Strategic Sample Selection*

Alfredo Di Tillio† Marco Ottaviani‡ Peter Norman Sørensen§

September 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? We show that anticipated selection locally reduces noise dispersion and thus increases informativeness if and only if the noise distribution is double logconvex, as with normal noise. The results are applied to the analysis of strategic sample selection by a biased researcher and extended to the case of uncertain and unanticipated selection. Our theoretical analysis offers applied research a new angle on the problem of selection in empirical studies, by characterizing when selective assignment based on untreated outcomes benefits or hurts the evaluator.

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

JEL codes: D82, D83, C72, C90

*Research funded by the European Research Council through Advanced Grant 29583 (EVALIDEA). We thank Matteo Camboni for outstanding research assistance. We also thank, without implication, Pierpaolo Battigalli, Francesco Corielli, Nicola Limodio, and seminar participants at Bocconi, Bruxelles, Caltech, Carlos III, Como, East Anglia, Gerzensee, Helsinki, Leicester, Lisbon, London Business School, Milan, Nottingham, Paris, Rome, 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

Empirical data are often nonrandomly selected, due to choices made by the subjects under investigation or sample inclusion decisions by data analysts.[1] Suppose a new treatment is given to the healthiest patient rather than to a random patient in a group. Because of selection, favorable outcomes are weaker evidence that the treatment is effective. But is the evaluator's assessment of the treatment effect more accurate with selective or with random assignment? A similar question arises when peremptory challenge gives a defendant the right to strike down a number of jurors. Given that the defendant selects the most favorable jurors, how is the quality of final judgement affected? When feeding a consumer review to potential buyers with limited attention, should an e-commerce platform post a random review or allow the merchant to cherry-pick one? And when testing a student in an exam, should the teacher ask a question at random or allow the student to select the most preferred question out of a batch?

These comparisons are all instances of one and the same problem—assessing the impact of selection in simple hypothesis testing. There is an unknown state θ that is either high or low—the effect of a new treatment, a defendant's innocence, the quality of a merchant's good, or a student's ability. An evaluator observes a stochastic signal about the state, $x = \theta + \varepsilon$, where ε is a noise term—the baseline health of an individual, a juror's bias, a reviewer's leniency, or a student's specific knowledge. Based on the signal, the evaluator then makes a decision either to accept—the correct choice in the high state—or reject—the correct choice in the low state. The issue is, in which of the following two scenarios does the evaluator make better decisions:

- **Random Experiment**. The signal is sampled randomly. The noise term ε is drawn from a known cumulative distribution F.
- **Selected Experiment**. The signal is selected—perhaps by another party, strategically—as the highest out of $k > 1$ random signals. Thus, $\varepsilon = \max\langle \varepsilon_1, \ldots, \varepsilon_k \rangle$, where $\varepsilon_1, \ldots, \varepsilon_k$ are independent draws from the same distribution F. The noise distribution of a selected experiment becomes $G = F^k$.

To address this issue we must compare the error rates the evaluator can achieve in the two experiments. Given a noise distribution, F or G, the evaluator's decision is the familiar trade-off between the probability α of a false positive—accepting in the low state—and the probability β of a false negative—rejecting in the high state. With a higher threshold of acceptance on the observed signal, α decreases, but β increases. The evaluator minimizes a weighted sum of the two, setting a higher threshold when the relative weight attached to α is higher. Consider the example of a uniform F, illustrated in Figure 1. The blue and red curves on the left are the distributions of the

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

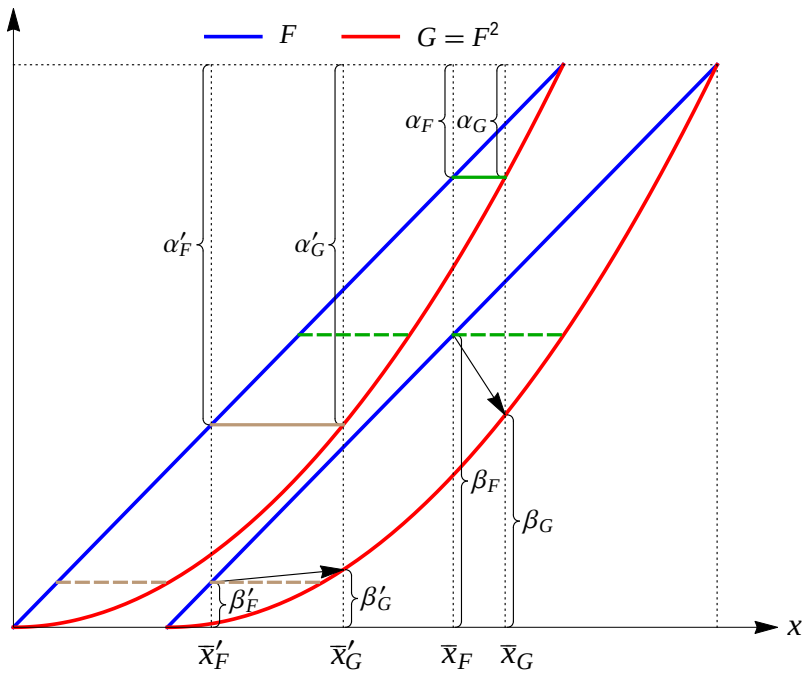


Figure 1: Local dispersion with a uniform noise.

random and selected ($k = 2$) signal in the low state. Those on the right, the distributions in the high state. An evaluator choosing the high threshold $\bar{x}_F$ in experiment F induces error rates α_F and β_F and is therefore at least as concerned about false positives as one choosing $\bar{x}'_F$, inducing error rates $\alpha'_F > \alpha_F$ and $\beta'_F < \beta_F$.

Consider now changing the experiment from F to $G = F^2$. An evaluator who was optimally choosing threshold $\bar{x}_F$ in experiment F stands to gain from the change: choosing threshold $\bar{x}_G$ in G gives as many false positives, $\alpha_G = \alpha_F$, but fewer false negatives, $\beta_G < \beta_F$. This is because at $\bar{x}_F$ the horizontal difference between the selected and the random signal distribution is smaller in the low state than in the high state—the solid green segment is shorter than the dashed green segment on the right. But the signal distributions in the two states are just translated versions of the noise distributions F and G. Thus, the gain is due to the fact that G is *locally less dispersed* than F at the top—the dashed green segment on the left is shorter than the solid green segment.

For an evaluator more concerned about false negatives, who was choosing a lower threshold like $\bar{x}'_F$ in experiment F, the mechanics of the change to G work just the opposite way, but the impact of the change is subtler. There are α'_F false positives in F, but getting as many ($\alpha'_G = \alpha'_F$) in G now requires increasing false negatives to $\beta'_G > \beta'_F$. At the bottom, G is locally more dispersed than F—the solid brown segment is longer than either dashed brown segment. From this observation, however, we cannot conclude that changing experiment from F to G damages the evaluator,

because the change leads to a new trade-off where α is reoptimized to a value different from α'_F.[2] The observation does reveal a structure though: plotting β against α for each experiment—we call this function the *information constraint* of the experiment—an alternating pattern emerges. The information constraint of G crosses that of F once and from below, as α increases beyond a certain value—reflecting the alternating pattern in local dispersion.

Our first main contribution is to develop a theory of local dispersion that can take advantage of this alternating pattern in the information constraint. By considering the convex envelope of the information constraints of F and G, in Theorem 1 we show that any alternating pattern of preference uniquely corresponds to an alternating sign pattern of the information constraint difference, and hence uniquely corresponds to a local dispersion pattern. Intuitively, every point where G switches from less to more locally dispersed than F corresponds to a critical weight attached to false positives—where the preferred experiment switches from G to F. For example, the uniform case discussed earlier represents the particular case where G is first more then less locally dispersed than F. This corresponds to the fact that as the weight attached to false positives increases beyond a critical value, the evaluator switches from preferring F to preferring G.

Our notion of local dispersion formalizes the idea that a noise distribution is more informative than another at some quantiles and less informative at others. Lehmann (1988) considered experiments ranked in the dispersive order—with one globally (everywhere locally) less dispersed than the other—and characterized when one experiment is preferred to the other for all possible evaluator's preferences, thus following Blackwell (1951, 1953). But, as the uniform example illustrates, in general the random experiment F and the selected experiment $G = F^k$ are incomparable in the sense of Blackwell—one cannot identify the experiment giving the evaluator the higher payoff without knowing anything about the evaluator's preferences. The notion of local dispersion strikes a different balance: it requires some knowledge of the evaluator's preferences—the strength of the evaluator's relative concern about false positives—but allows us to rank more experiments.

Given the characterization provided by Theorem 1, our main question about the welfare impact of selection becomes an issue of comparing random and selected experiments in terms of global or local dispersion. At a global level, our second main contribution identifies a necessary and sufficient condition for a selected experiment to be less dispersed than a random experiment. Theorem 2 shows that F^k is less dispersed than F, with dispersion monotonically decreasing in k, if and only if $-\log(-\log F)$ is a convex function. Gumbel's extreme value distribution has a double loglinear distribution function, so we can equivalently characterize double logconvexity of F as having a quantile density that is less elastic than Gumbel's. Given that less dispersion increases the evaluator's payoff, 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.

[2] Had we started from the assumption that $\bar{x}'_G$ were optimal for the selected experiment G, we could have concluded that the random experiment F is better than G.

Intuitively, the double logconvexity criterion says that neither the top tail of the distribution should be thicker than in the Gumbel distribution—for otherwise the thickening of the top tail resulting from selection would increase dispersion at the top of the distribution—nor the bottom tail should be thinner—for otherwise the thinning of the bottom tail created by selection would again increase dispersion, this time at the bottom. 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.

To assess the impact of selection in those cases where the double-log transformation of F is neither concave nor convex, we rely on both the local dispersion characterization in Theorem 1 and the double logconvexity criterion in Theorem 2. If, for instance, $-\log(-\log F)$ is first concave then convex, then selection harms evaluators with a strong preference against false negatives, as seen earlier with a uniform F. Overall, an evaluator more concerned about false negatives is harmed when the data distribution has sufficiently thin left tail (as for the uniform distribution). Symmetrically, an evaluator more concerned about false positives is harmed by selection when the right tail is sufficiently thick—this happens, for instance, with the Laplace distribution.

Our results have important implications for applied research. While typically thought of exclusively as a threat to internal validity, selective assignment based on untreated outcomes—or, more generally, some unobservable characteristics correlated with untreated outcomes—can actually benefit an evaluator who properly anticipates selection. Actually, as we discuss at the end of the paper, selection may benefit even an unwary evaluator who does not anticipate it. In addition, our analysis provides a useful practical criterion for a first assessment of the impact of selection in empirical studies. Since double logconvexity of F is equivalent to double logconvexity of F^k, a direct indication of the potential benefit or harm of selection can be obtained by checking if the empirical data distribution satisfies these properties—whether data are actually selected or not.

Drawing on extreme value theory, we also analyze the case with large selection, when k tends to infinity. 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 2) and information in the limit is less than full. We show that for all strictly logconcave distributions in the exponential power family the evaluator achieves full information in the limit with extreme selection. Thus, unless the noise distribution is exactly exponential in the upper tail (as for the Laplace distribution), extreme selection benefits the evaluator for all parameter values—however, the rate of convergence is extremely slow.

Selection naturally arises when the evidence is provided by a strategic researcher who observes a presample $x_1, ..., x_k$ of size k and then reports the most favorable realization $\max\langle x_1, ..., x_k\rangle$ to the evaluator. Given the presample size k and the fact that the evaluator uses an acceptance threshold

policy, it is indeed optimal for the researcher to report the highest realization. We provide a strategic foundation for sample selection by introducing a researcher who aims at convincing the evaluator that the true state is high—e.g. that a new 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 2, we show that equilibrium selective sampling benefits also the researcher in the empirically relevant case where the prior strongly favors rejection.[3] Instead, when the noise distribution has a 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.[4]

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. If selection is fully anticipated in equilibrium—for example because the researcher's cost of pre-sample collection is known—then pre-sample collection and selection are a pure rat race when noise follows a Gumbel distribution. In that case, correctly anticipated presample collection and selection have 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 the 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.[5]

Finally, we discuss the welfare impact of unanticipated and uncertain selection. 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 a random realization obtains the same payoff as when observing (and correctly expecting) a random realization. Intuitively, unexpected selection increases acceptance, but 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, we show that uncertainty in the extent of selection tends to damage the evaluator. This effect is most transparent in the Gumbel case, as indifference to selection hinges on the evaluator's exact knowledge of the extent of selection k.

Related Literature. Concerns about data selection and manipulation have long been voiced by

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

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

[5] 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.

the science and medicine literature and have led to important policy responses.[6] However, there is a dearth of modeling in the area.[7] An early exception is Blackwell and Hodges (1957), who analyze how an evaluator should optimally design a sequential experiment to minimize *selection bias*, a term they coined to represent the fraction of times a strategic researcher is able to correctly forecast the treatment assignment.[8] However, they did not model the information available to the researcher at the assignment stage. Moreover, the ensuing literature focused on exogenous selection bias and on how to adjust for it, rather than on its strategic origin and its impact on the quality of inference. Once we explicitly model information, we characterize situations in which selection actually benefits the evaluator, contrary to what Blackwell and Hodges (1957) stipulate.

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

In a complementary approach to modeling conflicts of interest in statistical testing, Banerjee, Chassang, and Snowberg (2017) propose a theory of a researcher facing an adversarial evaluator who will challenge any prior information. In another complementary approach, Tetenov (2016) analyzes an evaluator's optimal commitment to a decision rule when privately informed researchers select into costly testing. Instead, we focus on the impact of a researcher's manipulation of data on the welfare of an uncommitted evaluator.

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

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

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

[9] See also Henry (2009), Dahm, Gonzàlez, and Porteiro (2009), Felgenhauer and Schulte (2014), and Hoffmann, Inderst, and Ottaviani (2014) for persuasion models with endogenous information acquisition. 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.

2 Statistical Setup

A Bayesian 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 $p \in (0,1)$ to state θ_H. The evaluator faces a binary decision, 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 deciding, 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, called an *experiment*, with logconcave density f.[10] The evaluator optimally accepts if and only if the conditional expectation of the state given the signal realization is greater than or equal to R, that is,

$$pf(x-\theta_H)\theta_H + (1-p)f(x-\theta_L)\theta_L \geqslant \big[pf(x-\theta_H) + (1-p)f(x-\theta_L)\big]R.$$

Thus, the evaluator accepts when the likelihood ratio, $\ell_F(x) \equiv f(x-\theta_H)/f(x-\theta_L)$, is at least as large as the *acceptance hurdle* defined by

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

Logconcavity of f implies the monotone likelihood ratio property: $\ell_F(x)$ is increasing.[11] Thus, the evaluator accepts if and only if the signal realization x is greater than or equal to some (possibly unbounded) threshold $-\infty \leqslant \bar{x} \leqslant \infty$. The optimal threshold is $\bar{x}^*_F(\bar{\ell}) = \infty$ if $\ell_F(x) < \bar{\ell}$ for every x, and $\bar{x}^*_F(\bar{\ell}) = \inf\{x : \ell_F(x) \geqslant \bar{\ell}\}$ otherwise.

Reformulation. Any threshold $\bar{x}$ induces a false positive (type I error) rate $\alpha = 1 - F(\bar{x}-\theta_L)$, the probability of an incorrect acceptance in state θ_L, and a false negative (type II error) rate $\beta = F(\bar{x}-\theta_H)$, the probability of an incorrect rejection in state θ_H. Given the decision to accept above and reject below $\bar{x}$, conditional on state θ the evaluator's payoff is $F(\bar{x}-\theta)R + [1-F(\bar{x}-\theta)]\theta$. Thus, disregarding constants the evaluator's expected payoff becomes

$$-(1-p)(R-\theta_L)\alpha - p(\theta_H - R)\beta, \tag{2}$$

a linear and strictly decreasing function of α and β. The term $(1-p)(R-\theta_L)$ can be interpreted as the marginal cost of a false positive, and $p(\theta_H - R)$ as the marginal cost of a false negative. The acceptance hurdle $\bar{\ell}$ then measures the relative marginal cost of false positives.

The evaluator can achieve any pair (α, β) induced by some threshold, that is, any pair such that $\alpha = 1 - F(\bar{x}-\theta_L)$ and $\beta = F(\bar{x}-\theta_H)$ for some $-\infty \leqslant \bar{x} \leqslant \infty$. The two error types are then related

[10] A density f is logconcave if $\partial^2 \log f(x)/\partial x^2 \leqslant 0$; see Bagnoli and Bergstrom (2005) for a primer.

[11] Since we consider experiments with arbitrary values $\theta_H > \theta_L$, the monotone likelihood ratio property is not only necessary but also sufficient for logconcavity of the error distribution. See e.g. Lehmann and Romano (2005, p. 323).

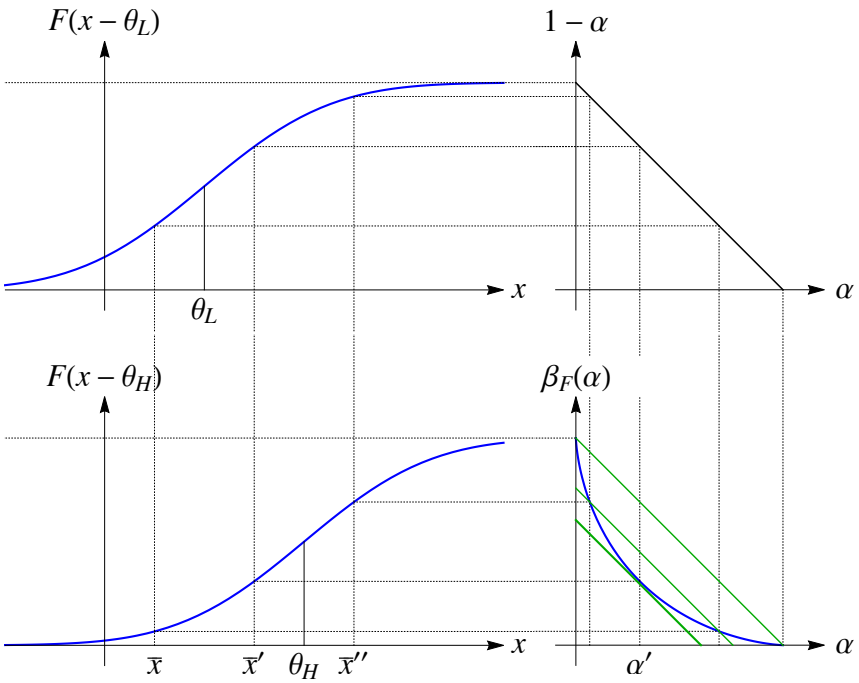


Figure 2: Reformulated problem and optimal solution in the normal case.

by the following function, which we call the *information constraint* of experiment F:

$$\beta = \beta_F(\alpha) = F\left(F^{-1}(1-\alpha) + \theta_L - \theta_H\right). \tag{3}$$

Figure 2 illustrates the construction of the information constraint from the cumulative distribution functions of the signal in the two states, when F is standard normal. Note that β_F is decreasing and convex. Indeed, by decreasing the threshold the evaluator accepts more often, increasing false positives but decreasing false negatives. Moreover, at any interior point $(\alpha, \beta_F(\alpha)) \in (0,1)^2$ the slope of the constraint is the negative of the likelihood ratio at the corresponding threshold:

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

To simplify the presentation of our results, it will be convenient to define the slope of the information constraint also for non-interior points on its graph. Specifically, we define $\beta_F'(0) = -\infty$ and $\beta_F'(\alpha) = 0$ for all $\alpha \in [0,1]$ such that $\beta_F(\alpha) = 0$.[12]

[12]What we call *information constraint* has appeared in varous guises and names in the statistical literature. For example, Jewitt (2007) considers the probability-probability plot which shows β as a function of $1-\alpha$, namely the parametric plot of the point $(F(x-\theta_L), F(x-\theta_H))$, as the signal realization x varies. Torgersen (1991) uses both names *β-functions* and *Neyman-Pearson functions* to describe $1-\beta$ as a function of α. The latter relationship is also known as the *receiver operating characteristic curve* or *ROC graph*.

Dividing (2) by $p(\theta_H - R)$ and using (1), 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 $\beta = \beta_F(\alpha)$. More simply, the evaluator chooses $\alpha \in [0,1]$ to minimize $\bar{\ell}\alpha + \beta_F(\alpha)$. Letting $\alpha_F^*(\bar{\ell})$ denote the optimal solution to the reformulated problem, we have $\alpha_F^*(\bar{\ell}) = 1 - F(\bar{x}_F^*(\bar{\ell}) - \theta_L)$, that is,

$$\alpha_F^*(\bar{\ell}) = \sup\left\{\alpha \in [0,1] : -\beta_F'(\alpha) \geqslant \bar{\ell}\right\}.$$

Figure 2 illustrates the optimal solution in the symmetric case where type I and type II errors have equal marginal costs, namely $\bar{\ell} = 1$. The green lines with slope $-\bar{\ell} = -1$ are the evaluator's indifference lines. The top line connects the points $(1,0)$ and $(0,1)$, the pairs of errors achievable by always accepting (setting threshold at $-\infty$) and always rejecting (setting threshold at ∞). The evaluator can do better than that, e.g. by setting the threshold at $\bar{x}$ or $\bar{x}''$, both of which lead to the intermediate indifference line in the figure. Finally, the lower indifference line, i.e. the unique (by convexity of β_F) line tangent to β_F, identifies the optimal solution $\alpha_F'^*(1)$, corresponding to the optimal threshold $\bar{x}_F'^*(1)$ in the original problem.

Finally, note that the information constraint and both the solution and the value of the evaluator's problem remain the same if the noise is translated by a constant c, so that its cumulative distribution function is $F(\varepsilon - c)$ instead of $F(\varepsilon)$.

Random vs. Selected Experiment. Our main goal in this paper is to compare the following two scenarios in terms of evaluator's welfare. In the first scenario, the evaluator observes a *random* data point: the noise term ε is drawn from some distribution F. In the second scenario, the evaluator observes a *selected* data point: the noise term is the maximum of $k > 1$ independent draws from F.

The case of selected data corresponds to the experiment F^k, with density $kF^{k-1}f$. The latter inherits logconcavity from f, so the evaluator again uses a cutoff rule, but the optimal threshold is higher than under F, and increasing in k. Indeed, the evaluator accepts if and only if

$$\ell_{F^k}(x) = [F(x-\theta_H)/F(x-\theta_L)]^{k-1}\ell_F(x) \geqslant \bar{\ell}, \tag{4}$$

and the term in square brackets is less than 1 for each x. By our reformulation, the evaluator's optimal decision minimizes $\bar{\ell}\alpha + \beta$ subject to the information constraint $\beta = \beta_{F^k}(\alpha)$.

Potential Outcomes Interpretation. Selection bias is an important concern in observational studies, as well as in the practice of controlled experiments; see Schulz (1995) and Berger (2005) for extensive accounts and examples of subversion of randomization in clinical trials.[13] Following

[13] 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.

Neyman (1923) and Rubin (1974, 1978), consider a population of individuals and two alternative treatments—a default, known treatment a and a new treatment b whose benefit beyond the default is unknown. Let $Y_i(t)$ denote the potential outcome of individual i when receiving treatment $t = a, b$. Assume that the unknown treatment effect on individual i, the difference $Y_i(b) - Y_i(a)$, is the same for each individual i and can take only two values. Letting $\varepsilon = Y_i(a)$ and $x = Y_i(b)$ it is easily seen that the random experiment F is nothing else than a randomized controlled trial with a random control group.[14] Analogously, experiment F^k corresponds to the selection-biased trial where the assignment to treatment is correlated with the untreated potential outcome—the treated individual's untreated outcome $\varepsilon = Y_i(a)$ is systematically higher than for a random individual (maximum in a group of k).

3 Comparing Experiments by Local Dispersion

Is the evaluator better off with a random experiment or with a selected experiment? Selection has two opposite effects on the acceptance probability in each state, and hence contrasting effects on the value of the evaluator's problem. The higher optimal threshold tends to lower acceptance in both states, decreasing α and increasing β. However, F^k first-order stochastically dominates F, which tends to raise acceptance in both states, increasing α and decreasing β. To shed light on these effects, in this section we take a step back, and address the more general problem of comparing any two experiments F and G, where G is not necessarily F^k.

3.1 Local Dispersion

Consider two experiments F and G with logconcave densities. If we know the parameters of the evaluator's problem and we compute the information constraints of the two experiments, β_F and β_G, then we can immediately determine the evaluator's preference: combining (1), (2), and (3), G is preferred to F if and only if

$$\bar{\ell}\alpha_G^*(\bar{\ell}) + \beta_G(\alpha_G^*(\bar{\ell})) \leqslant \bar{\ell}\alpha_F^*(\bar{\ell}) + \beta_F(\alpha_F^*(\bar{\ell})),$$

with strict preference or indifference if the inequality holds strictly or with equality, respectively. What is behind the evaluator's preference for one or the other experiment? Is there a method to determine this preference without computing information constraints or having detailed information about the problem parameters?

[14] More precisely, our analysis describes an experiment—random or selected—without a control group. But the case of an experiment with a randomly chosen control group is formally equivalent—observing ε_j for any other individual j carries no information about the treatment effect θ. However, adding a control group *can* benefit the evaluator in the selected experiment, when the untreated units are chosen among the $k-1$ that were not selected for treatment.

Following Blackwell (1951, 1953), suppose first that we are interested in characterizing those pairs of experiments F and G such that G is *globally* preferred to F, that is, preferred to F for *all* parameter values. It is easy to see that this global preference holds if and only if for all $\theta_H > \theta_L$ the information constraint of G lies entirely below that of F, or $\beta_G(\alpha) \leqslant \beta_F(\alpha)$ for all $\alpha \in [0,1]$. Lehmann (1988, Theorem 5.2) showed that the latter property is in turn equivalent to the following:[15] the quantile difference $G^{-1} - F^{-1}$ is weakly decreasing, i.e.

$$G^{-1}(v) - F^{-1}(v) \leqslant G^{-1}(u) - F^{-1}(u) \qquad \text{for all } 0 < u < v < 1. \tag{5}$$

This criterion defines that G is *less dispersed* than F, a notion of stochastic ordering proposed by Bickel and Lehmann (1979).[16] Intuitively, G is less spread out than F and thus provides better information about the state—for an example, see Figure 5 below.

Lehmann's result provides a sufficient condition that is intuitive, easy to check, and requires no knowledge of the problem parameters: given any value of θ_L, θ_H, p and R, if G is less dispersed than F then the evaluator must prefer G to F. Moreover, the condition is sharp, for no weaker condition can guarantee that G is always the preferred experiment—if G is *not* less dispersed than F, then there exist parameter values such that the evaluator strictly prefers F to G.

Despite its sharpness, Lehmann's criterion is often inapplicable—the ordering of experiments in terms of dispersion is only partial. If the quantile difference $G^{-1} - F^{-1}$ is not monotone, then for some $\theta_H > \theta_L$ the information constraints β_F and β_G must fail to lie one below the other, and the evaluator's preference then depends on the value of p and R. But suppose now that we *do* know something about the problem parameters and are therefore interested in determining the evaluator's preference in some *subset* of their possible values. For example, we might be interested in checking whether G is preferred to F when the acceptance hurdle $\bar{\ell}$ is sufficiently high—say, for fixed values of p, θ_L and θ_H and all sufficiently large values of R. Can we obtain a criterion that, like Lehmann's, is easy to check and does not need computing the information constraints, but allows comparison of experiments not ordered by dispersion?

We now show that a useful criterion for comparing experiments whose information constraints may cross can indeed be given, in terms of a notion that we call *local dispersion*. To introduce this notion, let $\delta = F^{-1}(v) - F^{-1}(u)$ and rewrite condition (5), that G is less dispersed than F, as follows:

$$F(F^{-1}(u) + \delta) \leqslant G(G^{-1}(u) + \delta) \qquad \text{for all } \delta > 0 \text{ and } 0 < u < 1. \tag{6}$$

[15] Briefly: let $v = 1 - \alpha$ and $u = \beta_F(\alpha)$. Like in the top part of Figure 1, G achieves no higher $\beta_G(\alpha)$ because $G^{-1}(v) - G^{-1}(u) \leqslant F^{-1}(v) - F^{-1}(u)$.

[16] Bickel and Lehmann (1979), in turn, credit Brown and Tukey (1946) for the essence of the definition. For applications of the notion of dispersion to economic problems, see Persico (2000) and Jewitt (2007). See also Quah and Strulovici (2009), who extend Lehmann's result from monotone procedures to a more general family of decision problems.

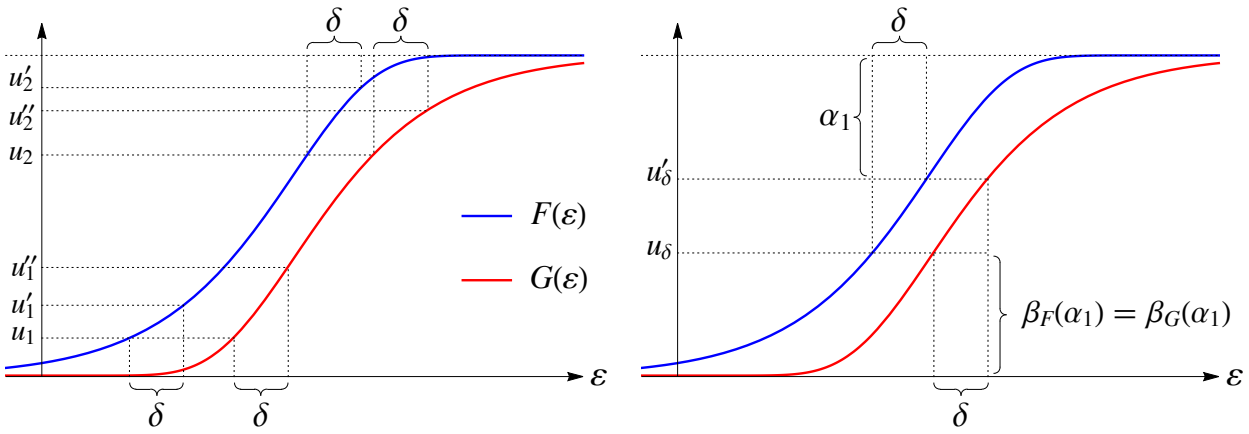


Figure 3: G is locally less δ-dispersed than F on $[0, u_\delta]$ and more δ-dispersed than F on $[u_\delta, 1]$.

Written this way, the condition says that, given any δ and u, the probability that in experiment F the noise takes a value between $F^{-1}(u)$ and $F^{-1}(u)+\delta$ is smaller than the probability that in experiment G the noise takes a value between $G^{-1}(u)$ and $G^{-1}(u)+\delta$. In other words, G is more concentrated in the interval $[G^{-1}(u), G^{-1}(u)+\delta]$ than F in the interval $[F^{-1}(u), F^{-1}(u)+\delta]$.

Our definition of local dispersion weakens both quantifiers in condition (6), requiring only that the inequality holds for a fixed value of δ and for all u in some interval.

Definition 1. Experiment G is *locally less δ-dispersed* than experiment F on $[u_1, u_2] \subseteq [0,1]$ if

$$F(F^{-1}(u)+\delta) \leqslant G(G^{-1}(u)+\delta) \qquad \text{for all } u_1 \leqslant u \leqslant u_2.$$

Figure 3 illustrates a case where G is locally less δ-dispersed than F for low values of u, and more δ-dispersed than F for large values of u.[17] Consider a low value of u, e.g. $u = u_1$. Going from $F^{-1}(u_1)$ to $F^{-1}(u_1)+\delta$, distribution F increases from u_1 to u_1'. But going from $G^{-1}(u_1)$ to $G^{-1}(u_1)+\delta$, distribution G increases more, from u_1 to $u_1'' > u_1'$. At large values of u, the opposite happens. For instance, starting from $u = u_2$, distribution F increases to u_2' while G increases to $u_2'' < u_2'$. At the critical value u_δ experiment G switches from being less to being more δ-dispersed than F. The quantile difference $G^{-1}(u) - F^{-1}(u)$ is decreasing at $u = u_\delta$, but moving horizontally by δ both F and G reach the value u_δ', where the quantile difference is increasing. On average, the two effects cancel—starting from u_δ, the two distributions grow by the same amount $u_\delta' - u_\delta$.

[17] The plots are drawn for the case where $\theta_H - \theta_L = 1$ and the information constraints β_F and β_G correspond to Gumbel's minimum and maximum extreme value distributions, respectively: $F(\varepsilon) = 1 - e^{-e^{\varepsilon}}$ and $G(\varepsilon) = e^{-e^{-\varepsilon}}$.

3.2 Comparison of Experiments

We are now ready to state our first main result—the characterization of how local dispersion of noise determines the value of information:

Theorem 1. *For all experiments F and G and for all $N \geqslant 1$ and $\theta_H > \theta_L$, letting $\delta = \theta_H - \theta_L$, the following conditions are equivalent:*

(L) There exist $0 = \ell_1 \leqslant \cdots \leqslant \ell_{2N+1} = \infty$ such that, for all $n = 1, \ldots, N$, the evaluator prefers F to G for $\bar{\ell} \in [\ell_{2n-1}, \ell_{2n}]$ and G to F for $\bar{\ell} \in [\ell_{2n}, \ell_{2n+1}]$.

(A) There exist $1 = \alpha_1 \geqslant \cdots \geqslant \alpha_{2N+1} = 0$ such that, for all $n = 1, \ldots, N$, $\beta_F(\alpha) \leqslant \beta_G(\alpha)$ for all $\alpha \in [\alpha_{2n}, \alpha_{2n-1}]$ and $\beta_F(\alpha) \geqslant \beta_G(\alpha)$ for all $\alpha \in [\alpha_{2n+1}, \alpha_{2n}]$.

(D) There exist $0 = u_1 \leqslant \cdots \leqslant u_{2N+1} = 1$ such that, for all $n = 1, \ldots, N$, F is locally less δ-dispersed than G on $[u_{2n-1}, u_{2n}]$ and more δ-dispersed than G on $[u_{2n}, u_{2n+1}]$.

According to Lehmann's (1988) Theorem 5.2, when two experiments are not ordered by dispersion we cannot determine the evaluator's preference without having information on the parameters of the problem. Theorem 1 shows, however, that we do not need much information to determine the value of information. The evaluator's preference depends on the problem parameters in a way that is both intuitive and easily predictable from the shape of the quantile difference, much like in Lehmann's global result.

Before discussing the intuition and the mechanics of Theorem 1, it is worth highlighting two special cases. First, letting $N = 1$, with $\ell_2 = \ell_1 = 0$ and $u_2 = u_1 = 0$, we obtain Lehmann's (1988) Theorem 5.2 as an immediate corollary. Second, letting again $N = 1$, but this time with $0 = \ell_1 < \ell_2 < \ell_3 = \infty$ and $0 = u_1 < u_2 < u_3 = 1$, our theorem characterizes when the information constraints of F and G have a single crossing—this happens in the example of Figure 3, which we will shortly revisit. In this single-crossing case, there is a simple criterion to identify the switch in local dispersion: if F and G have the same support and the quantile difference $G^{-1} - F^{-1}$ is first decreasing and then increasing, then, for every δ not too large, G is locally first less and then more δ-dispersed than F. Thus, while the definition of local dispersion depends on the value of δ, there is a simple case—which will be useful in the analysis of Section 4 below—where just looking at the quantile difference is enough. We report this conclusion in the following proposition.

Proposition 1. *Let $0 < u_1 < 1$. Let F and G be experiments such that $G^{-1}(u) - F^{-1}(u)$ is decreasing on $[0, u_1]$ and increasing on $[u_1, 1]$. Assume that F and G have a common (possibly unbounded) support $[\underline{\varepsilon}, \overline{\varepsilon}]$. For every $0 < \delta < \overline{\varepsilon} - \underline{\varepsilon}$ there exists $u_\delta \in [0,1]$ such that G is locally less δ-dispersed than F on $[0, u_\delta]$ and more δ-dispersed than F on $[u_\delta, 1]$.*

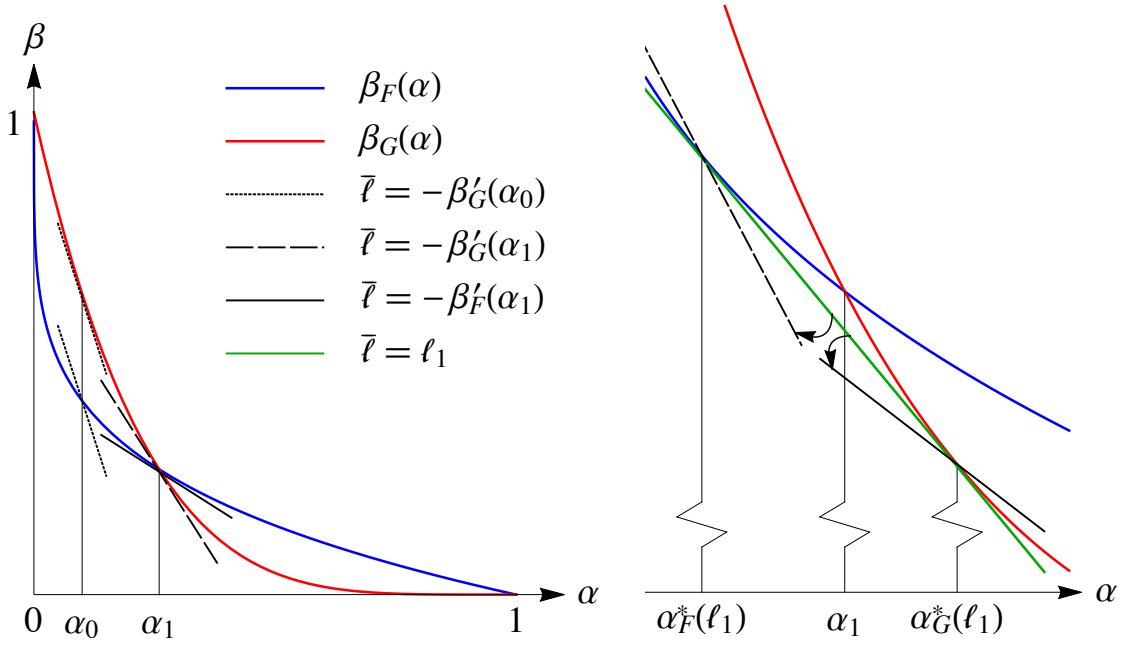


Figure 4: Building the preference pattern from the sign pattern of the constraint difference.

The intuition for the equivalence in Theorem 1 is based on a simple observation: asking for the inequality in (6) to hold only for $\delta = \theta_H - \theta_L$ (instead of all $\delta > 0$) and for all u in some interval is the same as asking that $\beta_G(\alpha) \leqslant \beta_F(\alpha)$ for all α in a certain corresponding interval. Thus, the alternating local dispersion pattern in condition (D) of Theorem 1 is equivalent to a suitably defined alternating sign pattern of the difference $\beta_F(\alpha) - \beta_G(\alpha)$, as in condition (A). But, crucially, each value of α where the information constraints cross corresponds to a critical value of $\bar{\ell}$ where the evaluator is indifferent and preference switches from one experiment to the other. Intervals between crossings instead correspond to intervals of values of $\bar{\ell}$ where the experiment with the lower information constraint is preferred. Thus, the pattern in condition (A) is in turn equivalent to the alternating preference pattern in condition (L).

Appendix A establishes and discusses in detail the intuitive but subtle equivalence between conditions (A) and (L) through two lemmas. Here we illustrate the main idea with the help of Figure 4, where F and G are the experiments already discussed in Figure 3. The left diagram gives the overall picture, the right diagram zooms in around the point where the information constraints cross. For low values of $\bar{\ell}$, and precisely for $\bar{\ell}$ between 0 and $-\beta'_F(\alpha_1)$, the preferred experiment is G, for G affords a higher payoff than F even to an evaluator who suboptimally chooses $\alpha^*_F(\bar{\ell})$. Symmetrically, when $\bar{\ell}$ is larger than $-\beta'_G(\alpha_1)$, the preferred experiment is F. For $\bar{\ell}$ between $-\beta'_F(\alpha_1)$ and $-\beta'_G(\alpha_1)$, however, the comparison is not immediate—with $\bar{\ell}$ in this range, β_G lies below β_F at $\alpha^*_F(\bar{\ell})$, and β_F lies below β_G at $\alpha^*_G(\bar{\ell})$, i.e. changing experiment without reoptimizing on α leads to a lower payoff. But since the information constraints have exactly one interior crossing, at α_1, Lemma 1 guarantees that preference switches exactly once, at the critical value ℓ_1

of $\bar{\ell}$. Thus, the optimizing evaluator still prefers G as $\bar{\ell}$ increases from $-\beta'_F(\alpha_1)$ to ℓ_1, while F remains preferred as $\bar{\ell}$ decreases from $-\beta'_G(\alpha_1)$ down to ℓ_1.

Lemma 2 in Appendix A allows us to go the other way around. Take any two experiments F and G and a critical value ℓ_1 where preference switches from G to F as $\bar{\ell}$ rises above ℓ_1. Then the convex envelope of β_F and β_G must be tangent to β_F at $\alpha^*_F(\ell_1)$ and to β_G at $\alpha^*_G(\ell_1)$. But then, by their continuity, the two constraints must cross at a point α_1 between $\alpha^*_F(\ell_1)$ and $\alpha^*_G(\ell_1)$.

Remark on Bayesian and Frequentist Hypothesis Testing. In the theory of statistical testing, the evaluator's optimal decision in an experiment F, to accept if and only if the observed signal x satisfies $\ell_F(x) \geqslant \bar{\ell}$, is called the *Bayes solution* or, in classical statistics terms, a *likelihood ratio test*, where $\alpha^*_F(\bar{\ell})$ is the *significance level* of the test and $1-\beta_F(\alpha^*_F(\bar{\ell}))$ its *power*.[18]

As we have already noted, a Bayesian evaluator prefers another experiment G to F for every value of $\bar{\ell}$ if and only if $\beta_G(\alpha) \leqslant \beta_F(\alpha)$ for every α. Thus, the Bayesian and the frequentist comparison agree when F and G are compared *globally* over all values of $\bar{\ell}$ or all significance levels: G is preferred to F by the Bayesian evaluator for every value of $\bar{\ell}$ if and only if, for every significance level α, the most powerful test based on G provides greater power than the most powerful test based on F.

The example in Figure 4 shows that the Bayesian and the frequentist approach differ, however, when F and G are compared *locally* in an interval of values of $\bar{\ell}$ or an interval of significance levels. Under experiment F, each value of α in the interval $[\alpha^*_F(\ell_1), \alpha_1]$ is the optimal solution for some value of $\bar{\ell}$ between $-\beta'_F(\alpha_1)$ and ℓ_1. Holding fixed the significance level at that value of α, the likelihood ratio test based on F provides greater power than that based on G, as β_G lies above β_F in the interval $[\alpha^*_F(\ell_1), \alpha_1]$. But our Bayesian evaluator, who does not keep fixed but rather optimizes on the type I error, nevertheless prefers G to F when $\bar{\ell}$ is between $-\beta'_F(\alpha_1)$ and ℓ_1.

4 Welfare Impact of Selection

In this section we use the notions of dispersion and local dispersion to address the main question in this paper: How is the welfare impact of selection related to the noise distribution and to the parameters of the evaluator's problem?

4.1 Global Impact of Selection

Toward a full answer to our question, we first provide a characterization of experiments F where the extent of selection, k, has a *global* and *monotone* effect: for all parameter values, increasing k

[18] See Torgersen (1991) for more general results.

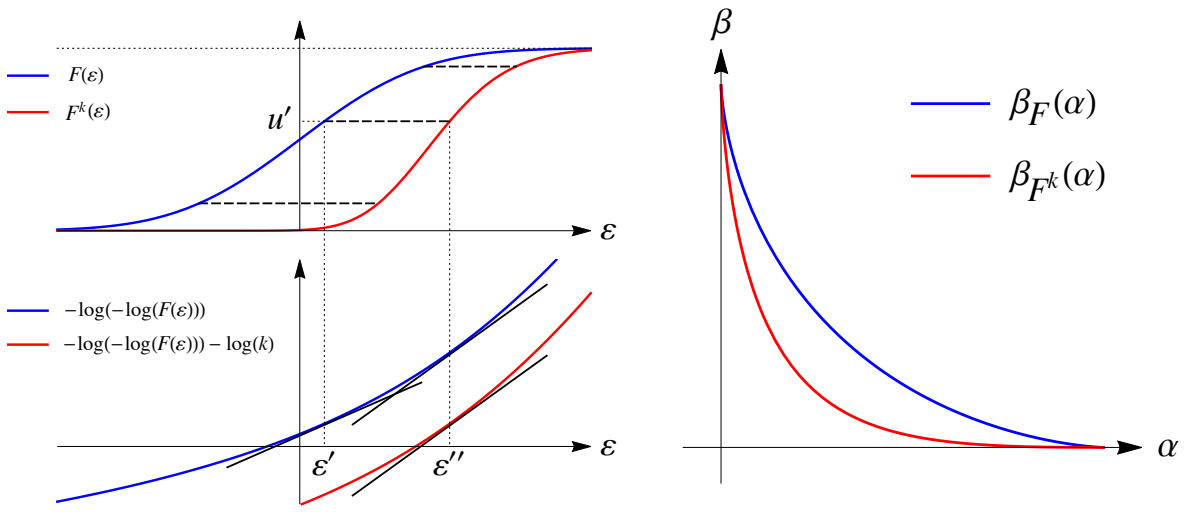


Figure 5: Selection in normal experiment: double-log transformation and convexity criterion.

makes the evaluator better off, or worse off—in other words, F^k is less dispersed the greater is k, or more dispersed the greater is k. This is our second main result:

Theorem 2. *F^k is less dispersed the greater is $k \geqslant 1$ if and only if $-\log(-\log F)$ is convex. Likewise, F^k is more dispersed the greater is $k \geqslant 1$ if and only if $-\log(-\log F)$ is concave.*

By Lehmann's (1988) Theorem 5.2, we obtain immediately the following corollary:

Proposition 2. *The evaluator prefers $F^{k'}$ to F^k (resp. F^k to $F^{k'}$) for all $k' \geqslant k \geqslant 1$ and all values of θ_L, θ_H, p and R, if and only if $-\log(-\log F)$ is convex (resp. concave).*

To gain intuition for the role of double logconvexity in our characterization, it is helpful to rewrite the condition that an experiment G is less dispersed than another, F, in yet another way, equivalent to (5) and (6) above:

$$f\big(F^{-1}(u)\big) \leqslant g\big(G^{-1}(u)\big) \qquad \text{for all } 0 < u < 1. \tag{7}$$

Thus, G is less dispersed than F if the slope of G at the quantile $G^{-1}(u)$ is steeper than the slope of F at the quantile $F^{-1}(u)$, for every u. Consider now comparing F with $G = F^k$. In order to compare the slopes of these distributions at their quantiles, as required by (7), it is convenient to first transform F and F^k in such a way that the transformed functions are parallel shifts of each other. The suitable transformation is the strictly increasing function $u \mapsto -\log(-\log u)$, because $-\log(-\log(F^k)) = -\log(-\log F) - \log k$. The transformation is illustrated in Figure 5 for the case of a standard normal experiment F.[19] Condition (7) requires that, for every $u' \in [0,1]$, the slope of

[19] The plots are drawn for $k = 8$, but of course any $k > 1$ gives the same qualitative result.

F computed at the quantile $\varepsilon' = F^{-1}(u')$ be less than the slope of F^k computed at the corresponding quantile $\varepsilon'' = (F^k)^{-1}(u')$. Since our double-log transformation is strictly increasing, this property is equivalent to the slope of $-\log(-\log F)$ at ε' being less than the slope of $-\log(-\log F^k)$ at ε''. But this is the same as saying that the slope of $-\log(-\log F)$ itself is greater at ε'' than at ε'. With $\varepsilon'' > \varepsilon'$, this is satisfied when $-\log(-\log F)$ is a convex function. Figure 5 shows that this is exactly what happens with a normal signal. In this case, the information constraint β_{F^k} lies everywhere below β_F, and further below the greater is k.

Variation of $k > 1$ lets us thus compare the slope at ε' with the slopes at any $\varepsilon'' > \varepsilon'$ in the support of F. This makes double logconvexity a necessary property. Let us remark that we described selection with a natural number k, but the interpretation—as well as the statement in the theorem—for real numbers $k > 1$ is equally valid. Increasing selection from k to $k' > k$ changes the experiment from F^k to $F^{k'} = (F^k)^{k'/k}$. This is akin to starting from F^k and applying selection of extent $k'/k > 1$. Our comparative statics result in Theorem 2 characterizes when this reduces dispersion. Note the implication that experiment F is such that selection monotonically benefits (or hurts) the evaluator if and only if F^k has the same property for every real number $k > 1$—both properties depend on double logconvexity of F.

There is only one distribution F such that $-\log(-\log F)$ is both convex and concave, i.e. linear, namely Gumbel's (maximum) extreme value distribution, $F(\varepsilon) = \exp(-\exp(-\varepsilon))$. This distribution, which plays a special role in the ensuing analysis, is such that for every k the experiment F^k is neither less nor more dispersed than F and the evaluator is therefore indifferent to selection. The following intuitive argument also leads to the same conclusion. With selection at extent k, the noise distribution is $F^k(\varepsilon) = \exp(-k\exp(-\varepsilon)) = F(\varepsilon - \log k)$. Thus, compared to a random sample, selection inflates noise by a constant, $\log(k)$. This implies that the information constraints in the random and in the selected experiment coincide: $\beta_F = \beta_{F^k}$. The evaluator adjusts for the constant inflation in the noise distribution, and is back to square one.[20]

Restatement in Terms of Shape of Reverse Hazard Rate. Taking its first derivative, the function $-\log(-\log F(\varepsilon))$ is easily seen to be convex if and only if the ratio $[f(\varepsilon)/F(\varepsilon)]/\log F(\varepsilon)$ of the reverse hazard rate to the reverse hazard function—sometimes also called cumulative reverse hazard rate—is decreasing.[21] But $[f(\varepsilon)/F(\varepsilon)]/\log F(\varepsilon)$ decreasing can also be seen as

$$\frac{d[f(\varepsilon)/F(\varepsilon)]/d\varepsilon}{f(\varepsilon)/F(\varepsilon)} \leqslant \frac{d\log F(\varepsilon)/d\varepsilon}{\log F(\varepsilon)},$$

which says that the rate at which the reverse hazard rate decreases—recall that by logconcavity of F, f/F decreases—is lower than the rate at which the reverse hazard function increases—note that the hazard function is negative and increasing for every random variable. Thus, selection benefits

[20] Recall that the value of the evaluator's problem is unaffected by a change in the location of the noise distribution.

[21] See e.g. Marshall and Olkin (2007) for definitions and properties of the hazard rate.

the evaluator, increasingly in k, if and only if the reverse hazard rate of the noise distribution decreases less fast than the reverse hazard function increases.

Restatement in Terms of Elasticity of Quantile Density Function. Note that the derivative of the transformation $u \mapsto -\log(-\log u)$ is given by $-1/(u \log u)$. Taking the next derivative, convexity of $-\log(-\log F)$ can be restated as

$$\frac{f'(\varepsilon)/f(\varepsilon)}{f(\varepsilon)/F(\varepsilon)} > \frac{1+\log F(\varepsilon)}{\log F(\varepsilon)} \qquad \text{for all } \varepsilon. \tag{8}$$

This condition has another interpretation. Denote the quantile function by $Q(u) = F^{-1}(u)$. Its derivative is the quantile density function $q(u) = 1/f\left(F^{-1}(u)\right)$, also known as Tukey's sparsity function.[22] The elasticity of the quantile density function is

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

For the Gumbel distribution, $Q(u) = -\log(-\log u)$ and hence $q(u) = -1/(u \log u)$. Thus,

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

Summing up, our key condition (8) can also be restated as saying that F has a quantile density function that is less elastic than the quantile density function of the Gumbel distribution.

Logistic Example. Besides the normal case discussed earlier, another instance where the evaluator prefers more selection is when noise is drawn from the logistic distribution, $F(\varepsilon) = 1/(1+e^{-\varepsilon})$. In this case, we have $Q(u) = \log[u/(1-u)]$ and hence $q(u) = 1/[u(1-u)]$. Thus,

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

The quantile density function is less elastic than Gumbel's, therefore any amount of selection benefits the evaluator, and benefits more as k increases.

Exponential Example. Our main example of the opposite case, where more selection hurts the evaluator, is the exponential distribution $F(\varepsilon) = 1 - e^{-\varepsilon}$, for $\varepsilon \geqslant 0$. In this case, $Q(u) = -\log(1-u)$ and hence $q(u) = 1/(1-u)$. Thus,

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

The quantile density function is more elastic than Gumbel's, therefore any amount of selection hurts the evaluator, ever more as k increases.

[22] See Parzen (2004) for an introduction to quantile probability modeling.

It is worth remarking that the exponential distribution is not the only double logconcave distribution with a logconcave density. For instance, given any $a < -1$, the distribution F such that

$$F(\varepsilon) = e^{\frac{1}{1+a}\left[\left(1-e^{-\varepsilon}\right)^{1+a}-1\right]} \qquad \text{for } \varepsilon > 0 \tag{9}$$

is such that both $-\log(-\log F)$ and $\log f$ are strictly concave.

Contribution to Stochastic Ordering of Order Statistics. Previous results in the literature on stochastic ordering 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.[23] Given that 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.[24] The novel characterization in Theorem 2 applies more generally to the relevant case of distributions with logconcave densities.

Empirics of Double Logconvexity. Given that $-\log(-\log F)$ and $-\log(-\log F^k)$ only differ by a constant, we conclude that F is double logconcave (double logconvex) if and only if F^k is double logconcave (double logconvex). From this selection-invariance property of double logconvexity/logconcavity we obtain a simple practical criterion to assess the possible impact of selection in empirical data. Whether selection is known to have occurred or not, empirical distributions of treated outcome with a double logconcave shape should "raise a flag": if selection did occur, then the analyst is bound to having less informative data, even when the analyst *is* aware of selection and correctly sets the acceptance standard. Instead, double logconvex data indicate that if the analyst does take selection into account, then selection actually results in a more informative experiment. In the online supplementary appendix Di Tillio, Ottaviani, and Sørensen (2017) we develop empirical tests for double logconvexity and illustrate how to apply the tests to some data sets drawn from recent experiments published in economics.

4.2 Locally Variable Impact of Selection

We now turn to discussing the cases where the welfare impact of selection is positive or negative depending on the parameters of the evaluator's problem. Here, our analysis exploits both the local dispersion charaterization provided by Theorem 1 and the double logconvexity criterion of Theorem 2. To keep things simple, we focus on experiments F such that, given any $k \geqslant 1$, F is first less and then more locally dispersed than F^k (or vice versa).

[23] According to Khaledi and Kochar (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 $i \leqslant j$ and $n-i \geqslant m-j$. Setting $i = n = 1$ and $j = m = k$, we have that the maximum of k i.i.d. variables with decreasing hazard rate is more dispersed than the original variable.

[24] Theorem 2 also covers distributions with decreasing hazard rate, where $-\log(-\log(F))$ is necessarily concave.

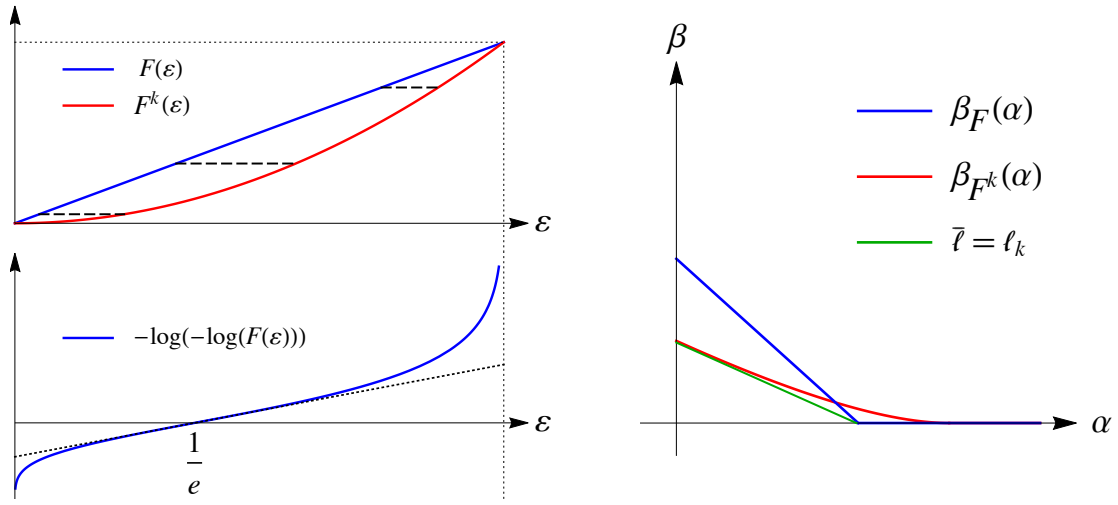


Figure 6: Selection in uniform experiment: concave then convex double-log transformation.

Proposition 3. *Let F be an experiment such that $-\log(-\log F)$ is first concave (resp. convex) and then convex (resp. concave). Then for every $k \geqslant 1$ there exists ℓ_k such that the evaluator prefers F to F^k (resp. F^k to F) for $\bar{\ell} \leqslant \ell_k$ and F^k to F (resp. F to F^k) for $\bar{\ell} \geqslant \ell_k$.*

By Theorem 2, when F is not everywhere double logconvex or everywhere double logconcave, random and selected experiment are incomparable in the sense of Lehmann (1988). The evaluator's preference for F or F^k then depends on the parameters of the specific problem at hand. However, thanks to the local criterion provided by Theorem 1, we can qualify the impact of selection without actually solving the problem. Proposition 3 offers a particularly simple criterion for making this assessment: if F is first double logconcave and then double logconvex, then the quantile difference $(F^k)^{-1}(u) - F^{-1}(u)$ is first increasing and then decreasing in u. Thus, the evaluator suffers or benefits from selection according to whether the acceptance hurdle is, respectively, below or above a certain critical value. The opposite is true if F is instead first double logconvex and then double logconcave—in this case, the quantile difference is first decreasing and then increasing.

Proposition 3 applies to two important special cases, which we now discuss.

Uniform Example. Suppose we are interested in comparing F with F^k when F is uniform, $F(\varepsilon) = \varepsilon$ for $\varepsilon \in [0,1]$. Thus, the double-log transformation of F is $-\log(-\log \varepsilon)$, which is concave for ε smaller and convex for ε greater than the inverse of Euler's number, $1/e$, as represented in Figure 6. According to Proposition 3, the evaluator is hurt by selection when more concerned about type II errors—the acceptance hurdle $\bar{\ell}$ is low—and benefits from selection when more concerned about type I errors—the acceptance hurdle $\bar{\ell}$ is high. Of course, we reach the same conclusion using Theorem 1: since $F^k(\varepsilon) = \varepsilon^k$, the quantile difference $(F^k)^{-1}(u) - F^{-1}(u) = u^{1/k} - u$ is easily seen to be bell-shaped, so F is locally first less then more dispersed than F^k.

Laplace Example. Noise drawn from a Laplace distribution, where $F(\varepsilon) = (1/2)e^{\varepsilon}$ for $\varepsilon < 0$ and

$F(\varepsilon) = 1 - (1/2)e^{-\varepsilon}$ for $\varepsilon \geqslant 0$, provides our second illustration of Proposition 3. In this case, the double-log transformation of F is convex for $\varepsilon < 0$ and concave for $\varepsilon > 0$. Thus, the evaluator's preference for selection is reversed compared to a uniform experiment. Here, the evaluator prefers F^k to F for low values of the acceptance hurdle, and F to F^k for large values. Thus, the information constraint difference $\beta_F(\alpha) - \beta_G(\alpha)$ is first positive, then negative.[25] Plotting the distribution functions F and F^k reveals a U-shaped quantile difference, in accordance with Theorem 1.

4.3 Extreme Selection

To conclude our analysis of the impact of selection on the evaluator's welfare, we examine the effect of extreme selection, letting $k \to \infty$. We draw 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. Let F be an experiment and suppose that, for some nondegenerate distribution $\bar{F}$ and some sequence of numbers $a_k > 0$ and b_k,

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

for every continuity point ε of $\bar{F}$. The fundamental theorem of extreme value theory says that $\bar{F}$ must belong to one of the following three types: Gumbel, Extreme Weibull or Frechet.[26]

Recall that the distribution of the noise term is systematically shifted upwards as k increases, in the sense of first-order stochastic dominance. Hence, the location normalization sequence b_k is growing. However, the evaluator can adjust for any translation of the noise distribution without any impact on payoff.

The limit impact of selection thus hinges on whether the scale normalization sequence a_k shrinks to zero or not. If $a_k \to 0$, then the noise distribution is less and less dispersed as k grows, providing the evaluator with arbitrarily precise information about the state. If instead we can choose a constant sequence a_k, then we can also choose $a_k = 1$ for all k, and extreme selection based on experiment F then amounts to a random experiment based on $\bar{F}$.

Proposition 4. *Let $F, \bar{F}$ be experiments and let $a_k > 0$ and b_k be sequences of numbers such that $F^k(b_k + a_k\varepsilon) \to \bar{F}(\varepsilon)$ at every continuity point of $\bar{F}$. If $a_k \to 0$, then the value of the evaluator's problem converges to the full information payoff ($\alpha = 0$, $\beta = 0$). If instead $a_k = 1$ for all k, then the value of the evaluator's problem converges to the value of the problem in experiment $\bar{F}$.*

It is well known that many familiar distributions are in the domain of attraction of the Gumbel distribution. Specifically, when F is normal—or half-normal, which has the same right tails—then

[25] In a Laplace experiment—random or selected—the likelihood ratio is bounded, hence the information constraint is not tangent to the axes as $\alpha \to 0$ and $\alpha \to 1$.

[26] See e.g. Leadbetter, Lindgren, and Rootzén (1983) for a primer on extreme value theory. Müller and Rufibach (2008) show that for every logconcave distribution F has a Gumbel or Extreme Weibull limiting distribution $\bar{F}$.

a_k must be decreasing to zero—the scale normalization sequence $a_k = (2\log k)^{-1/2}$ is appropriate in this case—and the limit experiment $\bar{F}$ is the Gumbel distribution.

More generally, consider a distribution F in the *exponential power* family, with density

$$f(\varepsilon) = \frac{s}{\Gamma(1/s)} e^{-|\varepsilon|^s},$$

where s is a shape parameter and Γ is the Gamma function. This family includes the Laplace ($s = 1$), normal ($s = 2$), and uniform ($s = \infty$) distributions discussed earlier in this section. Our next result shows that when the shape parameter s is strictly greater than 1, the scale normalization sequence a_k must be decreasing to zero, and the limiting distribution $\bar{F}$ is the Gumbel distribution.

Proposition 5. *Let F be an exponential power distribution with shape parameter $s > 1$. Then $F^k(b_k + a_k\varepsilon) \to e^{-e^{-\varepsilon}}$ for some sequence of constants b_k and $a_k \to 0$. Thus, the value of the evaluator's problem in experiment F^k converges to the full information payoff as $k \to \infty$.*

We find the conclusion in Proposition 5 striking, because it is known that when F is the exponential distribution—or the Laplace, since the two distributions have the same right tails—then F also converges to the Gumbel distribution, but we can take $a_k = 1$ for each k. (The same normalizing constants work for the generalized exponential distribution defined in (9).) Thus, while extreme selection leads to full information as $k \to \infty$ for any $b > 1$ in the exponential power family, the limit result is very different when $b = 1$. The globally 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.[27]

Note that Proposition 5 does not cover the uniform case ($s = \infty$). Indeed, in this case the relevant extreme value distribution is not the Gumbel distribution but rather the Extreme Weibull. However, it is immediate to see that extreme selection leads to full information also in the uniform case. If F is uniform on the interval $[\underline{\varepsilon}, \overline{\varepsilon}]$, letting $a_k = (\overline{\varepsilon} - \underline{\varepsilon})/k$ and $b_k = \overline{\varepsilon}$ we obtain $F^k(b_k + a_k\varepsilon) \to \bar{F}(\varepsilon)$ where $\bar{F}(\varepsilon) = e^{\varepsilon}$ for $\varepsilon < 0$ (Extreme Weibull). Since $a_k \to 0$, the full information result follows from Proposition 4. Intuitively, with noise bounded above by $\overline{\varepsilon}$, as $k \to \infty$ the observed signal becomes arbitrarily concentrated around $\theta + \overline{\varepsilon}$, revealing the true value of θ.

5 Strategic Selection

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 fully biased toward acceptance—the researcher suffers no loss for type I errors. Intuitively, this researcher wants to select an individual with a high noise term—e.g., in the treatment effect setting, a good untreated

[27] Of course, as s approaches 1 the convergence to full information becomes slower.

outcome—in order to push 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.

5.1 Selective Sampling Game

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.

Proposition 6. *There exists a Bayes Nash equilibrium where the researcher chooses maximal selection, $i \in \arg\max_{1 \leqslant j \leqslant k} \varepsilon_j$, and the evaluator accepts for signals x satisfying (4), i.e. such that*

$$\frac{F^{k-1}(x-\theta_H) f(x-\theta_H)}{F^{k-1}(x-\theta_L) f(x-\theta_L)} \geqslant \bar{\ell}. \tag{10}$$

Note that the evaluator's strategy is precisely the one we have analyzed until now, when observing a signal subjected to selection of extent k. The researcher's strategy is a best response because the evaluator will observe a higher signal and hence be more likely to accept.[28]

It is worth remarking that this behavior would continue to constitute an equilibrium if we instead modified stage 1 of the game to allow the researcher to observe outcomes $x_1, \ldots, x_k$. For any realization of state θ we have $\max_{1 \leqslant j \leqslant k} x_j = \theta + \max_{1 \leqslant j \leqslant k} \varepsilon_j$. The interpretation of the setup would be different: the researcher actually possesses information about θ before carrying out the strategic sample selection. Literally, the researcher in this version of the game is selectively reporting the outcome of already conducted trials.

Finally, we note that, even without assuming equilibrium, the outcome of the equilibrium described in Proposition 6 is the only outcome compatible with the assumption that (i) both players are rational, (ii) the researcher believes that the evaluator follows a cutoff rule, and (iii) the evaluator believes in (i) and (ii).

Equilibrium Impact of Selection on Researcher's Welfare. The equilibrium effect of selection on the evaluator's welfare works through its effect on the evaluator's optimal choice of type I and

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

type II errors—we have analyzed this effect in Section 4. By Proposition 6, the equilibrium impact of selection on the researcher's welfare also hinges on the direction of change in the pair (α, β) optimally chosen by the evaluator.

For any pair (α, β) that the evaluator may choose, the researcher's payoff is

$$p(1-\beta) + (1-p)\alpha.$$

Thus, a generic indifference curve of the researcher is a line of the form

$$\beta = \left(1 - \frac{u}{p}\right) + \frac{1-p}{p}\alpha,$$

where $0 \leqslant u \leqslant 1$ is the researcher's payoff. To assess the equilibrium impact of selection on the researcher's welfare, it is therefore enough to check whether the evaluator's optimal pair of error rates in experiment F^k lies above or below the researcher's indifference line going through the optimal pair in experiment F. The researcher benefits from selection if and only if the evaluator reacts to selection by choosing a new pair (α, β) which is below that line.

To formalize this argument, in the rest of this section we fix $\theta_H > \theta_L$ and a value p for the prior—thus fixing a family of researcher's indifference curves—and we investigate changes in the researcher's welfare, moving from experiment F to F^k, as the evaluator's safety payoff R varies between θ_L and θ_H. To emphasize the fact that the acceptance hurdle $\bar{\ell}$ here varies only when R varies, we use the notation $\ell(R)$ to denote the value of $\bar{\ell}$ corresponding to R. Thus,

$$\ell(R) = \frac{1-p}{p}\frac{R-\theta_L}{\theta_H - R}.$$

Define the function $\varphi : [0,1] \to [0,1]$ as follows:

$$\beta_{F^k}(\varphi(\alpha)) - \beta_F(\alpha) = \frac{1-p}{p}(\varphi(\alpha) - \alpha) \qquad \text{for all } 0 \leqslant \alpha \leqslant 1.$$

In other words, φ is defined so that the researcher is indifferent between the evaluator choosing a type I error equal to α in experiment F or equal to $\varphi(\alpha)$ in experiment F^k.

Proposition 7. *The researcher benefits from selection at R if and only if*

$$-\beta'_{F^k}(\varphi(\alpha^*_F(\ell(R)))) \geqslant \ell(R). \tag{11}$$

Let $0 \leqslant \ell_0 < \ell_1 \leqslant \infty$ and suppose that the evaluator prefers F to F^k (resp. F^k to F) for $\bar{\ell} \in [\ell_0, \ell_1]$ and is indifferent between F and F^k for $\bar{\ell} = \ell_1$. Then the researcher loses (resp. benefits) from selection at R such that $\ell(R) = \ell_1$.

It is immediate to see the first part of Proposition 7 at work in our leading example where the double logconvexity criterion of Theorem 2 is satisfied—the case of a normally distributed

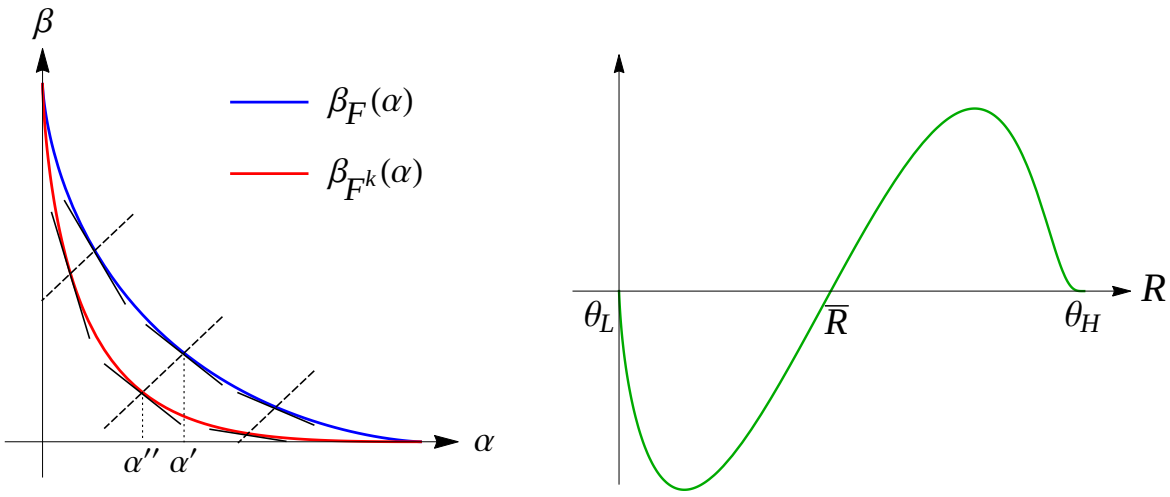


Figure 7: Impact of selection on researcher's welfare in normal experiment.

noise. In this case, as R varies between θ_L and θ_H, the evaluator's optimal solution in the random experiment, $\alpha^*_F(\ell(R))$, varies between $\alpha = 1$ and $\alpha = 0$, and the difference $\beta'_F(\alpha) - \beta'_{F^k}(\varphi(\alpha))$ is bell-shaped. Since $-\beta'_F(\alpha^*_F(\ell(R))) = \ell(R)$ for every $\theta_L \leqslant R \leqslant \theta_H$, we conclude by Proposition 7 that the researcher suffers from selection when R is below a critical value $\bar{R}$, and benefits from selection when R is above $\bar{R}$.

We illustrate this conclusion in Figure 7. The dashed lines are the researcher's indifference curves, while the green curve represents the researcher's payoff difference between selected and random experiment. When R is low, inequality (11) is violated—the researcher suffers from selection. When R is high, inequality (11) holds, and the researcher benefits. At the critical value $\bar{R}$, the researcher is indifferent between random and selected experiment—the respective optimal solutions $\alpha' = \alpha^*_F(\ell(R))$ and $\alpha'' = \alpha^*_{F^k}(\ell(R))$ lie on the same researcher's indifference curve.

Clearly, the opposite picture emerges when noise has an exponential distribution. In this case the information constraints β_F and β_{F^k} are arranged in the opposite order. With a Gumbel signal, selection has no effect on the error rates optimally chosen by the evaluator, so selection has zero impact on both the evaluator's and researcher's payoff.

When selection does not globally benefit (or hurt) the evaluator, the equilibrium impact of selection on the researcher's welfare is more subtle, but Proposition 7 again applies. Consider first the case of a uniform distribution. Recall that in this case—for a fixed value of p—the evaluator is worse off with F^k than with F for small values of R, but better off for large values of R (cf. Figure 6). As a consequence, the researcher's welfare changes sign twice. If R is either sufficiently small or sufficiently large, the researcher benefits from selection. However, at intermediate values of R the researcher loses from selection. Corresponding to the crossing between the information constraints there is a value of R where the evaluator is indifferent (Lemma 1 in Appendix A), and

the second claim in Proposition 7 then implies that the researcher loses from selection at this point.

Next, recall the case of the Laplace distribution. Symmetrically to the case of a uniform distribution, 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 harms the researcher for small or large values of R, but benefits for intermediate values.

In Section 4, we explored the possibility that the evaluator's choice converges to the bliss point ($\alpha = \beta = 0$) in the limit as k tends to infinity. Such a property makes it easy to check whether the researcher gains from extreme selection. In the limit, the researcher's payoff is p, since the evaluator correctly accepts in state θ_H. If, for lower k, the evaluator makes a choice satisfying $(1-p)\,\alpha > p\beta$, then the researcher is better off at k than when selection grows infinitely. This occurs when the evaluator has low reservation utility R and hence accepts often.

Conclusion. Summing up, the following intuitive conclusion emerges. Suppose that R is high, so the evaluator initially chooses α relatively low. The evaluator and researcher then have alignment of preferences over selection. If selection is good for the evaluator, as with normal or uniform noise, the researcher also stands to gain from selection. With exponential or Laplace noise, instead, both evaluator and researcher lose from selection—in this case selection causes a credibility crisis, which both parties would rather eliminate. When instead R is low, evaluator and researcher have conflicting preferences over selection.

5.2 Data Production and Selection Game

It is natural to endogenize the number k of subjects among which the researcher can select. We augment the game with a stage that takes place before stages 1 and 2:

0. The researcher privately chooses $k \geq 1$.

Note that this procedure is chosen before any realizations of ε_i are observed at the next stage. We assume that there is no credible way to directly reveal any information about the true k. The choice directly affects the researcher's payoff through a cost $C(k)$ 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 of this larger game, where the evaluator correctly anticipates the value of k optimally chosen by the researcher. In the subgame, the evaluator takes k as given and best responds by accepting when the signal realization x satisfies (10). In the first stage, the researcher correctly anticipates the evaluator's threshold $\bar{x}$, and chooses k in order to maximize the payoff

$$p\left(1 - F^k(\bar{x} - \theta_H)\right) + (1-p)\left(1 - F^k(\bar{x} - \theta_L)\right) - C(k). \tag{12}$$

Considering a deviation from equilibrium, the researcher 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.

Proposition 8. *The researcher's objective function is concave in k. Thus, any equilibrium* $(\bar{x},k)$ *solves the pair of equations* (10) *and*

$$-p\log\left(F\left(\bar{x}-\theta_H\right)\right)F^k\left(\bar{x}-\theta_H\right)-(1-p)\log\left(F\left(\bar{x}-\theta_L\right)\right)F^k\left(\bar{x}-\theta_L\right)=C'\left(k\right). \tag{13}$$

Note that, 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 (10) for the evaluator's best response $\bar{x}_k$. Then plug in $\bar{x}_k$ and k on the left hand side of (13) to determine the requisite $C'\left(k\right)$. Then choose the number $\gamma>0$ such that $C'\left(k\right)=2\gamma k$, and use the quadratic cost function $C\left(k\right)=\gamma k^2$. This observation provides a foundation for our approach so far where the number k was taken for given.

Researcher's Rat Race. The equilibrium described in Proposition 8 exhibits a rat race effect: when the evaluator correctly anticipates a greater degree k of selection, the researcher's cost $C\left(k\right)$ to manipulate the experiment is largely wasted. 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 $C\left(k\right)$, is then smaller when these costs are greater. As follows from our first remark, any k is consistent with equilibrium. 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, the endogenous costs will reduce this gain, perhaps to a loss.

Evaluator's Value of Commitment. The researcher's best response k may increase or decrease with the evaluator's standard $\bar{x}$, depending on parameters. The sign of this slope depends on the sign of the derivative of the left hand side in (13) with respect to $\bar{x}$,

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

which is positive when $F\left(\bar{x}-\theta_L\right)$ is sufficiently small, as happens when the prior strongly favors rejection. In that case, the best response k is an increasing function of $\bar{x}$. Conversely, when the prior strongly favors rejection, the best response k is a decreasing function of $\bar{x}$.[29]

[29] It can be easily verified that the best reply of the researcher is increasing for $\bar{x}<\theta_L+F^{-1}(e^{-1/k})$ and decreasing for $\bar{x}>\theta_H+F^{-1}(e^{-1/k})$.

Finally, we characterize the evaluator's optimal commitment to an ex-post suboptimal acceptance standard. Suppose R is large, so that the evaluator often rejects. The direction of commitment depends upon whether the evaluator stands to gain from more or less selection—this depends on F, as seen in Section 4. Given that then the researcher's best response is downward sloping, when the evaluator gains from greater k, it is optimal for the evaluator to commit to reduce the acceptance standard below the Nash level in order to induce the researcher to increase k. Conversely, when F is double logconcave at the top, the evaluator should optimally commit to a tougher acceptance standard, to reduce the extent of selection.

6 Extensions and Conclusion

Impact of Unanticipated Selection. So far, we considered situations in which the evaluator rationally predicts the correct extent of selection k and there is no uncertainty in k, 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. Here we relax these assumptions to consider more realistic scenarios.

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 accepting and rejecting, i.e. $\bar{\ell} = 1$.[30] Suppose that the noise distribution F 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 acceptance threshold that is optimal in the random experiment, namely $\bar{x}_F^* = \varepsilon_0 + (\theta_L + \theta_H)/2$, and consider how selection with $k = 2$ affects an unwary evaluator who maintains the acceptance standard unchanged at $\bar{x}_F^*$. The probability of acceptance clearly increases, resulting in a change in the evaluator's payoff—in the original problem, see (2) above—equal to

$$-(1-p)\underbrace{\left[F(\bar{x}_F^* - \theta_L) - F^2(\bar{x}_F^* - \theta_L)\right]}_{\text{increase in type I error}}(R - \theta_L) + p\underbrace{\left[F(\bar{x}_F^* - \theta_H) - F^2(\bar{x}_F^* - \theta_H)\right]}_{\text{reduction in type II error}}(\theta_H - R).$$

By equipoise, $(1-p)(R-\theta_L) = p(\theta_H - R)$. Thus, type I and type II errors are equally costly for the evaluator. By symmetry, $F(\bar{x}_F^* - \theta_L) + F(\bar{x}_F^* - \theta_H) = 1$. Thus, the increase in type I

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

error exactly equals the reduction in type II error. We conclude that the unwary evaluator, who anticipates no selection ($k = 1$), is indifferent between no selection and selection with $k = 2$.[31]

A fortiori, under symmetry and equipoise an increase in selection from $k = 1$ to $k = 2$ must necessarily benefit a rational evaluator.

Impact of Uncertain Selection. In a natural extension of the selection model, the number k is random. Again, information constraints can be generally used to compare experiments. However, the analysis in terms 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 $\lambda F^{k+1} + (1-\lambda) F^k$ 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.

Conclusion. Contrary to naive intuition, sample selection does not necessarily damage the evaluator. We characterize natural conditions under which the evaluator benefits. Increased selection benefits when the noise distribution is double logconvex—with a top tail thinner than the extreme value Gumbel distribution and the bottom tail thicker than Gumbel—a condition satisfied by normal noise, as well as a large set of exponential power distributions. For realistically unlikely propositions for which the researcher has an incentive to undertake experimentation in the first place, the top of the distribution matters, and sample selection is detrimental when the noise distribution is double logconcave—i.e., a relatively thick tail—at the top. Adding uncertainty or unawareness of selection adds further damage. 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.

At a methodological level, we also develop a generally applicable method for comparing the value of information structures depending on local dispersion. While in general selected data are not Blackwell comparable to random data, by constructing the convex envelope of the information constraints we characterize the welfare impact of selection on the basis of the local dispersion pattern of the conditional signal distribution and the parameters of the decision problem.

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 experi-

[31] As it is easy to verify, when the noise is normal for $\bar{\ell} \geqslant 1$ even an unwary evaluator (who wrongly anticipates random data, $k = 1$) benefits from observing selected data with $k = 2$.

mental subjects' unobserved actions. Also, given the work by Allcott (2015) on site selection bias, another open question is a general characterization of the impact of selection challenging external validity in the presence of heterogeneous treatment effects.

A Equivalent Comparison of Information Constraints

Lemmas 1 and 2 below spell out the technical details needed to establish the equivalence between conditions (A) and (L) in Theorem 1.

From Information to Preference. Lemma 1 is the building block in the construction of the evaluator's preference pattern from the sign pattern of the information constraint difference.

Lemma 1. *Let $0 \leqslant \alpha_0 < \alpha_1 \leqslant 1$. Let F and G be experiments such that $\beta_F(\alpha_1) = \beta_G(\alpha_1)$ and $\beta_F(\alpha) \leqslant \beta_G(\alpha)$ for all $\alpha \in [\alpha_0, \alpha_1]$. There exists $\ell_1 \in \left[-\beta'_F(\alpha_1), -\beta'_G(\alpha_1)\right]$ such that the evaluator prefers G to F for $\bar{\ell} \in \left[-\beta'_F(\alpha_1), \ell_1\right]$ and F to G for $\bar{\ell} \in \left[\ell_1, -\beta'_G(\alpha_0)\right]$.*

To understand the construction, consider again Figure 4. The points α_0 and α_1 in the figure satisfy the assumptions of Lemma 1, with β_F lying below β_G between α_0 and α_1 and then crossing β_G at α_1. Note that the type I and type II errors at the crossing point are exactly the ones determined by the switch in local dispersion at u_δ illustrated in Figure 3, for $\delta = \theta_H - \theta_L$.

First observe that, by the assumptions of the lemma, β_F cannot be steeper than β_G at α_1. Thus, the interval $\left[-\beta'_F(\alpha_1), -\beta'_G(\alpha_1)\right]$ is nonempty. At the left endpoint of the interval, $\bar{\ell} = -\beta'_F(\alpha_1)$, the evaluator prefers G to F. Indeed, by suboptimally choosing α_1 in experiment G, the evaluator gets as much as by optimally choosing α_1 under F, as illustrated by the solid black indifference line. Analogously, at the right endpoint of the interval, $\bar{\ell} = -\beta'_G(\alpha_1)$, and in fact all the way up to $\bar{\ell} = -\beta'_G(\alpha_0)$, by convexity of β_G the evaluator prefers F to G. For $\bar{\ell} = -\beta'_G(\alpha_1)$ this preference is illustrated by the dashed and dotted indifference lines.

Since the evaluator prefers G to F for $\bar{\ell} = -\beta'_F(\alpha_1)$ and F to G for $\bar{\ell} = -\beta'_G(\alpha_1)$, continuity of the problem's value guarantees existence of some ℓ_1 between $-\beta'_F(\alpha_1)$ and $-\beta'_G(\alpha_1)$, where the evaluator is indifferent between F and G. This indifference is illustrated by the green line in Figure 4. Starting from $\bar{\ell} = \ell_1$ and increasing $\bar{\ell}$ toward $-\beta'_G(\alpha_1)$, the indifference line going through the point $(\alpha^*_F(\ell_1), \beta_F(\alpha^*_F(\ell_1)))$ separates this point from the curve β_G (see clockwise arrow in the figure). This means that suboptimally choosing $\alpha^*_F(\ell_1)$ in experiment F gives the evaluator a higher payoff than choosing any α in experiment G. Thus, the evaluator prefers F to G for $\bar{\ell}$ between ℓ_1 and $-\beta'_G(\alpha_1)$. By analogous argument, when we decrease $\bar{\ell}$ from ℓ_1 toward $-\beta'_F(\alpha_1)$, as indicated by the counterclockwise arrow, the indifference line going through the point $(\alpha^*_G(\ell_1), \beta_G(\alpha^*_G(\ell_1)))$ separates this point from the curve β_F. Thus, the evaluator prefers G to F for $\bar{\ell}$ between $-\beta'_F(\alpha_1)$ and ℓ_1.

Finally, observe that in our example $\alpha_0 = 0$ also satisfies the hypothesis in Lemma 1. Indeed, F is the preferred experiment whenever $\bar{\ell} \geqslant \ell_1$.[32] Furthermore, $\beta_G(1) = \beta_F(1)$ and $\beta_G(\alpha) \leqslant \beta_F(\alpha)$ for all $\alpha \in [\alpha_1, 1]$. Thus, applying Lemma 1 again, the evaluator prefers G to F whenever $\bar{\ell} \leqslant -\beta_F'(\alpha_1)$. Putting our conclusions together, we have thus covered the whole set of possible values of the acceptance hurdle. The evaluator prefers G to F for $\bar{\ell} \in [0, \ell_1]$ and F to G for $\bar{\ell} \in [\ell_1, \infty]$.

From Preference to Information. Lemma 2 allows the reverse construction—inferring the crossing pattern of information constraints from the evaluator's preference pattern.

Lemma 2. *Let $0 \leqslant \ell_0 < \ell_1 \leqslant \infty$. Let F and G be experiments such that the evaluator prefers G to F for $\bar{\ell} \in [\ell_0, \ell_1]$ and is indifferent between F and G for $\bar{\ell} = \ell_1$. There exists $\alpha_1 \in [\alpha_F^*(\ell_1), \alpha_G^*(\ell_1)]$ such that $\beta_G(\alpha) \geqslant \beta_F(\alpha)$ for all $\alpha \in [\alpha_F^*(\ell_1), \alpha_1]$ and $\beta_G(\alpha) \leqslant \beta_F(\alpha)$ for all $\alpha \in [\alpha_1, \alpha_G^*(\ell_0)]$.*

To illustrate Lemma 2 we use again Figure 4, arguing that the information constraints of any two experiments F and G (not necessarily Gumbel's distributions) such that the evaluator prefers G to F for $\bar{\ell} \leqslant \ell_1$ and F to G for $\bar{\ell} \geqslant \ell_1$ must exhibit the single-crossing pattern in the figure.

Let $\ell_0 = 0$ and let ℓ_1 be as in Figure 4. The first observation to make is that $\alpha_F^*(\ell_1)$ cannot be greater than $\alpha_G^*(\ell_1)$, for otherwise the evaluator would strictly prefer F to G for $\bar{\ell}$ slightly below ℓ_1, contradicting the hypothesis in the lemma—-note that $\alpha_F^*(\ell_1) < \alpha_G^*(\ell_1)$ in the example depicted in Figure 4. Now consider the difference $\beta_G(\alpha) - \beta_F(\alpha)$, which is decreasing in the nonempty interval $[\alpha_F^*(\ell_1), \alpha_G^*(\ell_1)]$. At the left endpoint of this interval the difference is nonnegative, because for $\bar{\ell} = \ell_1$ the evaluator is indifferent between F and G—note that in the figure we have $\beta_G(\alpha_F^*(\ell_1)) > \beta_F(\alpha_F^*(\ell_1))$. At the right endpoint of the interval, and in fact all the way up to $\alpha_G^*(\ell_0) = 1$, the difference is nonpositive, because for $\bar{\ell}$ between ℓ_0 and ℓ_1 the evaluator prefers G to F—in the figure, β_G lies below β_F between $\alpha_G^*(\ell_1)$ and 1. The crossing point α_1 predicted by Lemma 2 then exists by continuity of the functions β_F and β_G.

The argument in the previous paragraph tells us that the difference $\beta_G(\alpha) - \beta_F(\alpha)$ is nonpositive for $\alpha \in [\alpha_1, 1]$ and nonnegative for $\alpha \in [\alpha_F^*(\ell_1), \alpha_1]$. Swapping the roles of F and G in Lemma 2 and applying the result again, this time to the interval $[\ell_1, \infty]$, we conclude that β_G must be above β_F also for $\alpha \leqslant \alpha_F^*(\ell_1)$, thus completing the picture.

B Proofs

Proof of Proposition 1. Let $\varepsilon' = F^{-1}(u_1)$ and fix any $0 < \delta < \overline{\varepsilon} - \underline{\varepsilon}$. Since $G^{-1}(u) - F^{-1}(u)$ is decreasing on $[0, u_1]$, we have $G^{-1}(F(\varepsilon + \delta)) - G^{-1}(F(\varepsilon)) \leqslant \delta$ for every $\varepsilon \in [\underline{\varepsilon}, \max\langle \underline{\varepsilon}, \varepsilon' - \delta \rangle]$. Similarly, since $G^{-1}(u) - F^{-1}(u)$ is increasing on $[u_1, 1]$, we have $G^{-1}(F(\varepsilon + \delta)) - G^{-1}(F(\varepsilon)) \geqslant \delta$ for every $\varepsilon \in [\min\{\varepsilon', \overline{\varepsilon} - \delta\}, \overline{\varepsilon} - \delta]$. Finally, again because $G^{-1}(u) - F^{-1}(u)$ is decreasing

[32] Recalling our definition of the slope of an information constraint at a non-interior point, we have $-\beta_G'(0) = \infty$.

on $[0,u_1]$ and increasing on $[u_1,1]$, the difference $G^{-1}(F(\varepsilon+\delta))-G^{-1}(F(\varepsilon)$ is increasing in the interval $[\max\langle\underline{\varepsilon},\varepsilon'-\delta\rangle,\min\langle\varepsilon',\overline{\varepsilon}-\delta\rangle]$. Thus, there exists ε_δ in this interval, such that $G^{-1}(F(\varepsilon+\delta))-G^{-1}(F(\varepsilon))$ is smaller than δ for $\varepsilon\in[\underline{\varepsilon},\varepsilon_\delta]$ and larger than δ for $\varepsilon\in[\varepsilon_\delta,\overline{\varepsilon}-\delta]$. Letting $u_\delta=F(\varepsilon_\delta)$, this is equivalent to $F(F^{-1}(u)+\delta)$ being smaller than $G(G^{-1}(u)+\delta)$ for $u\in[0,u_\delta]$ and greater than $G(G^{-1}(u)+\delta)$ for $u\in[u_\delta,1]$.

Proof of Lemma 1. First we show that the evaluator prefers F to G for $\bar\ell\in\big[-\beta'_G(\alpha_1),-\beta'_G(\alpha_0)\big]$. Since β_G is convex and $-\beta'_G(\alpha)<\bar\ell$ for $\alpha>\alpha^*_G(\bar\ell)$, we have $\alpha^*_G(\bar\ell)\geqslant\alpha_0$. Suppose first that $\alpha^*_G(\bar\ell)>\alpha_1$. Then by convexity of β_G we have $-\beta'_G(\alpha_1)=\bar\ell$ and hence $\beta_G(\alpha_1)-\beta_G(\alpha^*_G(\bar\ell))=\bar\ell\big[\alpha^*_G(\bar\ell)-\alpha_1\big]$. Since $\beta_F(\alpha_1)\leqslant\beta_G(\alpha_1)$, we thus have

$$\bar\ell\alpha_1+\beta_F(\alpha_1)\leqslant\bar\ell\alpha_1+\beta_G(\alpha_1)=\bar\ell\alpha^*_G(\bar\ell)+\beta_G(\alpha^*_G(\bar\ell)),$$

which proves that the evaluator prefers F to G. Next, suppose that $\alpha_0\leqslant\alpha^*_G(\bar\ell)\leqslant\alpha_1$. Then $\beta_F(\alpha^*_G(\bar\ell))\leqslant\beta_G(\alpha^*_G(\bar\ell))$ and hence $\bar\ell\alpha^*_G(\bar\ell)+\beta_F(\alpha^*_G(\bar\ell))\leqslant\bar\ell\alpha^*_G(\bar\ell)+\beta_G(\alpha^*_G(\bar\ell))$. Thus, also in this case the evaluator prefers F to G.

Next, we show that there exists $\ell_1\in\big[-\beta'_F(\alpha_1),-\beta'_G(\alpha_1)\big]$ such that the evaluator is indifferent between F and G for $\bar\ell=\ell_1$. First note that $-\beta'_F(\alpha_1)\leqslant-\beta'_G(\alpha_1)$, for $\beta_F(\alpha_1)=\beta_G(\alpha_1)$ and $\beta_F(\alpha)\leqslant\beta_G(\alpha)$ for every α in the nonempty interval $[\alpha_0,\alpha_1]$. Thus, the interval $\big[-\beta'_F(\alpha_1),-\beta'_G(\alpha_1)\big]$ is nonempty. Now observe that, since β_F is convex, $\beta_F(\alpha)-\beta_F(\alpha_1)\geqslant\beta'_F(\alpha_1)(\alpha-\alpha_1)$ for all $\alpha\in[0,1]$. Thus, $-\beta'_F(\alpha_1)\alpha_1+\beta_G(\alpha_1)=-\beta'_F(\alpha_1)\alpha_1+\beta_F(\alpha_1)\leqslant-\beta'_F(\alpha_1)\alpha+\beta_F(\alpha)$ for all $\alpha\in[0,1]$, so the evaluator prefers G to F for $\bar\ell=-\beta'_F(\alpha_1)$. Similarly, by convexity of β_G the evaluator prefers F to G for $\bar\ell=-\beta'_G(\alpha_1)$. The value of the evaluator's problem is continuous in $\bar\ell$ and the interval $\big[-\beta'_F(\alpha_1),-\beta'_G(\alpha_1)\big]$ is nonempty, so there exists ℓ_1 in this interval, such that the evaluator is indifferent between F and G for $\bar\ell=\ell_1$.

It remains to be shown that the evaluator prefers G to F for $\bar\ell\in\big[-\beta'_F(\alpha_1),\ell_1\big]$ and F to G for $\bar\ell\in\big[\ell_1,-\beta'_G(\alpha_1)\big]$. First note that, since the evaluator is indifferent between F and G for $\bar\ell=\ell_1$,

$$\ell_1\alpha+\beta_F(\alpha)\geqslant\ell_1\alpha^*_G(\ell_1)+\beta_G(\alpha^*_G(\ell_1))\qquad\forall\alpha\in[0,1],\tag{14}$$

$$\ell_1\alpha+\beta_G(\alpha)\geqslant\ell_1\alpha^*_F(\ell_1)+\beta_F(\alpha^*_F(\ell_1))\qquad\forall\alpha\in[0,1].\tag{15}$$

Next, note that $\ell_1\leqslant-\beta'_G(\alpha_1)$ and $\bar\ell\leqslant-\beta'_G(\alpha_1)$ imply $\alpha^*_G(\ell_1)\geqslant\alpha^*_G(-\beta'_G(\alpha_1))\geqslant\alpha_1$ and $\alpha^*_G(\bar\ell)\geqslant\alpha^*_G(-\beta'_G(\alpha_1))\geqslant\alpha_1$. Moreover, since $\ell_1\geqslant-\beta'_F(\alpha_1)$ and $\bar\ell\geqslant-\beta'_F(\alpha_1)$, there exist $\tilde\alpha_1\leqslant\alpha_1$ and $\hat\alpha_1\leqslant\alpha_1$ such that $\ell_1\tilde\alpha_1+\beta_F(\tilde\alpha_1)=\ell_1\alpha^*_F(\ell_1)+\beta_F(\alpha^*_F(\ell_1))$ and $\bar\ell\hat\alpha_1+\beta_F(\hat\alpha_1)=\bar\ell\alpha^*_F(\bar\ell)+\beta_F(\alpha^*_F(\bar\ell))$. Now, if $\bar\ell\leqslant\ell_1$ then by (14), using the fact that $\alpha^*_G(\ell_1)\geqslant\alpha_1\geqslant\hat\alpha_1$, we obtain

$$\bar\ell\alpha^*_F(\bar\ell)+\beta_F(\alpha^*_F(\bar\ell))=\bar\ell\hat\alpha_1+\beta_F(\hat\alpha_1)\geqslant\bar\ell\alpha^*_G(\ell_1)+\beta_G(\alpha^*_G(\ell_1)).$$

This proves that the evaluator prefers G to F for $\bar\ell\in\big[-\beta'_F(\alpha_1),\ell_1\big]$. Suppose now that $\bar\ell\geqslant\ell_1$. Then by (15), using the fact that $\alpha^*_G(\bar\ell)\geqslant\alpha_1\geqslant\tilde\alpha_1$, we obtain

$$\bar\ell\alpha^*_G(\bar\ell)+\beta_G(\alpha^*_G(\bar\ell))\geqslant\bar\ell\tilde\alpha_1+\beta_F(\tilde\alpha_1).$$

This proves that the evaluator prefers F to G for $\bar{\ell} \in \left[\ell_1, -\beta'_G(\alpha_1)\right]$.

Proof of Lemma 2. First we show that $\beta_G(\alpha) \leqslant \beta_F(\alpha)$ for all $\alpha \in [\alpha^*_G(\ell_1), \alpha^*_G(\ell_0)]$. This follows from the fact that $\beta_G(\alpha) \leqslant \beta_F(\alpha)$ for all $\alpha \in [0,1]$ such that $\ell_0 \leqslant -\beta'_G(\alpha) \leqslant \ell_1$, which we now prove. By convexity of β_G we have $\beta'_G(\alpha)(\alpha' - \alpha) \leqslant \beta_G(\alpha') - \beta_G(\alpha)$ for all $\alpha' \in [0,1]$. Assume by way of contradiction that $\beta_G(\alpha) > \beta_F(\alpha)$. Then $-\beta'_G(\alpha)\alpha + \beta_F(\alpha) < -\beta'_G(\alpha)\alpha' + \beta_G(\alpha')$ for all $\alpha' \in [0,1]$, that is, the evaluator strictly prefers F to G for $\bar{\ell} = -\beta'_G(\alpha)$. This is impossible, because the evaluator prefers G to F for $\bar{\ell} \in [\ell_0, \ell_1]$.

To conclude the proof, we show that there exists $\alpha_1 \in \left[\alpha^*_F(\ell_1), \alpha^*_G(\ell_1)\right]$ such that $\beta_G(\alpha) \geqslant \beta_F(\alpha)$ for all $\alpha \in \left[\alpha^*_F(\ell_1), \alpha_1\right]$ and $\beta_G(\alpha) \leqslant \beta_F(\alpha)$ for all $\alpha \in \left[\alpha_1, \alpha^*_G(\ell_1)\right]$. First we show that the interval $\left[\alpha^*_F(\ell_1), \alpha^*_G(\ell_1)\right]$ is nonempty. By way of contradiction, suppose that $\alpha^*_F(\ell_1) > \alpha^*_G(\ell_1)$, and let $\Delta = \beta_G(\alpha^*_F(\ell_1)) - \beta_G(\alpha^*_G(\ell_1)) + \ell_1 \left[\alpha^*_F(\ell_1) - \alpha^*_G(\ell_1)\right]$. Since β_G is convex and $-\beta'_G(\alpha) < \ell_1$ for all $\alpha > \alpha^*_G(\ell_1)$, we must have $\Delta > 0$. By optimality of $\alpha^*_G(\ell_1)$,

$$\beta_G(\alpha^*_G(\ell_1)) - \beta_G(\alpha) \leqslant \ell_1 \left[\alpha - \alpha^*_G(\ell_1)\right] \qquad \forall \alpha < \alpha^*_F(\ell_1),$$

and moreover, again by convexity of β_G,

$$\beta_G(\alpha^*_G(\ell_1)) - \beta_G(\alpha) \leqslant \ell_1 \left[\alpha - \alpha^*_G(\ell_1)\right] - \Delta \qquad \forall \alpha \geqslant \alpha^*_F(\ell_1).$$

Since the evaluator is indifferent between F and G for $\bar{\ell} = \ell_1$, the above inequalities imply

$$\begin{aligned} \beta_F(\alpha^*_F(\ell_1)) - \beta_G(\alpha) &\leqslant \ell_1 \left[\alpha - \alpha^*_F(\ell_1)\right] && \forall \alpha < \alpha^*_F(\ell_1), \\ \beta_F(\alpha^*_F(\ell_1)) - \beta_G(\alpha) &\leqslant \ell_1 \left[\alpha - \alpha^*_F(\ell_1)\right] - \Delta && \forall \alpha \geqslant \alpha^*_F(\ell_1). \end{aligned}$$

But $\Delta > 0$, hence for all $\delta > 0$ sufficiently small we have

$$\beta_F(\alpha^*_F(\ell_1)) - \beta_G(\alpha) < (\ell_1 - \delta) \left[\alpha - \alpha^*_F(\ell_1)\right] \qquad \forall \alpha \in [0,1],$$

which contradicts the assumption that the evaluator prefers G to F for all $\bar{\ell} \in [\ell_0, \ell_1]$. This proves that the interval $\left[\alpha^*_F(\ell_1), \alpha^*_G(\ell_0)\right]$ is nonempty. But the difference $\beta_G(\alpha) - \beta_F(\alpha)$ is decreasing on this interval, because for every $\alpha \in \left[\alpha^*_F(\ell_1), \alpha^*_G(\ell_1)\right]$ we have $-\beta'_G(\alpha) \geqslant \ell_1$ and $-\beta'_F(\alpha) \leqslant \ell_1$. Moreover, since the evaluator is indifferent between F and G for $\bar{\ell} = \ell_1$, we must have $\beta_G(\alpha) - \beta_F(\alpha) \geqslant 0$ for $\alpha = \alpha^*_F(\ell_1)$ and $\beta_G(\alpha) - \beta_F(\alpha) \leqslant 0$ for $\alpha = \alpha^*_G(\ell_1)$. Thus, by continuity of β_F and β_G there exists $\alpha_1 \in \left[\alpha^*_F(\ell_1), \alpha^*_G(\ell_1)\right]$ such that $\beta_G(\alpha) \geqslant \beta_F(\alpha)$ for all $\alpha \in \left[\alpha^*_F(\ell_1), \alpha_1\right]$ and $\beta_G(\alpha) \leqslant \beta_F(\alpha)$ for all $\alpha \in \left[\alpha_1, \alpha^*_G(\ell_1)\right]$.

Proof of Theorem 1. Changing variable from u to $\alpha = 1 - F(F^{-1}(u) + \theta_H - \theta_L)$, condition (D) becomes condition (A). We now show that (A) is equivalent to (L). If (A) holds then, by Lemma 1, there exist $0 \leqslant \ell_2 \leqslant \ldots \leqslant \ell_{2N} \leqslant \infty$ such that the evaluator prefers F to G for $\bar{\ell} \in [0, \ell_2] \cup \cdots \cup [\ell_{2N-1}, \ell_{2N}]$ and G to F for $\bar{\ell} \in [\ell_2, \ell_3] \cup \cdots \cup [\ell_{2N}, \ell_{2N+1}]$, i.e. (L) holds. Conversely, if (L) holds

then, by Lemma 2, there exist $1 \geqslant \alpha_2 \geqslant \cdots \geqslant \alpha_{2N} \geqslant 0$ such that $\beta_F(\alpha)$ is below $\beta_G(\alpha)$ for $\alpha \in [\alpha_{2N}, \alpha_{2N-1}] \cup \cdots \cup [\alpha_2, 1]$ and above $\beta_G(\alpha)$ for $\alpha \in [0, \alpha_{2N}] \cup \cdots \cup [\alpha_3, \alpha_2]$, i.e. (A) holds. Thus, (A), (L) and (D) are all equivalent.

Proof of Theorem 2. We provide first a direct argument and then one based on dispersion, i.e. using Lehmann's (1988) Theorem 5.2 (or our Theorem 1).

Given any $k \geqslant 1$, at the optimal threshold $\bar{x} = \bar{x}^*_{F^k}(\bar{\ell})$ the following first-order condition holds:

$$\frac{d\beta}{d\bar{x}} = kF^{k-1}(\bar{x}-\theta_H) f(\bar{x}-\theta_H) = \bar{\ell} kF^{k-1}(\bar{x}-\theta_L) f(\bar{x}-\theta_L) = -\bar{\ell}\frac{d\alpha}{d\bar{x}}. \tag{16}$$

As k varies, the changes in β and α may be derived as

$$\frac{d\beta}{dk} = \log(F(\bar{x}-\theta_H)) F^k(\bar{x}-\theta_H) + kF^{k-1}(\bar{x}-\theta_H) f(\bar{x}-\theta_H)\frac{d\bar{x}}{dk}, \tag{17}$$

$$\frac{d\alpha}{dk} = -\log(F(\bar{x}-\theta_L)) F^k(\bar{x}-\theta_L) - kF^{k-1}(\bar{x}-\theta_L) f(\bar{x}-\theta_L)\frac{d\bar{x}}{dk}. \tag{18}$$

Thus, the evaluator gains as k increases if and only if

$$0 > \frac{d\beta}{dk} + \bar{\ell}\frac{d\alpha}{dk} = \log(F(\hat{x}-\theta_H)) F^k(\hat{x}-\theta_H) - \bar{\ell}\log(F(\hat{x}-\theta_L)) F^k(\hat{x}-\theta_L),$$

where (16) was used to reduce the expression. Using (16) once more, we can rewrite the latter inequality as

$$\log(F(\hat{x}-\theta_H))\frac{F(\hat{x}-\theta_H)}{f(\hat{x}-\theta_H)} < \log(F(\hat{x}-\theta_L))\frac{F(\hat{x}-\theta_L)}{f(\hat{x}-\theta_L)}. \tag{19}$$

Since the derivative of $-\log(-\log F)$ is $-f/(F\log F) > 0$, and $\hat{x}-\theta_H < \hat{x}-\theta_L$, the evaluator gains as k increases if and only if the derivative of $-\log(-\log F)$ is increasing, i.e. if and only if $-\log(-\log F)$ is a convex function.

For the alternative argument, let $\varphi(\varepsilon) = -\log(-\log(F(\varepsilon)))$ for brevity, and observe that for all ε and $k \geqslant 1$ the horizontal distance between F^k and F, namely $(F^k)^{-1}(F(\varepsilon)) - \varepsilon$, is the same as the horizontal distance between the double-log transformations of F^k and F, namely $\varphi^{-1}(\varphi(\varepsilon) + \log(k)) - \varepsilon$. The derivative of the latter distance is

$$\frac{\varphi'(\varepsilon)}{\varphi'(\varphi^{-1}(\varphi(\varepsilon)+\log(k)))} - 1, \tag{20}$$

which is negative (resp. positive) for every ε and $k \geqslant 1$ if and only if φ is a convex (resp. concave) function. Thus, F^k is less dispersed than F if and only if φ is a convex function. The result now follows from Lehmann's (1988) Theorem 5.2.

Proof of Proposition 2. Immediate from Theorem 2 and Lehmann's (1988) Theorem 5.2.

Proof of Proposition 3. By Proposition 1, it is enough to show that if φ is first concave and then convex, then the expression in (20) is first positive and then negative. (The proof that, if φ

is first convex and then concave, then the expression in (20) is first negative and then positive, is analogous.) Thus, fix any ε_1 such that $\varphi(\varepsilon)$ is concave for $\varepsilon \leqslant \varepsilon_1$ and convex for $\varepsilon \geqslant \varepsilon_1$. Let $\varepsilon_0 = \varphi^{-1}(\varphi(\varepsilon_1) - \log(k))$. By the same argument used in the proof of Theorem 2, the horizontal difference $\varphi^{-1}(\varphi(\varepsilon) - \log(k)) - \varepsilon$ is increasing for $\varepsilon \leqslant \varepsilon_0$ and decreasing for $\varepsilon \geqslant \varepsilon_1$. For ε in the interval $[\varepsilon_0, \varepsilon_1]$ the slope $\varphi'(\varepsilon)$ is increasing while the slope $\varphi'(\varphi^{-1}(\varphi(\varepsilon) + \log(k)))$ is decreasing. Thus, in the interval $[\varepsilon_0, \varepsilon_1]$ the horizontal difference $\varphi^{-1}(\varphi(\varepsilon) - \log(k)) - \varepsilon$ is first increasing and then decreasing.

Proof of Proposition 4. Fix any $\delta \in (0, (\theta_H - \theta_L)/2)$. Let $\varepsilon_\delta > 0$ be such that $\bar{F}(\varepsilon_\delta) - \bar{F}(-\varepsilon_\delta) \geqslant 1 - \delta/2$. Choose $\hat{k}$ so that for all $k \geqslant \hat{k}$,

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

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

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

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 $x < \bar{x} = \theta_H + b_k - \delta$ gives

$$\begin{aligned} \alpha &= \Pr\left(x \geqslant \theta_H + b_k - \delta \;\middle|\; \theta_L\right) \leqslant 1 - \Pr\left(\theta_L + b_k - \delta \leqslant x \leqslant \theta_L + b_k + \delta \;\middle|\; \theta_L\right) \leqslant \delta \\ \beta &= \Pr\left(x < \theta_H + b_k - \delta \;\middle|\; \theta_H\right) \leqslant 1 - \Pr\left(\theta_H + b_k - \delta \leqslant x \leqslant \theta_H + b_k + \delta \;\middle|\; \theta_H\right) \leqslant \delta. \end{aligned}$$

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

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

Proof of Proposition 5. We show that if $\varepsilon_1, \ldots, \varepsilon_k$ are i.i.d. exponential power with shape $s > 1$, location and scale parameters 0 and 1, then $M_k = \max\langle \varepsilon_1, \ldots, \varepsilon_k \rangle$ satisfies $\Pr(M_k \leqslant a_k \varepsilon + b_k) \to e^{-e^{-\varepsilon}}$ for all ε, where

$$a_k = (s \log k)^{-\frac{s-1}{s}} \qquad \text{and} \qquad b_k = (s \log k)^{1/s} - \frac{\frac{s-1}{s} \log\log k + \log(2\Gamma[1/s])}{(s \log k)^{\frac{s-1}{s}}}.$$

Start by noticing that $f(\varepsilon)/\left(\varepsilon^{b-1}[1-F(\varepsilon)]\right) \to 1$ as $\varepsilon \to \infty$. Fix ε and define y_k for each $k \geqslant 1$ by $1-F(y_k) = e^{-\varepsilon}/k$, so that

$$\frac{e^{-\varepsilon}}{k}\frac{y_k^{s-1}}{f(y_k)} \to 1 \quad \text{as } k \to \infty. \tag{21}$$

We may assume $y_k > 0$ for all k. Then $f(y_k) = s^{\frac{s-1}{s}} e^{-y_k^s/s}/2\Gamma[1/s]$ and hence, by (21),

$$-\log k - \varepsilon + (s-1)\log y_k - \frac{s-1}{s}\log s + \log\left(2\Gamma[1/b]\right) + \frac{y_k^s}{s} \to 0. \tag{22}$$

From (22) we see that $-\log k + (s-1)\log y_k + y_k^s/s = -\log k + o\left(y_k^s/s\right) + y_k^s/s$ converges to a constant. Thus, $-s\log k/u_k^s + o\left(y_k^s/s\right)/(u_k^s/s) + 1$ converges to 0, i.e. $\log y_k = (1/s)(\log s + \log\log k) + o(1)$. Using this fact in (22), we obtain

$$\begin{aligned}\frac{y_k^s}{s} &= \log k + \varepsilon - \frac{s-1}{s}(\log s + \log\log k) + \frac{s-1}{s}\log s - \log\left(2\Gamma[1/s]\right) + o(1) \\ &= \log k + \varepsilon - \frac{s-1}{s}\log\log k - \log\left(2\Gamma[1/s]\right) + o(1).\end{aligned}$$

Equivalently,

$$\begin{aligned}y_k &= (s\log k)^{1/s}\left[1 + \frac{\varepsilon - \frac{s-1}{s}\log\log k - \log\left(2\Gamma[1/s]\right)}{\log k} + o\left(\frac{1}{\log k}\right)\right]^{1/s} \\ &= (s\log k)^{1/s}\left[1 + \frac{\varepsilon - \frac{s-1}{s}\log\log k - \log\left(2\Gamma[1/s]\right)}{s\log k} + o\left(\frac{1}{\log k}\right)\right] \\ &= (s\log k)^{1/s} + \frac{\varepsilon - \frac{s-1}{s}\log\log k - \log\left(2\Gamma[1/s]\right)}{(s\log k)^{\frac{s-1}{b}}} + o\left(\frac{1}{(\log k)^{\frac{s-1}{s}}}\right) = a_k\varepsilon + b_k + o(a_k).\end{aligned}$$

Thus, $\Pr(M_k \leqslant a_k\varepsilon + b_k + o(a_k)) \to e^{-e^{-\varepsilon}}$, as was to be shown.

Proof of Proposition 6. We prove the stronger claim that the strategy profile described in the proposition is the unique Bayes Nash equilibrium where the researcher always selects the same order statistic—in equilibrium, the highest. Since the density function $f(x)$ is logconcave, both the cumulative distribution function $F(x)$ and the reliability function $1-F(x)$ are logconcave. Moreover, the product of logconcave functions is logconcave. Thus, the density function of the mth smallest of k such (i.i.d.) random variables,

$$\frac{k!}{(m-1)!(k-m)!}F^{m-1}(x)\left[1-F(x)\right]^{k-m}f(x),$$

is logconcave, which implies that the evaluator accepts if and only if the signal is at or above some threshold $\bar{x}$. But then the best response of the researcher is to select the individual with highest noise term ($m = k$) because its distribution first-order stochastically dominates that of any other order statistic.

Proof of Proposition 7. The researcher benefits from selection at R if and only if

$$p\left[1-\beta_{F^k}(\alpha^*_{F^k}(\ell(R)))\right]+(1-p)\alpha^*_{F^k}(\ell(R)) \geqslant p\left[1-\beta_F(\alpha^*_F(\ell(R)))\right]+(1-p)\alpha^*_F(\ell(R)),$$

i.e.

$$\beta_{F^k}(\alpha^*_{F^k}(\ell(R)))-\beta_F(\alpha^*_F(\ell(R))) \leqslant \frac{1-p}{p}\left[\alpha^*_{F^k}(\ell(R))-\alpha^*_F(\ell(R))\right].$$

By definition of φ, the latter inequality is equivalent to $\alpha^*_{F^k}(\ell(R)) \geqslant \varphi(\alpha^*_F(\ell(R)))$, which is in turn equivalent to $\ell(R) \leqslant \beta'_{F^k}(\varphi(\alpha^*_F(\ell(R))))$. The first claim in the proposition then follows.

To prove the second claim, assume that the evaluator prefers F to F^k for $\bar{\ell} \in [\ell_0, \ell_1]$. By Lemma 2 we have $\alpha^*_{F^k}(\ell_2) \leqslant \alpha^*_F(\ell_2)$ and hence, by the evaluator's indifference for $\bar{\ell} = \ell_2$, also $\beta_{F^k}(\alpha^*_{F^k}(\ell_2)) \geqslant \beta_F(\alpha^*_F(\ell_2))$. Thus, the researcher loses from selection at R such that $\ell(R) = \ell_2$. The proof for the case where the evaluator prefers F^k to F for $\bar{\ell} \in [\ell_0, \ell_1]$ is analogous.

Proof of Proposition 8. Since the cost function is convex, it suffices to check that the first two terms in (12) are concave in k. It suffices to take k to be 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.

References

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

BAGNOLI, M. AND T. BERGSTROM (2005): "Log-Concave Probability and Its Applications," *Economic Theory*, 26(2), 445–469.

BANERJEE, A. V., S. CHASSANG, AND E. SNOWBERG (2017): "Decision Theoretic Approaches to Experiment Design and External Validity," in *Handbook of Field Experiments, Volume 1*, ed. by A. V. Banerjee and E. Duflo, North Holland, 141–174.

BERGER, V. W. (2005): *Selection Bias and Covariate Imbalances in Randomized Clinical Trials*, Wiley.

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.

BROWN, G. W. AND J. W. TUKEY (1946): "Some Distributions of Sample Means," *Annals of Mathematical Statistics*, 17(1), 1–12.

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.

DI TILLIO, A., M. OTTAVIANI, AND P. N. SØRENSEN (2017): "Testing for Double Logconvexity and Double Logconcavity," Online Supplementary Appendix for 'Strategic Sample Selection', https://www.dropbox.com/s/719s7d3imcyc1hs/DataAppendix.pdf?dl=0.

——— (forthcoming): "Persuasion Bias in Science: Can Economics Help?" *Economic Journal*.

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

FELGENHAUER, M. AND E. SCHULTE (2014): "Strategic Private Experimentation," *American Economic Journal: Microeconomics*, 6(4), 74–105.

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.

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.

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," *The Economic Journal*, 119(539), 1036–1064.

HENRY, E. AND M. OTTAVIANI (2015): "Research and the Approval Process: The Organization of Persuasion," Bocconi mimeo.

HOFFMANN, F., R. INDERST, AND M. OTTAVIANI (2014): "Persuasion through Selective Disclosure: Implications for Marketing, Campaigning, and Privacy Regulation," Bocconi mimeo.

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

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

KAMENICA, E. AND M. GENTZKOW (2011): "Bayesian Persuasion," *American Economic Review*, 101(6), 2590–2615.

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, 521–533.

LEHMANN, E. L. AND J. P. ROMANO (2005): *Testing Statistical Hypotheses*, Springer.

MARSHALL, A. W. AND I. OLKIN (2007): *Life Distributions*, Springer.

MÜLLER, S. AND K. RUFIBACH (2008): "On the Max-Domain of Attraction of Distributions with Log-Concave Densities," *Statistics and Probability Letters*, 78, 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.

PARZEN, E. (2004): "Quantile Probability and Statistical Data Modeling," *Statistical Science*, 19(4), 652–662.

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, 1949–1992.

RAYO, L. AND I. SEGAL (2010): "Optimal Information Disclosure," *Journal of Political Economy*, 118(5), 949–987.

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

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

SCHULZ, K. F. (1995): "Subverting Randomization in Controlled Trials," *Journal of the American Medical Association*, 274(18), 1456–1458.

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

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

TORGERSEN, E. (1991): *Comparison of Statistical Experiments*, Cambridge University Press.

WALD, A. (1950): *Statistical Decision Functions*, Wiley.