} where $k_1 > k$ is a simple operation that takes F^k to the power $k_1/k > 1$. As indicated by the horizontal guiding lines in the figure, Lehmann's dispersion notion requires that F^{k_1} is steeper than F^k when evaluated at the same quantile $F^{k_1}(\varepsilon_2) = F^k(\varepsilon_1)$. The slopes of the functions are easily found via the implicit function theorem, and the comparison where the relative power k_1/k tends to 1 provides condition (3).

If the distribution is logistic, $F(\varepsilon) = 1/(1+e^{-\varepsilon})$, we have $Q(u) = \log[u/(1-u)]$ and hence $q(u) = 1/[u(1-u)]$. Thus

$$\frac{uq'(u)}{q(u)} = -\frac{1-2u}{1-u} < -\frac{1+\log u}{\log u},$$

so F is more convex than the Gumbel distribution. Here, too, any amount of selection benefits the evaluator, and benefits even more as k increases.

For the exponential distribution, $F(\varepsilon) = 1 - e^{-\varepsilon}$ where $\varepsilon \geq 0$, so $Q(u) = -\log(1-u)$ and hence $q(u) = 1/(1-u)$. Thus

$$\frac{uq'(u)}{q(u)} = \frac{u}{1-u} > -\frac{1+\log u}{\log u}$$

so this distribution is more concave than the Gumbel distribution—any amount of selection harms the evaluator, ever more as k increases. This result for the exponential distribution also follows from the sufficient condition obtained by Khaledi and Kochar's (2000) Theorem 2.1, according to which for any distribution with *decreasing* hazard rate higher order statistics are more dispersed.[13] Given that logconcavity implies that the hazard rate is increasing by Prekopa's theorem, the only logconcave distribution for which Khaledi and Kochar's (2000) result applies is the exponential (loglinear) distribution, which has constant hazard rate.[14] Theorem 1's characterization applies more generally to logconcave distributions—intuitively, it shows that selection benefits the evaluator whenever the hazard rate is more increasing than in the Gumbel distribution.

4 Local Comparison of Experiments

Lehmann's comparison gives only a partial ordering of location experiments, because it imposes a uniform criterion over evaluator preferences. For example, as we shall see, the comparison fails for F and $G = F^k$ when F is uniform or Laplace. In these cases, the quantile difference is not monotonic. We develop a local version of Lehmann's criterion and use it to develop a complete characterization of when the evaluator gains or loses from selection.

Example: Selection with a Uniform Signal. How do we compare experiments whose quantile difference is not monotonic? Does the pattern of the quantile difference still characterize the evaluator's preference, and does it allow us to pinpoint the region of parameter values where G (or F) is preferred? Intuition based on Lehmann's logic suggests an answer to these questions.

To illustrate this intuition, we start with a simple example. Suppose we are interested in comparing F with $G = F^2$ when F is uniform, $F(\varepsilon) = \varepsilon$ for $\varepsilon \in [0,1]$. Since $F^2(\varepsilon) = \varepsilon^2$, the quantile difference $(F^2)^{-1}(u) - F^{-1}(u) = \sqrt{u} - u$ is bell-shaped, and hence F and F^2 are not

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

[14] Theorem 1 also covers distributions with decreasing hazard rate, for which $-\log(-\log(F))$ is necessarily concave.

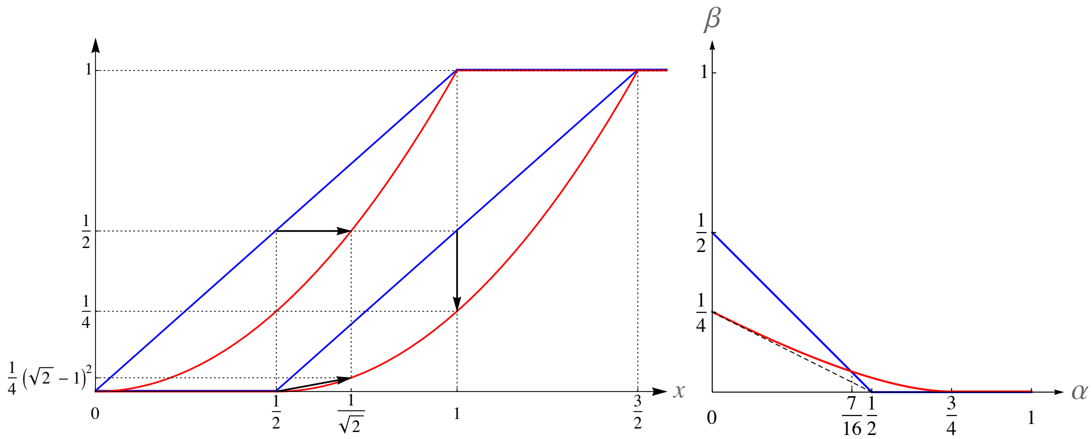

Figure 2: Comparison between F (blue) and F^2 (red) for the uniform distribution.

comparable in the sense of Lehmann.

Now suppose that $\theta_L = 0$ and $\theta_H = 1/2$, so there is some overlap in the supports of signal distributions $F(x-\theta_L)$ and $F(x-\theta_H)$, and hence also of signal distributions $F^2(x-\theta_L)$ and $F^2(x-\theta_H)$.[15] The four signal distributions are depicted in the panel on the left-hand side of Figure 2.

The information constraints of F and F^2 are, respectively, $\beta_F(\alpha) = 1/2 - \alpha$ (for $0 \leq \alpha \leq 1/2$) and

$$\beta_{F^2}(\alpha) = \left(\sqrt{1-\alpha} - 1/2\right)^2 \quad \text{(for } 0 \leq \alpha \leq 3/4\text{)}.$$

The constraints cross only once, with $\beta_{F^2}(\alpha) \leq \beta_F(\alpha)$ for $0 \leq \alpha \leq 7/16$ and $\beta_F(\alpha) \leq \beta_{F^2}(\alpha)$ for $7/16 \leq \alpha \leq 1$, as illustrated in the panel on the right-hand side of Figure 2. The picture tells us that the evaluator prefers F to F^2 when the relative weight attached to type I errors is less than $1/2$, and F^2 to F when it is greater than $1/2$. When the weight is exactly $1/2$, we have indifference, as the lowest iso-payoff line (the dashed line) touches both constraints.

The key observation is that, even without explicitly computing the information constraints, we can infer their single crossing by the bell-shape of the quantile difference. Consider again the

[15] Our argument does not depend on the particular choice of θ_L and θ_H. In case $\theta_H - \theta_L > 1$, the evaluator is indifferent between F and F^2, as each experiment perfectly reveals the state.

panel on the left-hand side of Figure 2. If the evaluator attaches a large relative weight to type I errors, then in experiment F the optimal threshold of acceptance is $\hat{x} = 1$, which corresponds to $\alpha = 0$ and $\beta = 1/2$. By adopting the same threshold, $\hat{x} = 1$, the evaluator obtains $\alpha = 0$ also in experiment F^2. But the type II error is lower, $\beta = 1/4 < 1/2$, as highlighted by the vertical arrow in the picture. By continuity of the information constraints, this implies that $\beta_{F^2}(\alpha)$ must lie below $\beta_F(\alpha)$ for small values of α. In other words, the evaluator prefers F^2 to F when the concern for type I errors is relatively high, that is, p is small or R is large.

Assume now that the evaluator attaches a small relative weight to type I errors. Then the optimal threshold for F becomes $\hat{x} = 1/2$, corresponding to $\alpha = 1/2$ and $\beta = 0$. As the picture illustrates, in order to obtain $\alpha = 1/2$ under experiment F^2 the evaluator must increase the threshold to $\hat{x} = 1/\sqrt{2}$. This time, however, the corresponding type II error is higher, $\beta = (\sqrt{2}-1)^2/4 > 0$. Thus, $\beta_{F^2}(\alpha)$ lies above $\beta_F(\alpha)$ when α is large. This means that the evaluator prefers F to F^2 when type I errors are relatively less important, that is, p is large or R is small.

Finally, the bell-shape of the quantile difference guarantees that there is no gap between the two regions of values of p and R identified above, that is, it guarantees that the information constraints of F^2 and F cross only once. Indeed, since $(F^2)^{-1}(u) - F^{-1}(u)$ is first strictly increasing and then strictly decreasing, there exists a unique pair of thresholds $\hat{x}$, $\hat{x}'$ such that, by adopting $\hat{x}$ in experiment F and $\hat{x}'$ in experiment G, both type I and type II errors are the same, that is, $F(\hat{x} - \theta_L) = G(\hat{x}' - \theta_L)$ and $F(\hat{x} - \theta_H) = G(\hat{x}' - \theta_H)$.

Local Dispersion. Let us now consider a generic pair of experiments F and G. The example above suggests that neighborhoods of values of u where G is *locally* less dispersed than F, that is, where the quantile difference $G^{-1}(u) - F^{-1}(u)$ is decreasing, should correspond to neighborhoods of values of α such that $\beta_G(\alpha) \leq \beta_F(\alpha)$. That is, they should correspond to regions of parameter values such that, for those parameter values, G is better than F. Moreover, if the neighborhood is a set of large values of u, then the corresponding region of parameters should, in fact, be one where the evaluator finds it harder to accept, that is, a region where p is small, or R is large.

Our next result shows that the argument illustrated in our example above indeed generalizes, to all cases where F and G have the same support and the slope of the quantile difference $G^{-1}(u)-F^{-1}(u)$ changes sign only once.[16] With multiple changes of sign, the intuition above remains essentially correct, but the correspondence between decreasing parts of the quantile difference and regions of parameters where the evaluator prefers G to F becomes subtler. We analyze this more general case in Appendix B.

Theorem 2. *Let $\bar{u} \in (0,1)$ and suppose that $G^{-1}(u)-F^{-1}(u)$ is increasing for $u \in (0,\bar{u})$ and decreasing for $u \in (\bar{u},1)$. Suppose also that F and G have either both unbounded supports, or both the same, bounded support. Then for every θ_L and θ_H there exists $\bar{K}>0$ such that F is preferred to G for*

$$\frac{1-p}{p}\frac{R-\theta_L}{\theta_H-R} > \bar{K},$$

and G is preferred to F for

$$\frac{1-p}{p}\frac{R-\theta_L}{\theta_H-R} < \bar{K}.$$

Proof. It suffices to show that there exists $\bar{\alpha}$ such that $\beta_G(\alpha)-\beta_F(\alpha)$ is nonnegative for $\alpha \geq \bar{\alpha}$ and nonpositive for $\alpha \leq \bar{\alpha}$. Fix θ_L and $\theta_H > \theta_L$. If the support of F and G is a bounded interval of length smaller than $\theta_H-\theta_L$, then both F and G perfectly reveal the state. In this case, $\beta_G(\alpha)-\beta_F(\alpha)=0$ for every α. The evaluator is indifferent between F and G for all values of q and R, and $\bar{K}$ can be chosen arbitrarily. Otherwise, since $G^{-1}(u)-F^{-1}(u)$ is increasing for $u \in (0,\bar{u})$ and decreasing for $u \in (\bar{u},1)$, there exists $\bar{\varepsilon}$ such that the difference

$$\left[G^{-1}(F(\varepsilon+\theta_H-\theta_L))-(\varepsilon+\theta_H-\theta_L)\right]-\left[G^{-1}(F(\varepsilon))-\varepsilon\right]$$

is nonnegative for $\varepsilon \leq \bar{\varepsilon}$ and nonpositive for $\varepsilon \geq \bar{\varepsilon}$. Changing variable to $\alpha = 1-F(\varepsilon+\theta_H-\theta_L)$, this is the same as saying that there exists $\bar{\alpha}$ such that $\beta_G(\alpha)-\beta_F(\alpha)$ is nonnegative for $\alpha \geq \bar{\alpha}$ and nonpositive for $\alpha \leq \bar{\alpha}$. □

Thus, when the quantile difference is first increasing and then decreasing (or vice versa), its slope at small or large values of u determines the evaluator's preference over G and F when the

[16] Graphically, like in the right panel of Figure 2, it is easy to see that the evaluator's preference as a function of the error trade-off parameter K switches from one experiment to the other if and only if the information constraints cross between the touch-points of the tangent with slope K.

relative weight attached to type I errors is small or large, respectively. For example, suppose that we would like to check whether G is preferred to F for a given p and for all R above a certain value (or for a given R and for all p below some value). The theorem highlights that checking whether this preference indeed holds only requires us to verify that $G^{-1}(u)-F^{-1}(u)$ is decreasing for u above a certain bound. An immediate corollary of this observation is that, perhaps remarkably, the same preference would then hold for any other distributions G' and F' with the same shape as G and F above that bound.

Local Ordering of Order Statistics. Thanks to the local condition developed above, we can now refine Theorem 1 by considering the following proposition. Since F and F^k have the same support, the result is an immediate consequence of Theorem 2. Afterwards, we will illustrate this method for the case of Laplace and uniform distributions.

Proposition 2. *Let $\bar{u} \in (0,1)$ and suppose that the elasticity of q is greater (less) than that of the Gumbel distribution for $u \in (0,\bar{u})$, and less (greater) for $u \in (\bar{u},1)$. Then for every $\ell > k \geq 1$ and every $\theta_H > \theta_L$ there exists $\bar{K} > 0$ such that the evaluator is worse (better) off with F^ℓ than with F^k when*

$$\frac{1-p}{p}\frac{R-\theta_L}{\theta_H-R} < \bar{K},$$

and better (worse) off when

$$\frac{1-p}{p}\frac{R-\theta_L}{\theta_H-R} > \bar{K}.$$

Intuitively, we may describe $(F^\ell)^{-1}(u)-(F^k)^{-1}(u)$ as bell-shaped when the elasticity of the quantile density of F is greater than that of the Gumbel distribution for small values of u, and smaller for large values of u. This case was illustrated by the uniform distribution in Section **??**.

We now illustrate the opposite case, where the elasticity of q is first smaller, then greater than that of the Gumbel distribution. Suppose again that $\ell = 2$ and $k = 1$, and consider the Laplace distribution where $F(\varepsilon) = (1/2)e^{\varepsilon}$ for $\varepsilon < 0$ and $F(\varepsilon) = 1-(1/2)e^{-\varepsilon}$ for $\varepsilon \geq 0$. In the left-hand side panel of Figure 3 we plot F (blue) and F^2 (red), which reveals the U-shape of $(F^2)^{-1}(u)-F^{-1}(u)$. In this example, there is a bound to the posterior belief after seeing any signal realization, and hence the information constraints are not tangent to the axes as $\alpha \to 0$

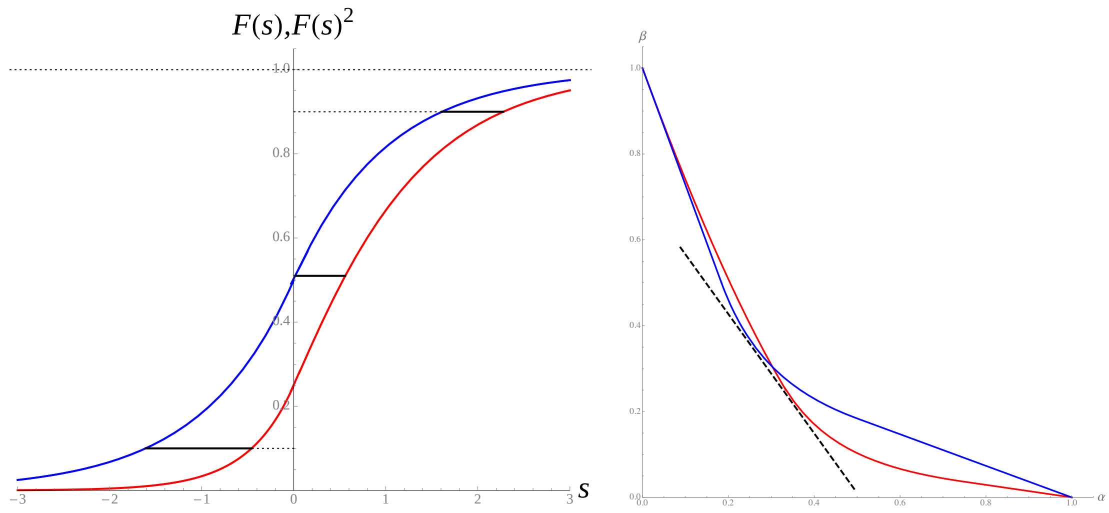

Figure 3: Comparison between F (blue) and F^2 (red) for the Laplace distribution.

and $\alpha \to 1$. As shown in the right-hand side panel of Figure 3, drawn for $\theta_L = 0$ and $\theta_H = 1$, the evaluator prefers F^2 to F when the ratio $(1-p)(R-\theta_L)/[p(\theta_H - R)]$ is small, and F to F^2 when it is large.

5 Extreme Selection

In this section, we examine the effect of extreme selection, $k \to \infty$, on the evaluator's payoff. To gain some initial intuition, recall from Theorem 1 that in the Gumbel distribution, the evaluator's payoff is constant in k. Next, the fundamental result in extreme value theory says that the distribution of the maximum of k i.i.d. random variables, properly adjusted for location and scale inflation, either does not converge weakly to any nondegenerate distribution (for any choice of adjustment) or it converges weakly to a distribution $\bar{F}$ that must belong to one of the following three types: Gumbel, Extreme Weibull or Frechet. (See e.g. Leadbetter et al. (1983) for a reference.) More precisely, for some sequence $a_k > 0$ and b_k,

$$F^k\left(b_k + a_k\varepsilon\right) \to \bar{F}\left(\varepsilon\right)$$

for every continuity point ε of $\bar{F}$.

Note that F^k is decreasing in k wherever $F \in (0,1)$. Hence, the distribution of ε is systematically shifted upwards as k increases, in the sense of first-order stochastic dominance. Hence, the location adjustment sequence $-b_k$ is growing. However, the evaluator can adjust for any translation of the error distribution without any impact on payoff.

The limit impact of selection thus hinges on whether the sequence a_k shrinks to zero or not. If $a_k \to 0$, the error distribution is less and less dispersed, providing the evaluator with arbitrarily precise information about the state.

To formalize this claim, denote by V^k the payoff of the evaluator when selection is at level k.

Theorem 3. *If $F^k(b_k + a_k\varepsilon) \to \bar{F}(\varepsilon)$ at every continuity point of $\bar{F}$, with $a_k \to 0$, then $V^k \to p\theta_H + (1-p)R$, the full information payoff. If $a_k = 1$ for all k, then V^k converges to the payoff from observing one non-selected observation with error distribution $\bar{F}$.*

It is well known that many familiar distributions are in the basin of attraction of the Gumbel distribution. Specifically, when F is normal—or half-normal, which has the same right tails—then a_n must be decreasing to zero ($a_n = (2\log n)^{-1/2}$ works), and $\bar{F}$ is the Gumbel distribution.

More generally, when F is exponential power (or folded exponential power) with shape parameter $b > 1$, a_n must be decreasing to zero ($a_n = (b\log n)^{-(b-1)/b}$ works, as we show in the appendix), and $\bar{F}$ is the Gumbel distribution. This result is striking, because it is also known that when F is the exponential distribution—or the Laplace distribution, since the two distributions have the same right tails—then it also converges to the Gumbel distribution, but we can take $a_k = 1$ for each k. This is the exponential power distribution with shape parameter $b = 1$.

So, while extreme selection leads to full information as $k \to \infty$ for any $b > 1$ in this family, the limit result is very different when $b = 1$. The uniformly negative impact of selection in the exponential case discussed earlier is, in this sense, non-generic, as any arbitrarily close distribution in the family reverses the conclusion.[17]

[17] Of course, as b approaches 1 from above, the convergence to full information gets slower.

6 Strategic Selection and Equilibrium Impact on Researcher

Sample selection of the sort considered above naturally arises as an equilibrium phenomenon in a strategic setting where the experiment is carried out by a researcher who is entirely indifferent to the type I error α. Intuitively, such a researcher biased toward acceptance wants to select an individual with a high error term—e.g., in the treatment effect setting, a good untreated outcome—in order to bias upward the experimental result and thus increase the chances of acceptance. As we have seen, an evaluator taking this behavior into account may suffer or benefit, compared to the case of a random sample. In this section, we verify that the posited behavior constitutes an equilibrium in this game, and we discuss how the researcher's ability to strategically select the sample affects the researcher's own welfare, too.

Model. Consider the following timeline:

1. The researcher privately observes $\varepsilon_1, \ldots, \varepsilon_k$ and then chooses $i \in \{1, \ldots, k\}$.
2. The evaluator observes $x_i = \theta + \varepsilon_i$ and then chooses whether to accept or reject.

As before, the evaluator receives a fixed payoff R when rejecting, and θ when accepting. The researcher receives 0 if the evaluator rejects, and 1 if the evaluator accepts.

Lemma 1. *There exists a Bayes Nash equilibrium where the researcher chooses maximal selection, $i \in \arg\max_{1 \le j \le k} \varepsilon_j$. The evaluator accepts for signals above the uniquely determined threshold $\hat{x}_k$, defined by*

$$\frac{F^{k-1}(\hat{x}_k - \theta_H) f(\hat{x}_k - \theta_H)}{F^{k-1}(\hat{x}_k - \theta_L) f(\hat{x}_k - \theta_L)} = \frac{1-p}{p} \frac{R - \theta_L}{\theta_H - R}. \tag{4}$$

Note that the evaluator's strategy is precisely the one we analyzed until now, when observing x subjected to selection of degree k. The researcher's strategy is a best response because the evaluator will observe a higher signal and hence be more likely to accept.[18]

[18] The researcher is indifferent when $\max\{\varepsilon_1, \ldots, \varepsilon_k\} < \hat{x} - \theta_H$ or $\min\{\varepsilon_1, \ldots, \varepsilon_k\} > \hat{x} - \theta_L$.

Distribution of Evaluator's Posterior Expectation. The equilibrium effect of changing the distribution of ε works through its effect on the distribution of the rational evaluator's posterior belief. Given $k \geq 1$ and an observation x, the posterior expectation of θ is

$$\pi_k(x) := \frac{pF^{k-1}(x-\theta_H)f(x-\theta_H)\theta_H + (1-p)F^{k-1}(x-\theta_L)f(x-\theta_L)\theta_L}{pF^{k-1}(x-\theta_H)f(x-\theta_H) + (1-p)F^{k-1}(x-\theta_L)f(x-\theta_L)}.$$

This is monotone increasing in the signal realization x, so it is easy to compute its c.d.f.,

$$\tilde{F}_k(\pi) := \int 1_{\{\pi_k(x)\leq\pi\}} dF^k(x).$$

Lemma 2. *The researcher's expected payoff is*

$$U^k(R) := 1 - \tilde{F}_k(R),$$

while the evaluator's expected payoff is

$$V^k(R) := R + \int_R^{\theta_H} [1-\tilde{F}_k(r)]dr.$$

The evaluator accepts if and only if $x \geq \hat{x}_k$, which is the same as $\pi_k(x) \geq R$. The researcher gains 1 at acceptance. The evaluator gains $\pi_k(x) - R$.

While payoffs depend on all exogenous parameters, here we have chosen to highlight their dependence on R. We will analyze the impact of selection on both the evaluator's and the researcher's welfare, as R varies between θ_L and θ_H (and hence as the relative weight attached to type I errors, the ratio $(1-p)(R-\theta_L)/[p(\theta_H - R)]$, varies between 0 and ∞).

Proposition 3. *Assume* $-\log(-\log(F))$ *is convex.* $V^k(R)$ *is increasing in k for every R. Moreover, for every pair* $k' > k \geq 1$ *there exist* $R', R'' \in (\theta_L, \theta_H)$ *such that* $U^{k'}(R) \leq U^k(R)$ *for* $R \in [\theta_L, R']$ *and* $U^{k'}(R) \geq U^k(R)$ *for* $R \in [R'', \theta_H]$.

In many special cases of interest, for $k' \geq k$ the c.d.f.s of $\pi_{k'}$ and π_k cross only once—other than at the extremes $R = \theta_L$ and $R = \theta_H$—and hence the thresholds R' and R'' in Proposition 3 coincide. When our convexity condition holds, the crossing occurs from above. The evaluator benefits from a larger k for every R, whereas the researcher benefits or loses according to whether R is above or below a threshold. This happens, for instance, in the case of a normally

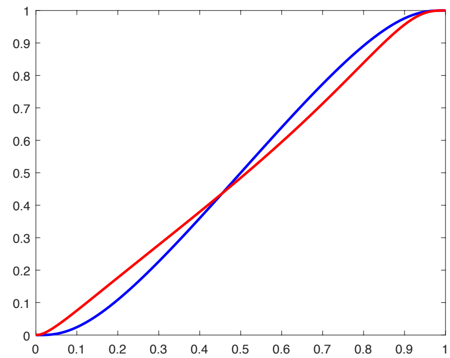

Figure 4: Welfare impact of selection in the normal case, illustrated by the posterior expectation's c.d.f.s $\tilde{F}_1$ (blue) and $\tilde{F}_2$ (red).

distributed error (cf. Figure 1), as illustrated in Figure 4 for $k = 1$ (blue) and $k = 2$ (red), with $\theta_L = 0$ and $\theta_H = 1$.

When greater selection is not uniformly better for the evaluator, the c.d.f.s of the posterior expectation corresponding to k and $k' > k$ must cross more than once, and the effect on the researcher is more subtle. Consider again the case of a Laplace distribution. Recall that, in this case, for the evaluator F^2 is worse than F for large R but better for small R (cf. Figure 3). A consequence of this is that the impact on the researcher's welfare changes sign twice. The researcher benefits from selection for small or large values of R, but loses for intermediate values. The left-hand side panel of Figure 5 plots the cumulative distributions of the posterior expectations π^2 (red) and π^1 (blue), illustrating this.

Next, recall the case of the uniform distribution discussed earlier, where the evaluator benefits from selection for large R, but fares worse for small R (cf. Figure 2) Similarly to the case of Laplace distribution, here the impact on the researcher's welfare also changes sign twice as R varies between θ_L and θ_H. However, the impact on the researcher's welfare, just like that on the evaluator's, is exactly the opposite. Selection benefits the researcher for small or large values of R, but harms for intermediate values, as illustrated in the right-hand side panel of Figure 5.

Finally, note that in the Gumbel case, the researcher's selection has no impact on the error

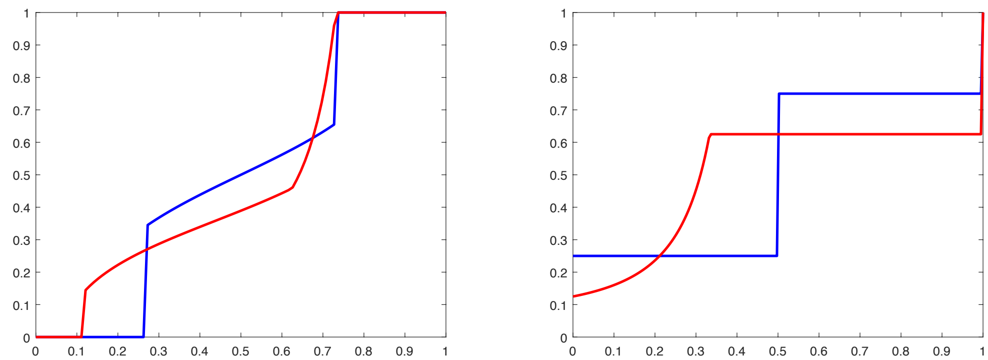

Figure 5: Comparison between $\tilde{F}_1$ (blue) and $\tilde{F}_2$ (red) for the Laplace distribution (left panel) and uniform distribution (right panel).

rates (α, β). In this case, properly anticipated selection has zero impact on the decision and thus on the information value (evaluator's payoff) and the acceptance probability (researcher's decision payoff).

Production of Data. Finally, it is natural to endogenize the number k of subjects among which the researcher can select. Suppose that the researcher can secretly choose this number ex ante, thus fixing the research procedure before errors $(\varepsilon_1, ..., \varepsilon_k)$ are drawn by nature. Suppose also that there is no credible way to directly reveal any information about the true k. The researcher bears costs $C(k)$ which we assume to be an increasing, convex function — if we restrict attention to natural numbers k, convexity means that $C(k+1) - C(k)$ is increasing in k.

We look for a pure strategy equilibrium, where the evaluator correctly anticipates the k that is optimally chosen by the researcher. In the subgame, the evaluator takes k as given and best responds with the acceptance threshold $\hat{x}$ defined by (4). In the first stage, the researcher correctly anticipates threshold $\hat{x}$, and chooses k in order to maximize the payoff

$$p\left(1 - F^k(\hat{x} - \theta_H)\right) + (1-p)\left(1 - F^k(\hat{x} - \theta_L)\right) - C(k). \tag{5}$$

Lemma 3. *The researcher's objective function is concave in k.*

Proof. We have already assumed that the cost function is convex, so it suffices to check that the first two terms are concave in k. It suffices to take k as a real number. The first derivative of a^k is $\log(a)\,a^k$ and the second derivative is $(\log(a))^2\,a^k$ which is positive when $a \notin \{0,1\}$. It is easy to see that the first terms are instead constant in k, if the base is zero or one. $\square$

Concavity implies that the best response is defined by one (or both) of the integers closest to solving the analytical first order condition

$$-p\log\left(F\left(\hat{x}-\theta_H\right)\right)F^k\left(\hat{x}-\theta_H\right)-(1-p)\log\left(F\left(\hat{x}-\theta_L\right)\right)F^k\left(\hat{x}-\theta_L\right)=C'(k)\,. \tag{6}$$

Considering a deviation from equilibrium, the researcher actually has a potential gain through the upward shift of the realized observation x. This is to be weighed against the cost of looking at more subjects, when already looking at k.

Lemma 4. *The equilibrium is characterized through the solution $(\hat{x},k)$ to the equation pair (4) and (6).*

Our main purpose is to remark three interesting properties of this equilibrium. We do not need to fully solve the model for this purpose.

First, keeping all other parts of the model fixed, for every $k \in \mathbb{N}$ there exists an increasing and convex cost function C such that k is the equilibrium choice of the researcher. Simply, fix k and solve (as before) equation (4) for the evaluator's best response $\hat{x}_k$. Then plug in $\hat{x}_k$ and k on the left hand side of (6) to determine the requisite $C'(k)$. Then choose the number $\gamma > 0$ such that $C'(k) = 2\gamma k$, and use the quadratic cost function $C(k) = \gamma k^2$. This observation provides a foundation for our approach so far where the number k was taken for given.

Second, and most importantly, this model exhibits a rat race effect: when the evaluator correctly anticipates a greater degree k of selection, the researcher's cost $C(k)$ to manipulate the experiment is largely wasted.[19] To see the cleanest instance of this, consider the Gumbel example. We have established above that the evaluator's and researcher's gross payoffs are independent of k. The researcher's total payoff, accounting for costs $C(k)$, is then smaller when these costs are greater. As follows from our first remark, any k is consistent with equilibrium.

[19] In general, rat race effects are described by Akerlof (1976) and Holmström (1999).

This does not imply that costs can be arbitrarily large, but it does prove that the costs can be positive. It also proves that the researcher may gain from tying the hands to be unable to augment k. The researcher would also gain from being able to credibly prove the chosen k to the evaluator. Going beyond the Gumbel example, these costs of manipulation could further harm a researcher who was already harmed by the evaluator's response. If the researcher stood to gain from manipulation in Proposition 3, the endogenous costs will reduce this gain, perhaps to a loss.

Third, note from (6) that the evaluator's best response k may increase or decrease with the evaluator's standard $\hat{x}$, depending on parameters. The sign of this slope depends on the sign of the derivative of the left hand side in (6) with respect to $\hat{x}$,

$$
\begin{aligned}
& p\left(1+k\log\left(F\left(\hat{x}-\theta_H\right)\right)\right)F^{k-1}\left(\hat{x}-\theta_H\right)f\left(\hat{x}-\theta_H\right) \\
& +(1-p)\left(1+k\log\left(F\left(\hat{x}-\theta_L\right)\right)\right)F^{k-1}\left(\hat{x}-\theta_L\right)f\left(\hat{x}-\theta_L\right),
\end{aligned}
$$

which is negative when $F\left(\hat{x}-\theta_L\right)$ is sufficiently small, as happens when the prior strongly favors acceptance. In that case, the best response k is a decreasing function of $\hat{x}$. Conversely, when the prior strongly favors rejection.

7 Impact of Unanticipated and Uncertain Selection

So far we consider situations in which the evaluator rationally predicts the correct k and there is no uncertainty in the extent of selection, for example because the parameters of the model (such as researcher's bias and cost of presampling) are known. This is the most optimistic scenario when evaluating the impact of selection. We now relax these assumptions to consider more realistic scenarios.

Impact on Unwary Evaluator. Consider an unwary evaluator who wrongly anticipates a lower k than true. Holding fixed the true k, clearly the evaluator is generally worse off by being unwary than being rational. More interestingly, it is ambiguous whether an unwary evaluator who expects k gains or loses when the true signal has $k' > k$. If the rational evaluator would prefer k' to k, this gain might be greater than the cost of irrationality.

In an important benchmark case, we find that the unwary evaluator is exactly indifferent to an increase of selection from $k=1$ to $k=2$. Consider a situation of equipoise, whereby at the prior the evaluator is indifferent between acceptance and rejection, $(1-p)(R-\theta_L) = p(\theta_H - R)$.[20] Suppose that the noise distribution is symmetric, so that for some ε_0 we have $F(\varepsilon_0+\varepsilon) = 1-F(\varepsilon_0-\varepsilon)$ for all ε. Start from the approval standard $\bar{x} = \varepsilon_0 + (\theta_L+\theta_H)/2$ optimal with no selection, $k=1$, and consider how an increase in selection to $k=2$ affect an unwary evaluator who does maintains the approval standard unchanged at $\bar{x}$. The probability of acceptance then clearly increases, resulting in a change in the evaluator's payoff equal to

$$-(1-p)\underbrace{\left[F(\bar{x}-\theta_L)-F^2(\bar{x}-\theta_L)\right]}_{\text{increase in FALSE POSITIVES}}(R-\theta_L)+p\underbrace{\left[F(\bar{x}-\theta_H)-F^2(\bar{x}-\theta_H)\right]}_{\text{reduction in FALSE NEGATIVES}}(\theta_H-R).$$

The two effects exactly cancel given that $F(\bar{x}-\beta_L) = 1-F(\bar{x}-\beta_H)$ for $\bar{x} = \varepsilon_0 + (\theta_L+\theta_H)/2$ by symmetry of F and equipoise, $(1-p)(R-\theta_L) = p(\theta_H-R)$:

Proposition 4. *Suppose that F is symmetric, $F(\varepsilon_0+\varepsilon) = 1-F(\varepsilon_0-\varepsilon)$, and that the evaluator attaches equal weight to type I and type II errors, $(1-p)(R-\theta_L) = p(\theta_H-R)$. Then the unwary evaluator who anticipates no selection, $k=1$, is indifferent to whether the true selection is $k=1$ or $k=2$.*

Thus under symmetry and equipoise an increase in selection from $k=1$ to $k=2$ necessarily benefits a rational evaluator.

Impact of Uncertain Selection. In a natural extension of the selection model, the number k is random. Again, the $\beta(\alpha)$ curve can be generally used to compare experiments. However, the application of dispersion is harder. Suppose that ε is drawn from $\lambda F^{k+1} + (1-\lambda)F^k$ where $\lambda \in (0,1)$ and F^{k+1} is less dispersed than F^k. It might be natural to conjecture that the evaluator is better off, the greater the weight λ attached to the less dispersed experiment. However, this is generally false. To see this note that when F is Gumbel, both F^k and F^{k+1} are Gumbel, but

[20] The condition of equipoise, requiring experimental subjects to be indifferent between treatment and control, is an ethical prerequisite for carrying out a randomized experiment.

$\lambda F^{k+1}+(1-\lambda)F^k$ is not Gumbel. In fact, for every $\lambda \in (0,1)$, $\lambda F^{k+1}+(1-\lambda)F^k$ is worse than F^k, for it is Blackwell worse than informing the evaluator about the outcome of the lottery over F^k and F^{k+1}. Intuitively, the equivalence of F^k with F^{k+1} rests on being able to remove a constant bias from the distribution of ε, but this is not feasible when it is random whether ε derives from one distribution or the other.[21]

8 Contribution to Literature

To the literature on optimal persuasion following Johnson and Myatt (2006), Rayo and Segal (2010) and Kamenica and Gentzkow (2011) we contribute a signal-jamming model of persuasion.[22] The researcher's choice of the size k of the presample is akin to the agent's effort choice in Holmström's (1999) classic career concern model. The twist here is that this effort results in private information, which the researcher then uses to select the reported information. As we show, information manipulation induces positive skewness in the distribution of treated outcomes. Contrary to naive intuition, the evaluator is not necessarily hurt by information manipulation; actually, we characterize natural conditions under which the evaluator benefits. In addition, we characterize situations in which the researcher ends up suffering from information manipulation like in a rat race, even if we abstract away from the cost of acquiring information.

In a pioneering game theoretic analysis, Blackwell and Hodges (1957) analyze how an evaluator should optimally design a sequential experiment to minimize selection bias, defined as the number of times an optimizing researcher is able to correctly forecast the treatment assignment.[23] In the context of our single experiment, we characterize situations in which the

[21] Some advance may still be feasible by using the arithmetic-mean to geometric-mean inequality, and observing that $F^{\lambda(k+1)}F^{(1-\lambda)k}$ is itself a power of F, as characterized in Theorem 1.

[22] See also Henry (2009), Dahm, González, and Porteiro (2009), and Felgenhauer and Schulte (2014) for persuasion models with endogenous information acquisition. In our setting, the researcher is constrained to disclose a single observation, as in the limited-attention models proposed by Fishman and Hagerty (1990) and Hoffmann, Inderst, and Ottaviani (2014).

[23] Blackwell and Hodges (1957) argue that selection bias is minimized by a truncated binomial design, according to which the initial allocations to treatment and control are selected independently with a fair coin, until half of the

selection bias that is present when the researcher is able to forecast the assignment actually benefits the evaluator, contrary to what Blackwell and Hodges (1957) stipulate.

In a complementary approach to modeling conflicts of interest in statistical testing, Tetenov (2016) analyzes a regulator's optimal commitment to a decision rule when privately informed proponents of innovations select into costly testing. Instead, we focus on the impact of a researcher's manipulation of the data on the welfare of an uncommitted evaluator. Henry and Ottaviani (2015) analyze a dynamic model of persuasion with costly information acquisition à la Wald (1950), where information is truthfully reported at the time of application.

While in general selected data are not Blackwell comparable to random data, we characterized the welfare impact of selection on the basis of dispersion. Our results depend on the features of the environment (conditional signal distribution and parameters of decision problem). Our notion of local dispersion, like Lehmann's, applies to the case where the evaluator observes only one outcome x. It is natural to generalize. The construction of the $\beta(\alpha)$ curve is generally applicable, and the convex envelope generally allows for the comparison of experiments, as explained in Section 4.

We leave to future work the design of experiments and policy responses in the presence of strategic selection. A natural starting point in this direction is Chassang, Padró i Miquel, and Snowberg's (2012) characterization of experimental design when outcomes are affected by experimental subjects' unobserved actions.

---

subjects are allocated to either treatment or control; from that point on, allocation is deterministic. Efron (1971), instead, characterizes the selection bias resulting from a biased coin design, according to which the probability of current assignment to treatment is higher if previous randomizations resulted in excess balance of controls over treatments.

A Appendix: Proofs

A.1 Proof of Theorem 3

Fix any $\delta \in (0, (\theta_H - \theta_L)/2)$. Let $\varepsilon_\delta > 0$ be such that $\bar{F}(\varepsilon_\delta) - \bar{F}(-\varepsilon_\delta) \geq 1 - \delta/2$. Choose $\hat{k}$ so that for all $k \geq \hat{k}$,

$$a_k \varepsilon_\delta < \delta, \qquad F^k(b_k + a_k \varepsilon_\delta) \geq \bar{F}(\varepsilon_\delta) - \frac{\delta}{4}, \qquad \text{and} \qquad F^k(b_k - a_k \varepsilon_\delta) \leq \bar{F}(-\varepsilon_\delta) + \frac{\delta}{4}.$$

Then, for each θ, since $x = \theta + b_k + a_k \varepsilon$,

$$\begin{aligned}
\Pr\left(\theta + b_k - \delta \leq x \leq \theta + b_k + \delta \,\middle|\, \theta\right) &\geq \Pr\left(\theta + b_k - a_k \varepsilon_\delta \leq x \leq \theta + b_k + a_k \varepsilon_\delta \,\middle|\, \theta\right) \\
&= F^k(b_k + a_k \varepsilon_\delta) - F^k(b_k - a_k \varepsilon_\delta) \\
&\geq \bar{F}(\varepsilon_\delta) - \frac{\delta}{4} - \bar{F}(-\varepsilon_\delta) - \frac{\delta}{4} \\
&\geq 1 - \delta.
\end{aligned}$$

In words, the distribution of observation x in state θ assigns at least probability $1 - \delta$ to a δ-ball around the point $\theta + b_k$. Now, rejecting if and only if $x < \hat{x} = \theta_H + b_k - \delta$ gives

$$\alpha = \Pr\left(x \geq \theta_H + b_k - \delta \,\middle|\, \theta_L\right) \leq 1 - \Pr\left(\theta_L + b_k - \delta \leq x \leq \theta_L + b_k + \delta \,\middle|\, \theta_L\right) \leq \delta$$

and

$$\beta = \Pr\left(x < \theta_H + b_k - \delta \,\middle|\, \theta_H\right) \leq 1 - \Pr\left(\theta_H + b_k - \delta \leq x \leq \theta_H + b_k + \delta \,\middle|\, \theta_H\right) \leq \delta.$$

As we can choose $\delta > 0$ arbitrarily small, we can make $\bar{\ell}\alpha + \beta$ arbitrarily small. The first claim in the theorem follows.

To prove the second claim, consider the pair of error rates (α, β) that result with threshold $\hat{x}$ when the error is drawn from $\bar{F}$. Let (α_k, β_k) be the error rates that result with threshold $\hat{x} - b_k$ when the error is draw from F^k. The convergence $F^k(b_k + \varepsilon) \to \bar{F}(\varepsilon)$ at both $\varepsilon = \hat{x} - \theta_H$ and $\hat{x} - \theta_L$ implies that the sequence (α_k, β_k) converges to (α, β). Since this is true for every $\hat{x}$, every point on the $\beta(\alpha)$ curve generated from $\bar{F}$ is a limit point for the corresponding curves for experiments F^k. For the convex curves in the compact space, this implies convergence of the $\beta(\alpha)$ functions. This implies convergence of the evaluator's payoff.

A.2 Extreme Selection in the Exponential Power Case

In this appendix we show that if $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. exponential power with shape b, location 0 and scale 1, then $M_n = \max\{\varepsilon_1, \ldots, \varepsilon_n\}$ satisfies

$$\Pr(M_n \leq a_n \varepsilon + b_n) \to e^{-e^{-\varepsilon}} \quad \forall \varepsilon \qquad \text{(Gumbel)},$$

where

$$a_n = (b \log n)^{-\frac{b-1}{b}}$$

and

$$b_n = (b \log n)^{1/b} - \frac{\frac{b-1}{b} \log\log n + \log(2\Gamma[1/b])}{(b \log n)^{\frac{b-1}{b}}}.$$

Here, Γ denotes the Gamma function.

Remark 1. *For $b = 1$ (Laplace) we have $a_n = 1$ and*

$$b_n = \log n - \log(2\Gamma[1]) = \log n - \log 2.$$

Remark 2. *For $b = 2$ (normal) we have $a_n = (2 \log n)^{-1/2}$ and*

$$\begin{aligned} b_n &= (2\log n)^{1/2} - \frac{\frac{1}{2}\log\log n + \log(2\Gamma[1/2])}{(2\log n)^{1/2}} \\ &= (2\log n)^{1/2} - \frac{\log\log n + 2\log\left(2\sqrt{\pi}\right)}{2(2\log n)^{1/2}} \\ &= (2\log n)^{1/2} - \frac{\log\log n + \log 4\pi}{2(2\log n)^{1/2}}, \end{aligned}$$

as in Leadbetter et al. (Note that a_n in that book corresponds to $1/a_n$ here.)

Remark 3. *The calculations do not cover the case $b = \infty$ (uniform) because in this case the relevant extreme value distribution is not Type I (Gumbel) but rather Type III (Weibull).*

The proof follows closely and generalizes the one given by Leadbetter et al. (1983) for the normal case $b = 2$.

Start by noticing that

$$\frac{f(\varepsilon)}{\varepsilon^{b-1}[1 - F(\varepsilon)]} \to 1 \quad \text{as } \varepsilon \to \infty.$$

Fix ε and define y_n for each $n \geq 1$ by

$$1 - F(y_n) = \frac{e^{-\varepsilon}}{n},$$

so that

$$\frac{e^{-\varepsilon}}{n} \frac{y_n^{b-1}}{f(y_n)} \to 1 \quad \text{as } n \to \infty. \tag{7}$$

We may assume $y_n > 0$ for all n. Then

$$f(y_n) = \frac{b^{\frac{b-1}{b}}}{2\Gamma[1/b]} e^{-y_n^b/b},$$

and hence, by (7),

$$-\log n - \varepsilon + (b-1)\log y_n - \frac{b-1}{b}\log b + \log(2\Gamma[1/b]) + \frac{y_n^b}{b} \to 0. \tag{8}$$

From (8) we see that

$$-\log n + (b-1)\log y_n + \frac{y_n^b}{b} = -\log n + o\left(y_n^b/b\right) + \frac{y_n^b}{b} \to \text{a constant},$$

and hence that

$$-\frac{b\log n}{u_n^b} + \frac{o\left(y_n^b/b\right)}{u_n^b/b} + 1 \to 0,$$

that is

$$\frac{b\log n}{u_n^b} \to 1,$$

or

$$b\log y_n - \log b - \log\log n \to 0,$$

that is

$$\log y_n = \frac{1}{b}(\log b + \log\log n) + o(1).$$

Using this fact in (8), we obtain

$$\begin{aligned}
\frac{y_n^b}{b} &= \log n + \varepsilon - \frac{b-1}{b}(\log b + \log\log n) + \frac{b-1}{b}\log b - \log(2\Gamma[1/b]) + o(1) \\
&= \log n + \varepsilon - \frac{b-1}{b}\log\log n - \log(2\Gamma[1/b]) + o(1),
\end{aligned}$$

or

$$y_n^b = b\log n\left[1 + \frac{\varepsilon - \frac{b-1}{b}\log\log n - \log(2\Gamma[1/b])}{\log n} + o\left(\frac{1}{\log n}\right)\right],$$

or

$$
\begin{aligned}
y_n &= (b\log n)^{1/b}\left[1+\frac{\varepsilon-\frac{b-1}{b}\log\log n-\log\left(2\Gamma[1/b]\right)}{\log n}+o\left(\frac{1}{\log n}\right)\right]^{1/b} \\
&= (b\log n)^{1/b}\left[1+\frac{\varepsilon-\frac{b-1}{b}\log\log n-\log\left(2\Gamma[1/b]\right)}{b\log n}+o\left(\frac{1}{\log n}\right)\right] \\
&= (b\log n)^{1/b}+\frac{\varepsilon-\frac{b-1}{b}\log\log n-\log\left(2\Gamma[1/b]\right)}{(b\log n)^{\frac{b-1}{b}}}+o\left(\frac{1}{(\log n)^{\frac{b-1}{b}}}\right) \\
&= a_n\varepsilon+b_n+o\left(a_n\right).
\end{aligned}
$$

Thus

$$
\Pr\left(M_n\le a_n\varepsilon+b_n+o\left(a_n\right)\right)\to e^{-e^{-\varepsilon}}.
$$

A.3 Proof of Proposition 3

It follows directly from Theorem 1 that $V^k(R)$ is increasing in k. Now, $V^{k'}(R)-V^k(R)=\int_R^{\theta_H}[\tilde{F}_k(r)-\tilde{F}_{k'}(r)]dr$. Hence, we must have $\tilde{F}_{k'}(R)\le\tilde{F}_k(R)$ for all R sufficiently close to θ_H. By definition of $U^k(R)$, then $U^{k'}(R)\ge U^k(R)$. A similar argument works at the other end of the distribution. To see this, first recall that the expectation of π_k is the same for every k, because it must equal the prior expectation of θ. In other words, for every k we have

$$
\int_{\theta_L}^{\theta_H}[1-\tilde{F}_k(r)]dr=p\theta_H+(1-p)\theta_L.
$$

Hence, $V^{k'}(R)-V^k(R)=\int_{\theta_L}^{R}[\tilde{F}_{k'}(r)-\tilde{F}_k(r)]dr$. It follows that $\tilde{F}_{k'}(r)\ge\tilde{F}_k(r)$ and thus $U^{k'}(R)\le U^k(R)$ for R sufficiently close to θ_L.

A.4 Proof of Proposition 4

The two assumptions imply that the unwary evaluator uses a symmetric threshold, $\bar{x}=\varepsilon_0+(\theta_L+\theta_H)/2$. The likelihood ratio of this threshold signal is

$$
\frac{f\left(\bar{x}-\theta_H\right)}{f\left(\bar{x}-\theta_L\right)}=\frac{f\left(\varepsilon_0+(\theta_L-\theta_H)/2\right)}{f\left(\varepsilon_0-(\theta_L-\theta_H)/2\right)}=1,
$$

and $F(\bar{x}-\theta_L)+F(\bar{x}-\theta_H)=1$ is satisfied. Now, the unwary's utility gain due to selection is

$$(1-p)\left[F^2(\bar{x}-\theta_L)-F(\bar{x}-\theta_L)\right](R-\theta_L)+p\left[F^2(\bar{x}-\theta_H)-F(\bar{x}-\theta_H)\right](R-\theta_H).$$

By assumption, $(1-p)(R-\theta_L)=p(\theta_H-R)$, so the utility gain reduces to

$$(1-p)(R-\theta_L)\left[F^2(\bar{x}-\theta_L)-F(\bar{x}-\theta_L)+F(\bar{x}-\theta_H)-F^2(\bar{x}-\theta_H)\right].$$

This is zero since $F(\bar{x}-\theta_L)+F(\bar{x}-\theta_H)=1$.

B Appendix: Local Comparison of Experiments with Multiple Crossings

In this appendix we complement the characterization provided in Theorem 2 by dealing with comparisons between experiments based on distributions that may not have the same support (up to a location shift), or whose quantile difference may exhibit multiple increasing or decreasing parts. As we explain shortly, dispensing with the requirements in our earlier theorem makes it necessary to meet another requirement, namely that the difference $\theta_H - \theta_L$ be not too large.

Let us illustrate why this additional requirement is needed, when the conditions in Theorem 2 are not met. To see what happens when two experiments F and G do not have the same support, suppose F is the uniform distribution on $[0,1]$ and G is a piecewise-uniform distribution on $[0,5/4]$, say $G(\varepsilon) = \varepsilon/2$ for $\varepsilon \in [0,1]$ and $G(\varepsilon) = 1/2 + 2(\varepsilon - 1)$ for $\varepsilon \in [1,5/4]$. Clearly, the quantile difference $G^{-1}(u) - F^{-1}(u)$ is strictly increasing for $u < 1/2$ and strictly decreasing for $u > 1/2$. However, the information constraint of F is strictly below that of G if $\theta_H - \theta_L$ is sufficiently large. Indeed, $\beta_F(\alpha) = 1 - \alpha - (\theta_H - \theta_L)$ while $\beta_G(\alpha) = 5/8 - (\theta_H - \theta_L)/2 - \alpha/4$. Thus, $\beta_G(\alpha) > \beta_F(\alpha)$ for all α, provided that $\theta_H > \theta_L + 3/4$. Intuitively, the difference between θ_H and θ_L is too large relative to the decreasing part of the quantile difference, which is never relevant for the comparison between F and G. Similar examples can be constructed, where the other requirement in Theorem 2, a single change of sign in the slope of the quantile difference, is violated.

Theorem 4. *Let $N \geq 1$ and $0 = u_1 \leq \cdots \leq u_{2N+1} = 1$. Suppose that, for all $n = 1, \ldots, N$,*

$$f\big(F^{-1}(u)\big) \leq g\big(G^{-1}(u)\big) \qquad \text{for all } u \in (u_{2n-1}, u_{2n}),$$
$$f\big(F^{-1}(u)\big) \geq g\big(G^{-1}(u)\big) \qquad \text{for all } u \in (u_{2n}, u_{2n+1}),$$

Suppose also that, for all $m = 1, \ldots, 2N$,

$$\theta_H - \theta_L \leq \max\Big\{F^{-1}(u_{m+1}) - F^{-1}(u_m)\ ,\ G^{-1}(u_{m+1}) - G^{-1}(u_m)\Big\}.$$

Then there exist $0 = K_1 \leq \cdots \leq K_{2N+1} = \infty$ *such that, for all* $n = 1, \ldots N$,

$$\begin{aligned} &G \textit{ is preferred to } F \quad \textit{if} \quad \frac{1-p}{p}\frac{R-\theta_L}{\theta_H - R} \in \left(K_{2n-1}, K_{2n}\right), \\ &F \textit{ is preferred to } G \quad \textit{if} \quad \frac{1-p}{p}\frac{R-\theta_L}{\theta_H - R} \in \left(K_{2n}, K_{2n+1}\right). \end{aligned}$$

Proof. It suffices to show that there exist $0 = \alpha_{2N+1} \leq \cdots \leq \alpha_1 = 1$ such that, for all $n = 1, \ldots, N$, we have $\beta_G(\alpha) \leq \beta_F(\alpha)$ for all $\alpha \in (\alpha_{2n}, \alpha_{2n-1})$ and $\beta_G(\alpha) \geq \beta_F(\alpha)$ for all $\alpha \in (\alpha_{2n+1}, \alpha_{2n})$.

Define, for every $v \in (0,1)$,

$$u(v) = \max\left\{\beta_F(1-v)\,,\, \beta_G(1-v)\right\}.$$

The bound on the difference $\theta_H - \theta_L$ implies that there exist $0 = v_1 \leq \cdots \leq v_{2N+1} = 1$ such that, for all $n = 1, \ldots, N$,

$$\begin{aligned} G^{-1}(v) - F^{-1}(v) &\leq G^{-1}(u(v)) - F^{-1}(u(v)) \qquad \text{for all } v \in (v_{2n-1}, v_{2n}), \\ G^{-1}(v) - F^{-1}(v) &\geq G^{-1}(u(v)) - F^{-1}(u(v)) \qquad \text{for all } v \in (v_{2n}, v_{2n+1}). \end{aligned}$$

But for each $\alpha \in (0,1)$ we have

$$G^{-1}(1-\alpha) - F^{-1}(1-\alpha) = G^{-1}(\beta_G(\alpha)) - F^{-1}(\beta_F(\alpha))$$

and hence the inequality $\beta_G(\alpha) \leq \beta_F(\alpha)$ is equivalent to both of the following:

$$\begin{aligned} G^{-1}(1-\alpha) - F^{-1}(1-\alpha) &\leq G^{-1}(\beta_F(\alpha)) - F^{-1}(\beta_F(\alpha)), \\ G^{-1}(1-\alpha) - F^{-1}(1-\alpha) &\leq G^{-1}(\beta_G(\alpha)) - F^{-1}(\beta_G(\alpha)). \end{aligned}$$

We conclude that, for all $n = 1, \ldots, N$,

$$\begin{aligned} \beta_G(\alpha) &\leq \beta_F(\alpha) \qquad \text{for all } \alpha \in (1 - v_{2n}, 1 - v_{2n-1}) =: (\alpha_{2n}, \alpha_{2n-1}), \\ \beta_G(\alpha) &\geq \beta_F(\alpha) \qquad \text{for all } \alpha \in (1 - v_{2n+1}, 1 - v_{2n}) =: (\alpha_{2n+1}, \alpha_{2n}). \end{aligned}$$

□

References

Akerlof, George (1976), "The Economics of Caste and of the Rat Race and Other Woeful Tales," *Quarterly Journal of Economics*, 90(4), 599–617.

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

Berger, Vance W. (2005), *Selection Bias and Covariate Imbalances in Randomized Clinical Trials*, Wiley.

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

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

Brown, George W. and John W. Tukey (1946), "Some Distributions of Sample Means," *Annals of Mathematical Statistics*, 17(1), 1–12.

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

Dahm, Matthias, Paula González, and Nicolás Porteiro (2009), "Trials, Tricks and Transparency: How Disclosure Rules Affect Clinical Knowledge," *Journal of Health Economics*, 28(6), 1141–1153.

Di Tillio, Alfredo, Marco Ottaviani, and Peter N. Sørensen (2015), "Persuasion Bias in Science: Can Economics Help?" mimeo.

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

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

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

Head, Megan L., Luke Holman, Rob Lanfear, Andrew T. Kahn, and Michael D. Jennions (2015), "The Extent and Consequences of P-hacking in Science," *PLoS Biology*, 13(3), 1–15.

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

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

Henry, Emeric and Marco Ottaviani (2015), "Research and the Approval Process: The Organization of Persuasion," Bocconi mimeo.

Hoffmann, Florian, Roman Inderst, and Marco Ottaviani (2014), "Persuasion through Selective Disclosure: Implications for Marketing, Campaigning, and Privacy Regulation," Bocconi mimeo.

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

Jewitt, Ian (2007), "Information Order in Decision and Agency Problems," Nuffield College mimeo.

Johnson, Justin P., and David P. Myatt (2006), "On the simple economics of advertising, marketing, and product design," *American Economic Review*, 96(3), 756–784.

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

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

Khaledi, Baha-Eldin and Subhash Kochar (2000), "On Dispersive Ordering between Order Statistics in One-Sample and Two-Sample Problems," *Statistics and Probability Letters*, 46(3), 257–261.

Leadbetter, M. R., Georg Lindgren, and Holger Rootzén (1983), *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York.

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

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

Schulz, Kenneth F. (1995), "Subverting Randomization in Controlled Trials," *Journal of American Medical Association*, 274(18), 1456–1458.

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

Simonsohn, Uri, Leif D. Nelson, and Joseph P. Simmons (2014), "*P*-curve: A Key to the File Drawer," *Journal of Experimental Psychology: General*, 143(2), 534–547.

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

Torgersen, Erik (1991), *Comparison of Statistical Experiments*, Cambridge University Press.

Wald, Abraham (1950), *Statistical Decision Function*, Wiley.