From Blackwell Dominance in Large Samples to Rényi Divergences and Back Again*

Xiaosheng Mu† Luciano Pomatto‡ Philipp Strack§ Omer Tamuz¶

September 4, 2020

Abstract

We study repeated independent Blackwell experiments; standard examples include drawing multiple samples from a population, or performing a measurement in different locations. In the baseline setting of a binary state of nature, we compare experiments in terms of their informativeness in large samples. Addressing a question due to Blackwell (1951), we show that generically an experiment is more informative than another in large samples if and only if it has higher Rényi divergences.

We apply our analysis to the problem of measuring the degree of dissimilarity between distributions by means of divergences. A useful property of Rényi divergences is their additivity with respect to product distributions. Our characterization of Blackwell dominance in large samples implies that every additive divergence that satisfies the data processing inequality is an integral of Rényi divergences.

1 Introduction

Statistical experiments form a general framework for modeling information: Given a set Θ of parameters, an *experiment* P produces an observation distributed according to P_θ, given the true parameter value $\theta \in \Theta$. Blackwell's celebrated theorem (Blackwell, 1951) provides a partial order for comparing experiments in terms of their informativeness.

*We are grateful to the co-editor and three referees for their comments and suggestions. In addition we would like to thank Kim Border, Laura Doval, Federico Echenique, Tobias Fritz, Drew Fudenberg, George Mailath, Massimo Marinacci, Margaret Meyer, Marco Ottaviani and Peter Norman Sørensen for helpful discussions.

†Princeton University. Email: xmu@princeton.edu. Xiaosheng Mu acknowledges the hospitality of Columbia University and the Cowles Foundation at Yale University, which hosted him during parts of this research.

‡Caltech. Email: luciano@caltech.edu.

§Yale University. Email: philipp.strack@yale.edu.

¶Caltech. Email: tamuz@caltech.edu. Omer Tamuz was supported by a grant from the Simons Foundation (#419427), a Sloan research fellowship, and a BSF award (#2018397).

As is well known, requiring two experiments to be ranked in the Blackwell order is a demanding condition. Consider the problem of testing a binary hypothesis $\theta \in \{0, 1\}$, based on random samples drawn from one of two experiments P or Q. According to Blackwell's ordering, P is more informative than Q if, for every test performed based on observations produced by Q, there exists another test based on P that has lower probabilities of both Type-I and Type-II errors (Blackwell and Girshick, 1979). This is a difficult condition to satisfy, especially in the case where only one sample is produced by each experiment.

In many applications, an experiment does not consist of a single observation but of multiple i.i.d. samples. For example, a new vaccine is typically tested on multiple patients, and a randomized control trial assessing the effect of an intervention usually involves many subjects. We study a weakening of the Blackwell order that is appropriate for comparing experiments in terms of their large sample properties. Our starting point is the question, first posed by Blackwell (1951), of whether it is possible for n independent observations from an experiment P to be more informative than n observations from another experiment Q, even though P and Q are not comparable in the Blackwell order. The question was answered in the affirmative by Stein (1951), Torgersen (1970) and Azrieli (2014).[1] However, identifying the precise conditions under which this phenomenon occurs has remained an open problem.

We say that P dominates Q *in large samples* if for every n large enough, n independent observations from P are more informative, in the Blackwell order, than n independent observations from Q. We focus on a binary set of parameters Θ, and show that generically P dominates Q in large samples if and only if the experiment P has higher Rényi divergences than Q (Theorem 1). Rényi divergences are a one-parameter family of measures of informativeness for experiments; introduced and characterized axiomatically in Rényi (1961), we show that they capture the informativeness of an experiment in large samples. For any two experiments comparable in terms of Rényi divergences, we also provide a simple bound on the sample size that ensures that larger samples of independent experiments are comparable in the Blackwell order (Theorem 4).

The proof of this result crucially relies on two ingredients. First, we use techniques from large deviations theory to compare sums of i.i.d. random variables in terms of stochastic dominance. In addition, we provide and apply a new characterization of the Blackwell order: We associate to each experiment a new statistic, *the perfected log-likelihood ratio*, and show that the comparison of these statistics in terms of first-order stochastic dominance is in fact equivalent to the Blackwell order.

We apply our characterization of Blackwell dominance in large samples to the problem of quantifying the extent to which two probability distributions are dissimilar. This is a common problem in econometrics and statistics, where formal measures quantifying the

[1] Even though Stein (1951) is frequently cited in the literature for a first example of this type, we could not gain access to that paper.

difference between distributions are referred to as *divergences*.[2] Well known examples include total variation distance, the Hellinger distance, the Kullback-Leibler divergence, Rényi divergences, and more general f-divergences.

Rényi divergences satisfy two key properties. The first is *additivity*: Rényi divergences decompose into a sum when applied to pairs of product distributions. Additivity captures a principle of non-interaction across independent domains, as the total divergence of two unrelated pairs does not change when they are considered together as a bundle. Additivity is a natural property, and in applications it is a crucial simplification for studying i.i.d. processes. A second desirable property is described by the *data-processing inequality*, which stipulates that the distributions of two random variables X and Y are at least as dissimilar as those of $f(X)$ and $f(Y)$, for any transformation f. As we show, this property is closely related to monotonicity with respect to the Blackwell order.

Using our main result, we show that every additive divergence that satisfies the data-processing inequality and a mild finiteness condition is an integral (i.e., the limit of positive linear combinations) of Rényi divergences (Theorem 2). This result is an improvement over the original characterization of Rényi (1961), as well as more modern ones (Csiszár, 2008), because it shows that additivity alone pins down a single class of divergences without making any further assumptions on the functional form.

The study most closely related to ours is Moscarini and Smith (2002). In their order, an experiment P dominates another experiment Q if for for every finite decision problem, a large enough sample of observations from an experiment P will achieve higher expected payoff than a sample of the same size of observations from Q. In contrast to the order proposed by Blackwell and analyzed in this paper, their definition allows for the critical sample size to depend on the decision problem, and considers a restricted class of decision problems. We provide a detailed discussion of this and other related work in §6.

The paper is organized as follows. In §2 we provide our main definitions. §3 contains the characterization of Blackwell dominance in large samples, with proof deferred to §5. In §4 we characterize additive divergences. Finally, we further discuss our results and their relation to the literature in §6.

2 Model

2.1 Statistical Experiments

A state of the world θ can take two possible values, 0 or 1. A *Blackwell-Le Cam experiment* $P = (\Omega, P_0, P_1)$ consists of a sample space Ω, which we assume to be a Polish space,

[2] See, e.g., Sawa (1978); White (1982); Critchley et al. (1996); Kitamura and Stutzer (1997); Hong and White (2005); Ullah (2002). See Kitamura et al. (2013) for a recent application of α-divergences, which are a reformulation of Rényi divergences.

and a pair of Borel probability measures (P_0, P_1) defined over Ω, with the interpretation that $P_\theta(A)$ is the probability of observing $A \subseteq \Omega$ in state $\theta \in \{0,1\}$. This framework is commonly encountered in simple hypothesis tests as well as in information economics. In §6 we discuss the case of experiments for more than two states: we obtain necessary conditions for dominance in large samples and explain the obstacles to a full characterization.

Given two experiments $P = (\Omega, P_0, P_1)$ and $Q = (\Xi, Q_0, Q_1)$, we can form the *product experiment* $P \otimes Q$ given by

$$P \otimes Q = (\Omega \times \Xi, P_0 \times Q_0, P_1 \times Q_1).$$

where $P_\theta \times Q_\theta$, given $\theta \in \{0,1\}$, denotes the product of the two measures. Under the experiment $P \otimes Q$ the realizations produced by both P and Q are observed, and the two observations are independent (conditional on the true state). For instance, if P and Q consist of drawing samples from two different populations, then $P \otimes Q$ consists of the joint experiment where a sample from each population is drawn. We denote by

$$P^{\otimes n} = P \otimes \cdots \otimes P$$

the n-fold product experiment where n independent observations are generated according to the experiment P.

Consider now a Bayesian decision maker whose prior belief assigns probability $1/2$ to the state being 1. To each experiment $P = (\Omega, P_0, P_1)$ we associate a Borel probability measure π over $[0,1]$ that represents the distribution over posterior beliefs induced by the experiment. Formally, let $p(\omega)$ be the posterior belief that the state is 1 given the realization $\omega \in \Omega$:

$$p(\omega) = \frac{\mathrm{d}P_1(\omega)}{\mathrm{d}P_1(\omega) + \mathrm{d}P_0(\omega)}.$$

Furthermore, define for every Borel set $B \subseteq [0,1]$

$$\pi_\theta(B) = P_\theta\left(\{\omega \,:\, p(\omega) \in B\}\right)$$

as the probability that the posterior belief will belong to B, given state θ. We then define $\pi = (\pi_0 + \pi_1)/2$ as the unconditional measure over posterior beliefs.

Throughout the paper we restrict our attention to experiments where the measures P_0 and P_1 are mutually absolutely continuous, so that no signal realization $\omega \in \Omega$ perfectly reveals either state. We say that P is *trivial* if $P_0 = P_1$, and *bounded* if the derivative $\mathrm{d}P_1/\mathrm{d}P_0$ is bounded above and bounded away from 0.

2.2 The Blackwell Order

We first review the main concepts behind Blackwell's order over experiments (Bohnenblust, Shapley, and Sherman, 1949; Blackwell, 1953). Consider two experiments P and Q and

their induced distribution over posterior beliefs denoted by π and τ, respectively. The experiment P *Blackwell dominates* Q, denoted $P \succeq Q$, if

$$\int_0^1 v(p)\,\mathrm{d}\pi(p) \geq \int_0^1 v(p)\,\mathrm{d}\tau(p) \tag{1}$$

for every convex function $v\colon (0,1) \to \mathbb{R}$. Equivalently, $P \succeq Q$ if π is a mean-preserving spread of τ. We write $P \succ Q$ if $P \succeq Q$ and $Q \not\succeq P$. So, $P \succ Q$ if and only if (1) holds with a strict inequality whenever v is strictly convex, i.e. π is a mean-preserving spread of τ and $\pi \neq \tau$.

As is well known, each convex function v can be seen as the indirect utility induced by some decision problem. That is, for each convex v there exists a set of actions A and a utility function u defined on $A \times \{0,1\}$ such that $v(p)$ is the maximal expected payoff that a decision maker can obtain in such a decision problem given a belief p. Hence, $P \succeq Q$ if and only if in every decision problem, an agent can obtain a higher payoff by basing her action on the experiment P rather than on Q.

Blackwell's theorem shows that the order $\succeq$ can be equivalently defined by "garbling" operations: Intuitively, $P \succeq Q$ if and only if the outcome of the experiment Q can be generated from the experiment P by compounding the latter with additional noise, without adding further information about the state.[3]

As discussed in the introduction, we are interested in understanding the large sample properties of the Blackwell order. This motivates the next definition.

Definition 1 (Large Sample Order)**.** An experiment P dominates an experiment Q *in large samples* if there exists an $n_0 \in \mathbb{N}$ such that

$$P^{\otimes n} \succeq Q^{\otimes n} \quad \text{for every} \quad n \geq n_0. \tag{2}$$

This order was first defined by Azrieli (2014) under the terminology of *eventual sufficiency*. The definition captures the informal notion that a large sample drawn from P is more informative than an equally large sample drawn from Q. Consider, for instance, the case of hypothesis testing. The experiment P dominates Q in the Blackwell order if and only if for every test based on Q there exists a test based on P that has weakly lower probabilities of both Type-I and Type-II errors. Definition 1 extends this notion to large samples, in line with the standard paradigm of asymptotic statistics: P dominates Q if every test based on n i.i.d. realizations of Q is dominated by another test based on n i.i.d. realizations of P, for sufficiently large n. When the two experiments are statistics of a

[3]Formally, given two experiments $P = (\Omega, P_0, P_1)$ and $Q = (\Xi, Q_0, Q_1)$, $P \succeq Q$ if and only if there is a measurable kernel (also known as "garbling") $\sigma : \Omega \to \Delta(\Xi)$, where $\Delta(\Xi)$ is the set of probability measures over Ξ, such that for every θ and every measurable $A \subseteq \Xi$, $Q_\theta(A) = \int \sigma(\omega)(A)\,\mathrm{d}P_\theta(\omega)$. In other terms, there is a (perhaps randomly chosen) measurable map f with the property that for both $\theta = 0$ and $\theta = 1$, if X is a random quantity distributed according to P_θ then $Y = f(X)$ is distributed according to Q_θ.

common experiment, dominance in the large sample order implies that one statistic will eventually contain all the information captured by the other.

As shown by Blackwell (1951, Theorem 12), dominance of P over Q implies dominance of $P^{\otimes n}$ over $Q^{\otimes n}$, for every n. So dominance in large samples is an extension of the Blackwell order. This extension is strict, as shown by examples in Torgersen (1970) and Azrieli (2014).

2.3 Rényi Divergence and the Rényi Order

Our main result relates Blackwell dominance in large samples to a well-established notion of informativeness due to Rényi (1961). Given two probability measures μ, ν on a measurable space Ω and a parameter $t > 0$, the Rényi t-divergence is given by

$$R_t(\mu\|\nu) = \frac{1}{t-1} \log \int_\Omega \left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(\omega)\right)^{t-1} \mathrm{d}\mu(\omega) \tag{3}$$

when $t \neq 1$, and, ensuring continuity,

$$R_1(\mu\|\nu) = \int_\Omega \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(\omega)\right) \mathrm{d}\mu(\omega). \tag{4}$$

Equivalently, $R_1(\mu\|\nu)$ is the Kullback-Leibler divergence between the measures μ and ν. As t increases, the value of R_t increases and is continuous whenever it is finite. The limit value as $t \to \infty$, which we denote by $R_\infty(\mu\|\nu)$, is the essential maximum of $\log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)$, the logarithm of the ratio between the two densities.

As a binary experiment precisely consists of a pair of probability measures, we can apply this definition straightforwardly to experiments. Given an experiment $P = (\Omega, P_0, P_1)$, a state θ, and parameter $t > 0$, the Rényi t-divergence of P under θ is

$$R_P^\theta(t) = R_t(P_\theta\|P_{1-\theta}). \tag{5}$$

Intuitively, observing a sample realization for which the likelihood ratio $\mathrm{d}P_\theta/\mathrm{d}P_{1-\theta}$ is high constitutes evidence that favors state θ over $1-\theta$. For instance, in the case of $t = 2$, a higher value of $R_P^\theta(2)$ describes an experiment that, in expectation, more strongly produces evidence in favor of the state θ when this is the correct state. Varying the parameter t allows to consider different moments for the distribution of likelihood ratios. Rényi divergences have found applications to statistics and information theory (Liese and Vajda, 2006; Csiszár, 2008), machine learning (Póczos et al., 2012; Krishnamurthy et al., 2014), computer science (Fritz, 2017), and quantum information (Horodecki et al., 2009; Jensen, 2019). The Hellinger transform (Torgersen, 1991, p. 39), another well known measure of informativeness, is a monotone transformation of the Rényi divergences of an experiment.

The two Rényi divergences R_P^1 and R_P^0 of an experiment are related by the identity

$$R_P^1(t) = \frac{t}{1-t} R_P^0(1-t). \tag{6}$$

Hence the values of $R_P^\theta(t)$ for $t \in [0, 1/2]$ are determined by the values of $R_P^{1-\theta}(t)$ on the interval $[1/2, 1]$. Thus, it suffices to consider values of t in $[1/2, \infty]$.

Definition 2 (Rényi Order)**.** An experiment P dominates an experiment Q in the *Rényi order* if it holds that for all $\theta \in \{0, 1\}$ and all $t > 0$

$$R_P^\theta(t) > R_Q^\theta(t)\,.$$

The Rényi order is a extension of the (strict) Blackwell order. In the proof of Theorem 1 below, we explicitly construct a one-parameter family of decision problems with the property that dominance in the Rényi order is equivalent to higher expected payoff with respect to each decision problem in this family. See §5.1 for details.

A simple calculation shows that if $P = S \otimes T$ is the product of two experiments, then for every state θ,

$$R_P^\theta = R_S^\theta + R_T^\theta\,.$$

A key implication is that P dominates Q in the Rényi order if and only if the same relation holds for their n-th fold repetitions $P^{\otimes n}$ and $Q^{\otimes n}$, for any n. Hence, the Rényi order compares experiments in terms of properties that are unaffected by the number of samples. Because, in turn, the Rényi order extends the Blackwell order, it follows that dominance in the Rényi order is a necessary condition for dominance in large samples.

As a final remark on the definition of the Rényi order, it is important to require the comparison for both states $\theta = 0$ and $\theta = 1$, as there exist pairs of experiments P and Q such that $R_P^1(t) > R_Q^1(t)$ for every t, but $R_P^0(t) < R_Q^0(t)$ for some t.[4]

3 Characterization of the Large Sample Order

We say two bounded experiments P and Q form a *generic* pair if the essential maxima of the log-likelihood ratios $\log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}$ and $\log \frac{\mathrm{d}Q_1}{\mathrm{d}Q_0}$ are different, and if their essential minima are also different. This holds, for example, if for each of the two experiments the set of signal realizations is finite, and there is no posterior beliefs that can be induced by both experiments.

Theorem 1. *For a generic pair of bounded experiments P and Q, the following are equivalent:*

[4] A simple example involves the following pair of binary experiments:

	ω	ω'
P_0	1/3	2/3
P_1	2/3	1/3

	ω	ω'
Q_0	6/9	3/9
Q_1	8/9	1/9

where the entries represent conditional probabilities. Direct computation shows that $R_P^1(t) > R_Q^1(t)$ for every $t > 0$, while $R_P^0(t) < R_Q^0(t)$ for $t > 2$.

(i). *P dominates Q in large samples.*

(ii). *P dominates Q in the Rényi order.*

That (ii) implies (i) means that for every two experiments P and Q that are ranked in the Rényi order, there exists a sample size n such that n or more independent samples of P and Q are ranked in the Blackwell order. The proof of the theorem also establishes an upper bound on n; however, as stating this bound requires several additional concepts we defer this result to Theorem 4 in §5.7. The complete proof of Theorem 1 appears in §5 below.

We mention that Theorem 1 remains true so long as the dominated experiment Q is bounded (whereas P need not be bounded); see §J in the appendix for discussion of this and another generalization. On the other hand, the theorem does not remain true if we remove the genericity assumption. In §I in the appendix we discuss the knife-edge case where the maxima or the minima of the log-likelihood ratios are equal. We demonstrate a non-generic pair of experiments P and Q such that P dominates Q in the Rényi order, but P does not dominate Q in large samples. Given this example, it seems difficult to obtain an applicable characterization of large sample dominance without imposing some genericity condition.

A natural alternative definition of "Blackwell dominance in large samples" would require $P^{\otimes n} \succeq Q^{\otimes n}$ to hold for *some* n, but the resulting order is in fact equivalent under our genericity assumption. This is a consequence of Theorem 1, because $P^{\otimes n_0} \succeq Q^{\otimes n_0}$ for any n_0 implies P dominates Q in the Rényi order, which in turn implies $P^{\otimes n} \succeq Q^{\otimes n}$ for all large n.[5]

3.1 Examples

In this section we illustrate Theorem 1 by means of two examples of pairs of experiments that are not Blackwell ranked, but are ranked in large samples.

Example 1. We first introduce a new example of two such experiments P and Q. The first experiment P appears in Smith and Sørensen (2000). The signal space is the interval $[0,1]$, and the measures P_0 and P_1 are absolutely continuous with densities $f_0(s) = 1$ and $f_1(s) = 1/2 + s$. Our second experiment Q is binary, with signal space $\{0,1\}$. The measure Q_0 assigns probability $1/2$ to both signals, while the other measure is $Q_1(1) = p$ and $Q_1(0) = 1 - p$.

For $p = 0.625$, P Blackwell dominates Q, as witnessed by the garbling from $[0,1]$ to $\{0,1\}$ that maps all signal realizations above $1/2$ to 1 and all realizations below $1/2$ to

[5] However, it is not true that $P^{\otimes n_0} \succeq Q^{\otimes n_0}$ for some n_0 implies $P^{\otimes n} \succeq Q^{\otimes n}$ for all $n \geq n_0$. The case of $\alpha = 0.305$, $\beta = 0.1$ in Example 2 below provides an example where $P^{\otimes 2}$ Blackwell dominates $Q^{\otimes 2}$, but $P^{\otimes 3}$ does not dominate $Q^{\otimes 3}$.

0. For larger p, P is no longer Blackwell dominant. To see this, consider the decision problem in which the prior belief is uniform, the set of actions is the set of states, and the payoff is one if the action matches the state and zero otherwise. It is easy to check that for $p > 0.625$, the experiment Q yields a larger expected payoff.

Nevertheless, if we choose $p = 0.63$, then as Figure 1 below suggests, P dominates Q in the Rényi order even though the two experiments are not Blackwell ranked.[6] Thus, by

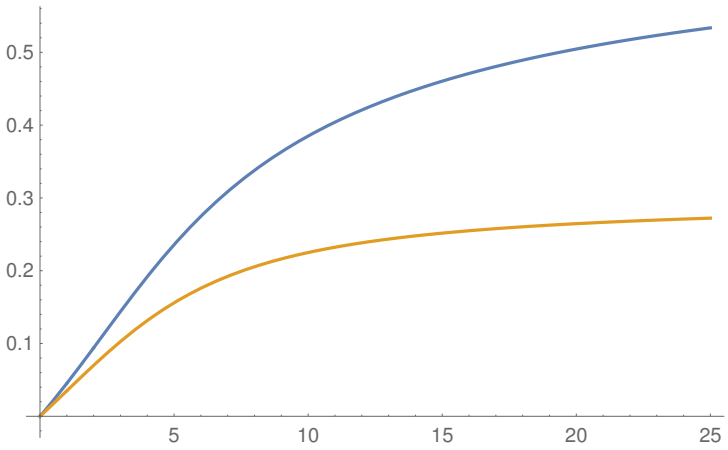

Figure 1: The Rényi divergences R_P^0 (blue), and R_Q^0 (orange) for $p = 0.63$ in Example 1. The comparison between R_P^1 and R_Q^1 yields a similar graph.

Theorem 1, there is some n so that n independent samples from P Blackwell dominate n independent samples from Q.

The next proposition generalizes the example, showing that a binary experiment Q with the same properties can be constructed for (almost) any experiment P.

Proposition 1. *Let P be a bounded experiment with induced distribution over posteriors π. Assume that the support of π has cardinality at least 3. Then there is a binary experiment Q such that P and Q are not Blackwell ranked, and P dominates Q in large samples.*

The proof of this proposition crucially relies on Theorem 1.

[6] The Rényi divergences as defined in (5) are computed to be

$$R_P^0(t) = \frac{1}{t-1}\log\left(\frac{(3/2)^{2-t} - (1/2)^{2-t}}{2-t}\right); \quad R_P^1(t) = \frac{1}{t-1}\log\left(\frac{(3/2)^{t+1} - (1/2)^{t+1}}{t+1}\right)$$

and

$$R_Q^0(t) = \frac{1}{t-1}\log\left(2^{-t}\cdot\left(p^{1-t} + (1-p)^{1-t}\right)\right); \quad R_Q^1(t) = \frac{1}{t-1}\log\left(2^{t-1}\cdot\left(p^t + (1-p)^t\right)\right).$$

Example 2 and a conjecture by Azrieli (2014). We next apply Theorem 1 to revisit an example due to Azrieli (2014) and to complete his analysis. The example provides a simple instance of two experiments that are not ranked in Blackwell order but become so in large samples. Despite its simplicity, the analysis of this example is not straightforward, as shown by Azrieli (2014). We will show that applying the Rényi order greatly simplifies the analysis and elucidates the logic behind the example.

Consider the following two experiments P and Q, parametrized by β and α, respectively. In each matrix, entries are the probabilities of observing each signal realization given the state θ:

P :

θ	x_1	x_2	x_3
0	β	$\frac{1}{2}$	$\frac{1}{2}-\beta$
1	$\frac{1}{2}-\beta$	$\frac{1}{2}$	β

Q :

θ	y_1	y_2
0	α	$1-\alpha$
1	$1-\alpha$	α

The parameters satisfy $0 \leq \beta \leq 1/4$ and $0 \leq \alpha \leq 1/2$. The experiment Q is a symmetric, binary experiment. The experiment P with probability $1/2$ yields a completely uninformative signal realization x_2, and with probability $1/2$ yields an observation from another symmetric binary experiment. As shown by Azrieli (2014, Claim 1), the experiments P and Q are not ranked in the Blackwell order for parameter values $2\beta < \alpha < 1/4 + \beta$.

Azrieli (2014) points out that a necessary condition for P to dominate Q in large samples is that the Rényi divergences are ranked at $1/2$, that is $R^1_P(1/2) > R^1_Q(1/2)$.[7] In addition, he conjectures it is also a sufficient condition, and proves it in the special case of $\beta = 0$. We show that for the experiments in the example, the fact that the Rényi divergences are ranked at $1/2$ is enough to imply dominance in the Rényi order, and therefore, by Theorem 1, dominance in large samples. This settles the above conjecture in the affirmative.

Proposition 2. *In this example, suppose $R^1_P(1/2) > R^1_Q(1/2)$. Then $R^1_P(t) > R^1_Q(t)$ for all $t > 0$ and by symmetry $R^0_P(t) > R^0_Q(t)$, hence P dominates Q in large samples.*

3.2 A Quantification of Blackwell Dominance in Large Samples

The characterization in Theorem 1 makes it possible to quantify the extent to which one experiment Blackwell dominates another in large samples. We start with the observation that any two experiments, even if not ranked according to dominance in large samples,

[7] As in his paper, this condition can be written in terms of the parameter values as

$$\sqrt{\alpha(1-\alpha)} > \sqrt{\beta(\frac{1}{2}-\beta)} + \frac{1}{4}.$$

Thus, when $\alpha = 0.1$ and $\beta = 0$ for example, the experiment P does not Blackwell dominate Q but does dominate it in large samples, as shown by Azrieli (2014).

can be compared by applying different samples sizes. For example, suppose P and Q are not comparable, but $P^{\otimes 50}$ Blackwell dominates $Q^{\otimes 100}$. Then 50 samples from P are more informative than 100 from Q, and thus, in an intuitive sense, P is at least twice as informative as Q, for large enough samples.

Our formal definition is based on the fact that for any two bounded non-trivial experiments P and Q, there exist positive integers n, m such that $P^{\otimes n}$ Blackwell dominates $Q^{\otimes m}$. Reasoning as above, P will be at least m/n times as informative as Q in large samples. We can then consider the largest ratio m/n for which this comparison holds. This leads to a well defined measure of dominance, which we refer to as the *dominance ratio* P/Q of P with respect to Q:

$$P/Q = \sup\left\{\frac{m}{n} \,:\, P^{\otimes n} \succeq Q^{\otimes m}\right\}.$$

Thus, in large samples, each observation from P contributes at least as much as P/Q observations from Q.

An immediate consequence of Theorem 1 is the following characterization of P/Q in terms of the Rényi divergences of the two experiments.

Proposition 3. *Let P and Q be non-trivial, bounded experiments. Then*

$$P/Q = \inf_{\substack{\theta\in\{0,1\}\\ t>0}} \frac{R^\theta_P(t)}{R^\theta_Q(t)}.$$

Furthermore, the dominance ratio P/Q is always positive.[8]

As discussed, P/Q can be interpreted as an asymptotic *lower bound* on the information produced by one observation from P relative to Q. On the other hand, we also have the asymptotic *upper bound* $(Q/P)^{-1}$, where Q/P is the dominance ratio of Q with respect to P. We remark that the two bounds are in general (in fact, generically) not equal. However, Proposition 3 shows that $P/Q \leq (Q/P)^{-1}$ always holds.

3.3 The Blackwell Order in the Presence of Additional Information

The large sample order compares the informativeness of repeated experiments. A related problem is to compare the informativeness of one-shot experiments when additional independent sources of information may be present.

[8]This characterization, together with Theorem 1, implies that the following natural alternative definition of P/Q is equivalent:

$$P/Q = \sup\left\{a > 0 \,:\, P^{\otimes n} \succeq Q^{\otimes \lceil an\rceil} \text{ for } \textit{all } n \text{ large enough}\right\}$$

where $\lceil an\rceil$ denotes the smallest integer greater than or equal to an.

Consider a decision maker choosing which of two experiments P and Q to conduct, *on top of* an independent source of information R. The resulting choice is between the compound experiments $P \otimes R$ and $Q \otimes R$. It is intuitive, and immediate from Blackwell's garbling characterization, that if P dominates Q in the Blackwell order, then the same relation must hold between the two compound experiments.

One might expect that if P and Q are incomparable, then no additional independent experiment R can make the compound experiments comparable. Instead, we show that $P \otimes R$ can dominate $Q \otimes R$ even though the two original experiments P and Q were not comparable. Moreover, for generic experiments, this occurs precisely when P has higher Rényi divergences than Q.

Proposition 4. *Let P and Q be a generic pair of bounded experiments. Then the following are equivalent:*

(i). *There exists a bounded experiment R such that $P \otimes R \succeq Q \otimes R$.*

(ii). *P dominates Q in the Rényi order.*

Proposition 4 suggests that in general, whether two experiments are Blackwell ordered depends on *what* additional sources of information are available. We note that whenever an experiment R makes P dominant over Q (when each is combined with R), then the same holds for any experiment R' that is more informative than R. It is an interesting question for future work to fully characterize the set of experiments R that make P dominant.

Proposition 4 follows by combining the characterization in Theorem 1 together with the observation that if P dominates Q in the large sample order, then there exists an R such that $P \otimes R$ Blackwell dominates $Q \otimes R$. The latter fact is a consequence of an order-theoretic result from the quantum information literature (Duan et al., 2005; Fritz, 2017, see Lemma 4 in the appendix).

4 A Characterization of Additive Divergences

In this section we apply the characterization of Blackwell dominance in large samples to study measures for quantifying the degree of dissimilarity between distributions, also known as *divergences*. Examples of divergences include total variation distance, the Hellinger distance, the Kullback-Leibler divergence, Rényi divergences, and more general f-divergences.

A key property of Rényi divergences is additivity. Consider two domains Ω_1 and Ω_2, a pair of measures μ_1, ν_1 defined on Ω_1, and a pair of measures μ_2, ν_2 on Ω_2. Additivity states that when the two domains are considered in conjunction, the divergence between the product measures $\mu_1 \times \mu_2$ and $\nu_1 \times \nu_2$, which are both defined on $\Omega_1 \times \Omega_2$, is the sum of the divergences of the two pairs. In words, this condition says that the total divergence of two unrelated pairs should not change when they are considered together as a bundle.

Another property of Rényi divergences, which it in fact shares with all the above examples of divergences, is the *data processing inequality*, which captures the idea that discarding some information decreases dissimilarity.

We show that every additive divergence that satisfies the data-processing inequality is an integral of Rényi divergences. The proof relies on the characterization of the large sample order together with functional analytic techniques. Since this result does not assume any functional form of the divergence, it improves over the existing characterizations such as in Rényi (1961) and Csiszár (2008).

The result has potential applications for modeling experiments as economic commodities. In recent years, there has been growing interest in modeling the cost and pricing of information. By interpreting a divergence as a cost function over experiments, additivity reflects an assumption of constant marginal costs in information production (an assumption discussed in detail in Pomatto et al., 2018). By interpreting a divergence as a pricing function over experiments, additivity captures a notion of linearity, appropriate for pricing information in competitive markets.

4.1 Additive Divergences

Given a Polish space Ω, we denote by $\mathcal{B}(\Omega)$ its Borel σ-algebra and by $\Delta(\Omega)$ the collection of Borel probability measures on $\mathcal{B}(\Omega)$. Given another Polish space Ξ, a measurable function $f\colon \Omega \to \Xi$ and a probability measure $\mu \in \Delta(\Omega)$, we denote by $f_*(\mu)$ the push-forward probability measure in $\Delta(\Xi)$ defined as $[f_*(\mu)](E) = \mu(f^{-1}(E))$ for all $E \in \mathcal{B}(\Xi)$.

Consider, for each Ω, a map

$$D_\Omega\colon \Delta(\Omega) \times \Delta(\Omega) \to \mathbb{R}_+ \cup \{+\infty\},$$

and let $D = (D_\Omega)$ be the collection obtained by varying Ω. We say D is a *divergence* if $D_\Omega(\mu, \mu) = 0$ for all Ω and all $\mu \in \Delta(\Omega)$.

A divergence satisfies the *data processing inequality* if for any measurable $f\colon \Omega \to \Xi$ it holds that

$$D_\Xi(f_*(\mu), f_*(\nu)) \leq D_\Omega(\mu, \nu).$$

The data processing inequality captures the idea that the distributions of two random variables X and Y are at least as dissimilar as those of $f(X)$ and $f(Y)$; applying a common deterministic mapping f can only make the distributions more similar.[9] It is a natural concept in signal processing and information theory, and closely related to the Blackwell order over experiments. Indeed, we can see a pair of probability measures as an experiment

[9]Note that the data processing inequality implies that D is invariant to measurable isomorphisms: If f is a bijection then $D_\Xi(f_*(\mu), f_*(\nu)) = D_\Omega(\mu, \nu)$. Thus the dissimilarity between measures does not depend on the particular labelling of the domain.

(P_0, P_1), and hence a divergence D as a functional over experiments. The data-processing inequality states that the value of D decreases when applying a deterministic garbling.

We say that the divergence D is *additive* if

$$D_{\Omega\times\Xi}(\mu_1 \times \mu_2, \nu_1 \times \nu_2) = D_\Omega(\mu_1, \nu_1) + D_\Xi(\mu_2, \nu_2).$$

We will henceforth drop the subscript from $D_\Omega(\mu, \nu)$, and write $D(\mu, \nu)$ whenever there is no risk of confusion.

We call a pair μ, ν of measures as *bounded* if there exists an $M > 0$ such that for any measurable $A \subseteq \Omega$, $\nu(A) \geq \mu(A)/M$ and $\mu(A) \geq \nu(A)/M$. Equivalently, $\mathrm{d}\mu/\mathrm{d}\nu$ is supported on $[1/M, M]$, and hence bounded from above and bounded away from 0. We will restrict our attention to divergences that take finite values on bounded pairs of experiments.

4.2 Representation Theorem

Our representation theorem shows that all additive divergences that are finite on bounded experiments arise from linear combinations of Rényi divergences.

Theorem 2. *Let D be an additive divergence that satisfies the data processing inequality and is finite on bounded experiments. Then there exist two finite Borel measures m_0, m_1 on $[1/2, \infty]$ such that for every bounded pair μ, ν it holds that*

$$D(\mu, \nu) = \int_{[1/2,\infty]} R_t(\mu\|\nu)\,\mathrm{d}m_0(t) + \int_{[1/2,\infty]} R_t(\nu\|\mu)\,\mathrm{d}m_1(t), \tag{7}$$

with R_t given by (3) and (4).

Varying the two measures m_0 and m_1 leads to some important special cases. When both are finitely supported, D is a linear combination of Rényi divergences. Any additive divergence D (finite on bounded experiments) is hence a limit of such combinations. When m_0 and m_1 are Dirac probability measures concentrated on 1, D reduces to twice the Jensen-Shannon divergence, which is the symmetric counterpart of the Kullback-Leibler divergence. When instead m_0 is a Dirac probability measure concentrated on 1 and m_1 is set to have total mass zero, D reduces to the Kullback-Leibler divergence.

Note that the lower integration bound in (7) is 1/2. This is because, as discussed, the values of $R_t(\mu\|\nu)$ are related to the values of $R_{1-t}(\nu\|\mu)$. Hence it suffices to consider values of t above 1/2.

Proof Sketch of Theorem 2. The first key idea is to see a bounded pair of probability measures as a bounded experiment (P_0, P_1), and hence see a divergence D as a functional over experiments. When D is additive, the data processing inequality implies monotonicity with respect to the Blackwell order.

The next crucial step is to leverage Theorem 1 to show that additivity renders D monotone in the Rényi order. Indeed, if (P_0, P_1) dominates (Q_0, Q_1) in the Rényi order, then, by Theorem 1, there exists a number n of repetitions such that (P_0^n, P_1^n) dominates (Q_0^n, Q_1^n) in the Blackwell order. Hence, by combining Blackwell monotonicity and additivity, we obtain that D must satisfy

$$nD(P_0, P_1) = D(P_0^n, P_1^n) \geq D(Q_0^n, Q_1^n) = nD(Q_0, Q_1).$$

Hence, D is monotone in the Rényi order.

We deduce from this that D is a monotone functional $F(R_P^0, R_P^1)$ of the Rényi divergences of the experiment. Additivity of D implies F is also additive. We then use tools from functional analysis to show that F extends to a positive linear functional, leading to the integral representation of Theorem 2.

5 Proof of Theorem 1

The proof of Theorem 1 is organized as follows. In §5.1 we first show that the Rényi order is necessary for the large sample order. The remaining subsections demonstrate sufficiency. In §5.3 we provide a novel characterization of Blackwell dominance, showing that it is equivalent to first-order stochastic dominance of appropriate statistics of the two experiments. §5.5 applies this observation, together with techniques from large deviations theory. Omitted proofs are deferred to the appendix.

5.1 Dominance in Large Samples Implies Dominance in the Rényi Order

As discussed above, the comparison of Rényi divergences between two experiments is independent of the number of samples. Thus it suffices to show that the Rényi order extends the strict Blackwell order.[10] We do this by constructing decision problems with the property that higher expected payoff in these problems translates into higher Rényi divergences.

For each $t > 1$, the function $v_1(p) = 2p^t(1-p)^{1-t}$ defined for $p \in (0,1)$ is strictly convex, because its second derivative in p is $2t(t-1)p^{t-2}(1-p)^{-1-t}$. Thus $v_1(p)$ is the indirect utility function induced by some decision problem. Moreover, we have that

$$\int_0^1 v_1(p)\,\mathrm{d}\pi(p) = \int_\Omega \left(\frac{\mathrm{d}P_1(\omega)}{\mathrm{d}P_0(\omega)}\right)^{t-1} \mathrm{d}P_1(\omega) = \mathrm{e}^{(t-1)R_P^1(t)}. \tag{8}$$

To see this, recall that π_θ is the distribution over posteriors induced by P, conditional on state $\theta \in \{0,1\}$, and that

$$\mathrm{d}\pi(p) = \frac{1}{2}(\mathrm{d}\pi_1(p) + \mathrm{d}\pi_0(p)) \qquad \text{and} \qquad \mathrm{d}\pi_1(p) = \frac{p}{1-p}\,\mathrm{d}\pi_0(p). \tag{9}$$

[10] Since by assumption the two experiments P and Q form a generic pair, Blackwell dominance of $P^{\otimes n}$ over $Q^{\otimes n}$ necessarily implies strict Blackwell dominance.

Thus $\mathrm{d}\pi(p) = \frac{1}{2p}\,\mathrm{d}\pi_1(p)$, which allows us to write

$$\int_0^1 v_1(p)\,\mathrm{d}\pi(p) = \int_0^1 2p^t(1-p)^{1-t}\cdot\frac{1}{2p}\,\mathrm{d}\pi_1(p) = \int_0^1 \left(\frac{p}{1-p}\right)^{t-1}\,\mathrm{d}\pi_1(p).$$

The first equality in (8) then follows from a change of variable from signal realizations ω to posterior beliefs $p = \frac{\mathrm{d}P_1(\omega)}{\mathrm{d}P_1(\omega)+\mathrm{d}P_0(\omega)}$ (with the probability measure changing from P_1 to π_1, holding fixed the true state $\theta = 1$).

The second equality in (8) follows from the definition of Rényi divergences. Thus (8) holds, which shows that in the decision problem with indirect utility function $v_1(p)$, the ex-ante expected payoff is a monotone transformation of the Rényi divergence $R^1_P(t)$. Hence, experiment P yields higher expected payoff in this decision problem than Q if and only if $R^1_P(t) > R^1_Q(t)$.

Similarly, for $t \in (0,1)$ we consider the indirect utility function $v_2(p) = -2p^t(1-p)^{1-t}$, which is now strictly convex due to the negative sign (its second derivative is $2t(1-t)p^{t-2}(1-p)^{-1-t}$). Then

$$\int_0^1 v_2(p)\,\mathrm{d}\pi(p) = -\mathrm{e}^{(t-1)R^1_P(t)}$$

is again a monotone transformation of the Rényi divergence. So P yields higher expected payoff in this decision problem only if $R^1_P(t) > R^1_Q(t)$.

For $t = 1$, we consider the indirect utility function $v_3(p) = 2p\log(\frac{p}{1-p})$, which is strictly convex with a second derivative of $2p^{-1}(1-p)^{-2}$. We have

$$\int_0^1 v_3(p)\,\mathrm{d}\pi(p) = \int_0^1 \log\left(\frac{p}{1-p}\right)\,\mathrm{d}\pi_1(p) = \int_\Omega \log\left(\frac{\mathrm{d}P_1(\omega)}{\mathrm{d}P_0(\omega)}\right)\,\mathrm{d}P_1(\omega) = R^1_P(1).$$

Thus P yields higher expected payoff in this problem if and only if $R^1_P(1) > R^1_Q(1)$.

Summarizing, the above family of decision problems shows that P strictly Blackwell dominates Q only if $R^1_P(t) > R^1_Q(t)$ for all $t > 0$. Since the two states are symmetric, another set of necessary conditions is that $R^0_P(t) > R^0_Q(t)$ for all $t > 0$. Hence dominance in the Rényi order is necessary for Blackwell dominance and (due to additivity of Rényi divergences) also for dominance in large samples.

5.2 Repeated Experiments and Log-Likelihood Ratios

We turn to the proof that dominance in the Rényi order is (generically) sufficient for dominance in large samples. Recall that $P^{\otimes n}$ Blackwell dominates $Q^{\otimes n}$ if and only if the former induces a distribution over posterior beliefs that is a mean-preserving spread of the latter. However, the distribution over posteriors induced by a product experiment can be difficult to analyze directly. A more suitable approach consists in studying the distribution of the induced log-likelihood ratio

$$\log\frac{\mathrm{d}P_\theta}{\mathrm{d}P_{1-\theta}}.$$

As is well known, given a repeated experiment $P^{\otimes n} = (\Omega^n, P_0^n, P_1^n)$, its log-likelihood ratio satisfies, for every realization $\omega = (\omega_1, \ldots, \omega_n)$ in Ω^n,

$$\log \frac{\mathrm{d}P_1^n}{\mathrm{d}P_0^n}(\omega) = \sum_{i=1}^{n} \log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}(\omega_i).$$

Moreover, the random variables

$$X_i(\omega) = \log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}(\omega_i) \quad i = 1, \ldots, n$$

are i.i.d. under P_θ^n, for $\theta \in \{0, 1\}$. Focusing on the distributions of log-likelihood ratios will allow us to transform the study of repeated experiments to the study of sums of i.i.d. random variables.

5.3 From Blackwell Dominance to First-Order Stochastic Dominance

Expressing posterior beliefs in terms of log-likelihood ratios simplifies the analysis of repeated experiments. However, it is not obvious that the Blackwell order admits a simple interpretation in this domain.

We provide a novel characterization of the Blackwell order, expressed in terms of the distributions of the log-likelihood ratios. Given two experiments $P = (\Omega, P_0, P_1)$ and $Q = (\Xi, Q_0, Q_1)$ we denote by F_θ and G_θ, respectively, the cumulative distribution function of the log-likelihood ratios conditional on state θ. That is,

$$F_\theta(a) = P_\theta\left(\left\{\log \frac{\mathrm{d}P_\theta}{\mathrm{d}P_{1-\theta}} \leq a\right\}\right) \quad \text{for all} \;\; a \in \mathbb{R}, \; \theta \in \{0, 1\}. \tag{10}$$

The c.d.f. G_θ is defined analogously using Q_θ.

We associate to P a new quantity, which we call the *perfected log-likelihood ratio* of the experiment. Define

$$\tilde{L}_1 = \log \frac{\mathrm{d}P_1}{\mathrm{d}P_0} - E$$

where E is a random variable that, under P_1, is independent from $\log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}$ and distributed according to an exponential distribution with support $\mathbb{R}_+$ and cumulative distribution function $1 - \mathrm{e}^{-x}$ for all $x \geq 0$. We denote by $\tilde{F}_1$ the cumulative distribution function of $\tilde{L}_1$ under P_1. That is, $\tilde{F}_1(a) = P_1(\{\tilde{L}_1 \leq a\})$ for all $a \in \mathbb{R}$.

More explicitly, $\tilde{F}_1$ is the convolution of the distribution F_1 with the distribution of $-E$, and thus can be defined as

$$\tilde{F}_1(a) = \int_{\mathbb{R}} P_1(\{-E \leq a - u\}) \, \mathrm{d}F_1(u) = F_1(a) + \mathrm{e}^a \int_{(a,\infty)} \mathrm{e}^{-u} \, \mathrm{d}F_1(u). \tag{11}$$

The next result shows that the Blackwell order over experiments can be reduced to first-order stochastic dominance of the corresponding perfected log-likelihood ratios.

Theorem 3. *Let P and Q be two experiments, and let $\tilde{F}_1$ and $\tilde{G}_1$, respectively, be the associated distributions of perfected log-likelihood ratios. Then*

$$P \succeq Q \quad \textit{if and only if} \quad \tilde{F}_1(a) \leq \tilde{G}_1(a) \textit{ for all } a \in \mathbb{R}.$$

Proof. Let π and τ be the distributions over posterior beliefs induced by P and Q, respectively. As is well known, Blackwell dominance is equivalent to the requirement that π is a mean-preserving spread of τ. Equivalently the functions defined as

$$\Lambda_\pi(p) = \int_{[0,p]} (p-q)\,\mathrm{d}\pi(q) \quad \text{and} \quad \Lambda_\tau(p) = \int_{[0,p]} (p-q)\,\mathrm{d}\tau(q) \tag{12}$$

must satisfy $\Lambda_\pi(p) \geq \Lambda_\tau(p)$ for every $p \in (0,1)$.

We now express (12) in terms of the distributions of log-likelihood ratios F_1 and G_1. We have

$$\Lambda_\pi(p) = p\left(1 - \int_{(p,1]} 1\,\mathrm{d}\pi(q)\right) - \int_{[0,p]} q\,\mathrm{d}\pi(q). \tag{13}$$

To transform the relevant integrals into those that condition on state 1, we recall that (9) implies $\mathrm{d}\pi(q) = \frac{1}{2q}\,\mathrm{d}\pi_1(q)$. We then obtain from (13) that

$$2\Lambda_\pi(p) = p\left(2 - \int_{(p,1]} \frac{1}{q}\,\mathrm{d}\pi_1(q)\right) - \int_{[0,p]} \mathrm{d}\pi_1(q).$$

Next, we change variable from posterior beliefs to log-likelihood ratios. Letting $a = \log\frac{p}{1-p}$ and accordingly $u = \log\frac{q}{1-q}$, we have

$$2\Lambda_\pi(p) = \frac{\mathrm{e}^a}{1+\mathrm{e}^a}\left(2 - \int_{(a,\infty)} \frac{1+\mathrm{e}^u}{\mathrm{e}^u}\,\mathrm{d}F_1(u)\right) - F_1(a). \tag{14}$$

Since

$$\int_{(a,\infty)} \frac{1+\mathrm{e}^u}{\mathrm{e}^u}\,\mathrm{d}F_1(u) = \int_{(a,\infty)} \mathrm{e}^{-u}\,\mathrm{d}F_1(u) + 1 - F_1(a),$$

(14) leads to

$$2\Lambda_\pi(p) = \frac{\mathrm{e}^a}{1+\mathrm{e}^a} - \frac{F_1(a)}{1+\mathrm{e}^a} - \frac{\mathrm{e}^a}{1+\mathrm{e}^a}\int_{(a,\infty)} \mathrm{e}^{-u}\,\mathrm{d}F_1(u) = \frac{\mathrm{e}^a}{1+\mathrm{e}^a} - \frac{\tilde{F}_1(a)}{1+\mathrm{e}^a},$$

where the final equality follows from (11). It then follows that $\Lambda_\pi(p) \geq \Lambda_\pi(p)$ if and only if $\tilde{F}_1(a) \leq \tilde{G}_1(a)$ for $a = \log\frac{p}{1-p}$. Requiring this for all $p \in (0,1)$ yields the theorem. □

Intuitively, transferring probability mass from lower to higher values of $\log(\mathrm{d}P_\theta/\mathrm{d}P_{1-\theta})$ leads to an experiment that, conditional on the state being θ, is more likely to shift the decision maker's beliefs towards the correct state. Hence, one might conjecture that Blackwell dominance of the experiments P and Q is related to stochastic dominance of

the distributions F_θ and G_θ. However, since the likelihood ratio $\mathrm{d}P_1/\mathrm{d}P_0$ must satisfy the change of measure identity $\int \frac{\mathrm{d}P_0}{\mathrm{d}P_1}\,\mathrm{d}P_1 = 1$, the distribution F_1 must satisfy

$$\int_{\mathbb{R}} \mathrm{e}^{-u}\,\mathrm{d}F_1(u) = 1.$$

Because the function e^{-u} is strictly decreasing and convex, and the same identity must hold for G_1, it is impossible for F_1 to stochastically dominate G_1. Theorem 3 shows that a more useful comparison is between the perfected log-likelihood ratios.[11]

The next lemma simplifies the study of perfected log-likelihood ratios, by showing that their first-order stochastic dominance can be deduced from comparisons of the original distributions F_θ and G_θ over subintervals.

Lemma 1. *Consider two experiments P and Q. Let F_θ and G_θ, respectively, be the distributions of the corresponding log-likelihood ratios, and $\tilde{F}_1$ and $\tilde{G}_1$ be the distributions of the perfected log-likelihood ratios. The following holds:*

(i). *If $F_1(a) \leq G_1(a)$ for all $a \geq 0$, then $\tilde{F}_1(a) \leq \tilde{G}_1(a)$ for all $a \geq 0$.*

(ii). *If $F_0(a) \leq G_0(a)$ for all $a \geq 0$, then $\tilde{F}_1(a) \leq \tilde{G}_1(a)$ for all $a \leq 0$.*

5.4 Large Deviations

The main step in the proof of Theorem 1 relies on the theory of large deviations. Large deviations theory studies low probability events, and in particular the odds with which an i.i.d. sum deviates from its expectation. The Law of Large Numbers implies that for a random variable X, the probability of the event $\{X_1 + \cdots + X_n > na\}$ is low for $a > \mathbb{E}[X]$ and large n, where $X_1, \ldots, X_n$ are i.i.d. copies of X. A crucial insight due to Cramér (1938) is that the order of magnitude of the probability of this event is determined by the *cumulant generating function* of X, defined as

$$K_X(t) = \log \mathbb{E}[\mathrm{e}^{tX}]$$

for every $t \in \mathbb{R}$.

As is well known, K_X is strictly convex whenever X is not a constant. We denote by

$$K_X^*(a) = \sup_{t \in \mathbb{R}} t \cdot a - K_X(t) \quad a \in \mathbb{R}, \tag{15}$$

its Fenchel conjugate. Two facts we will repeatedly apply are that for every $a \in (\min[X], \max[X])$ the problem (15) has a unique solution $t \in \mathbb{R}$, and such t is non-negative if and only if $a \geq \mathbb{E}[X]$. Moreover, $K_X^* \geq 0 \cdot a - K_X(0) = 0$ is non-negative.

[11] It might appear puzzling that two distributions F_1 and G_1 that are not ranked by stochastic dominance become ranked after the addition of the same independent random variable. In a different context and under different assumptions, the same phenomenon is studied by Pomatto, Strack, and Tamuz (2019).

Cramér's Theorem establishes that for each threshold $a > \mathbb{E}[X]$, the exponential rate at which the probability of the event $\{X_1 + \cdots + X_n > na\}$ vanishes with n is equal to the value $K_X^*(a)$ taken by the Fenchel conjugate at a. In this paper we are interested in comparing the probabilities of large deviations across different random variables. Consider, to this end, two random variables X and Y and a threshold a strictly greater than $\mathbb{E}[X]$ and $\mathbb{E}[Y]$. If

$$K_Y^*(a) > K_X^*(a),$$

then the probability of the event $\{X_1 + \cdots + X_n > na\}$ vanishes more slowly than the probability of the event $\{Y_1 + \cdots + Y_n > na\}$. Thus there exists n sufficiently large such that

$$\mathbb{P}\left[X_1 + \cdots + X_n > na\right] \geq \mathbb{P}\left[Y_1 + \cdots + Y_n > na\right].$$

The next proposition establishes a general version of this fact, while also providing a specific number of repetitions sufficient to rank the probability of the two events.

Proposition 5. *Let X and Y be random variables taking values in $[-b, b]$ and let $X_1, \ldots, X_n$, $Y_1, \ldots, Y_n$ be i.i.d. copies of X and Y respectively. Suppose $a \geq \mathbb{E}[Y]$, and $\eta > 0$ satisfies $K_Y^*(a) - \eta > K_X^*(a + \eta)$. Then for all $n \geq 4b^2(1 + \eta)\eta^{-3}$, it holds that*

$$\mathbb{P}\left[X_1 + \cdots + X_n > na\right] \geq \mathbb{P}\left[Y_1 + \cdots + Y_n > na\right]. \tag{16}$$

The condition $K_Y^*(a) - \eta > K_X^*(a + \eta)$ ensures that the rate at which the probability of the events $\{Y_1 + \ldots + Y_n > na\}$ vanish with n is larger by a factor of at least η than the rate of the events $\{X_1 + \ldots + X_n > n(a + \eta)\}$. Larger values of η make this condition more demanding, and imply that a smaller number of repetitions is sufficient to guarantee (16) to hold.

5.5 Application to the Rényi Order

Now consider two experiments $P = (\Omega, P_0, P_1)$ and $Q = (\Xi, Q_0, Q_1)$. Denote the corresponding log-likelihood ratios

$$X^\theta = \log \frac{\mathrm{d}P_\theta}{\mathrm{d}P_{1-\theta}} \quad \text{and} \quad Y^\theta = \log \frac{\mathrm{d}Q_\theta}{\mathrm{d}Q_{1-\theta}}$$

defined over the probability spaces (Ω, P_θ) and (Ξ, Q_θ), respectively. Thus, for instance, X^1 is the log-likelihood ratio of state 1 to state 0, distributed conditional on state 1, and X^0 is the log-likelihood ratio of state 0 to 1, distributed conditional on state 0.

The cumulant generating function of the log-likelihood ratio is a simple transformation of the Rényi divergences, as defined in (3), (4) and (5):

$$K_{X^\theta}(t) = t \cdot R_P^\theta(t + 1). \tag{17}$$

Likewise $K_{Y^\theta}(t) = t \cdot R^\theta_Q(t+1)$. Hence, if P dominates Q in the Rényi order then the following relation must hold between the cumulant generating functions:

$$K_{X^\theta}(t) > K_{Y^\theta}(t) \quad \text{for } t > 0 \tag{18}$$

$$K_{X^\theta}(t) < K_{Y^\theta}(t) \quad \text{for } -1 < t < 0. \tag{19}$$

At $t = 0$ we have $K_{X^\theta}(0) = K_{Y^\theta}(0) = 0$, but $K'_{X^\theta}(0) > K'_{Y^\theta}(0)$ must hold by (17) and the assumption that $R^\theta_P(1) > R^\theta_Q(1)$. It is well known that $K'_{X^\theta}(0) = \mathbb{E}[X^\theta]$, which by definition is the Kullback-Leibler divergence between P^θ and $P^{1-\theta}$. Hence we also have

$$\mathbb{E}[X^\theta] > \mathbb{E}[Y^\theta] > 0.$$[12]

The Fenchel conjugate is an order-reversing operation: From (15) we see that if $K_X \geq K_Y$ pointwise, then the corresponding conjugates satisfy $K^*_Y \geq K^*_X$ pointwise. The relation between K_{X^θ} and K_{Y^θ} established in (18) and (19) is more complicated, and implies the following ranking of their conjugates:

$$K^*_{Y^\theta}(a) > K^*_{X^\theta}(a) \quad \text{for} \quad \mathbb{E}[X^\theta] \leq a \leq \max[Y^\theta]$$
$$K^*_{Y^\theta}(a) < K^*_{X^\theta}(a) \quad \text{for} \quad 0 \leq a \leq \mathbb{E}[Y^\theta].$$

This is the content of the next lemma, which in addition shows that the differences between the Fenchel conjugates admit a uniform bound.

Lemma 2. *Suppose P and Q are a generic pair of bounded experiments such that P dominates Q in the Rényi order. Let (X^θ) and (Y^θ) be the corresponding log-likelihood ratios. Then there exists $\eta \in (0,1)$ such that in both states $\theta \in \{0,1\}$*

$$K^*_{Y^\theta}(a) - \eta > K^*_{X^\theta}(a+\eta) \quad \textit{for} \quad \mathbb{E}[X^\theta] - \eta \leq a \leq \max[Y^\theta]$$
$$K^*_{Y^\theta}(a-\eta) < K^*_{X^\theta}(a) - \eta \quad \textit{for} \quad 0 \leq a \leq \mathbb{E}[Y^\theta] + \eta.$$

These estimates will allow us to apply the previous Proposition 5 and make uniform comparisons of large deviation probabilities. In the range $a \in (\mathbb{E}[Y^\theta] + \eta, \mathbb{E}[X^\theta] - \eta)$ that is not covered by Lemma 2, large deviation techniques are not necessary and it will be sufficient to apply more elementary estimates.

5.6 Rényi Order Implies Large Sample Order

We now complete the proof of Theorem 1 and show that if two experiments are ranked in the Rényi order then they are also ranked in the large sample order. By Theorem 3

[12] Throughout the proof we assume Q is a non-trivial experiment, so that $\mathbb{E}[Y^\theta]$ being the Kullback-Leibler divergence between Q^θ and $Q^{1-\theta}$ is strictly positive. This is without loss, as P clearly dominates Q (in large samples) in case Q is trivial.

we need to show that there exists a sample size n_0 such that for all $n \geq n_0$, the perfected log-likelihood ratios of n independent draws from P and Q are ordered in terms of first-order stochastic dominance.

More concretely, consider the log-likelihood ratios X^θ and Y^θ (for a single sample) as defined above, with distributions F_θ and G_θ conditional on state θ. Let F_θ^{*n} be the n-th convolution power of F_θ, which represents the distribution of log-likelihood ratios under the product experiment $P^{\otimes n}$; similarly define G_θ^{*n}. By Lemma 1, it suffices to show that for $n \geq n_0$ it holds that

$$F_1^{*n}(na) \leq G_1^{*n}(na) \quad \text{for all } a \geq 0 \tag{20}$$

and

$$F_0^{*n}(na) \leq G_0^{*n}(na) \quad \text{for all } a \geq 0. \tag{21}$$

Below we show (20); the argument for (21) is identical after relabelling the states. Assume that X^1 and Y^1 take values in $[-b, b]$. We will set $n_0 = 8b^2\eta^{-3}$, where $\eta \in (0,1)$ is as given in Lemma 2. For future use, we note that $\mathbb{E}[X^1] - \eta > \mathbb{E}[Y^1]$.[13]

Let $X_1^1, \ldots, X_n^1$ be i.i.d. copies of X^1 and $Y_1^1, \ldots, Y_n^1$ be i.i.d. copies of Y^1. We can restate (20) as

$$\mathbb{P}\left[X_1^1 + \cdots + X_n^1 \leq na\right] \leq \mathbb{P}\left[Y_1^1 + \cdots + Y_n^1 \leq na\right], \quad \text{for all } a \geq 0. \tag{22}$$

To prove this, we divide into four ranges of values of a:

Case 1: $a \geq \max[Y^1]$**.** In this case the right-hand side of (22) is 1, and hence the result follows trivially.

Case 2: $\mathbb{E}[X^1] - \eta \leq a < \max[Y^1]$**.** From Lemma 2 we have that

$$K^*_{Y^1}(a) - \eta > K^*_{X^1}(a + \eta).$$

As $a \geq \mathbb{E}[X^1] - \eta > \mathbb{E}[Y^1]$, we can directly apply Proposition 5 and conclude that (22) holds for all $n \geq 4b^2(1+\eta)\eta^{-3}$. Since $\eta < 1$, it holds for all $n \geq n_0 = 8b^2\eta^{-3}$.

Case 3: $\mathbb{E}[Y^1] + \eta \leq a < \mathbb{E}[X^1] - \eta$**.** By the Chebyshev inequality,

$$\mathbb{P}\left[X_1^1 + \cdots + X_n^1 \leq na\right] \leq \mathbb{P}\left[X_1^1 + \cdots + X_n^1 \leq n(\mathbb{E}[X^1] - \eta)\right] \leq \frac{\operatorname{Var}(X_1^1 + \cdots + X_n^1)}{n^2\eta^2}.$$

Since $\operatorname{Var}(X_1^1 + \cdots + X_n^1) = n \operatorname{Var}(X^1) \leq nb^2$, we have that

$$\mathbb{P}\left[X_1^1 + \cdots + X_n^1 \leq na\right] \leq \frac{b^2}{n\eta^2}.$$

[13] Otherwise, the first part of Lemma 2 would apply to $a = \mathbb{E}[Y^1]$, leading to $0 - \eta > K^*_{X^1}(a + \theta)$. This is impossible as K^* is non-negative.

By a similar argument,

$$\mathbb{P}\left[Y_1^1 + \cdots + Y_n^1 \leq na\right] \geq 1 - \frac{b^2}{n\eta^2}.$$

Hence for all $n \geq 2b^2\eta^{-2}$ we have

$$\mathbb{P}\left[X_1^1 + \cdots + X_n^1 \leq na\right] \leq \mathbb{P}\left[Y_1^1 + \cdots + Y_n^1 \leq na\right].$$

As $n_0 = 8b^2\eta^{-3}$ is bigger, (22) holds for $n \geq n_0$.

Case 4: $0 \leq a < \mathbb{E}[Y^1] + \eta$. By Lemma 2 we have that

$$K^*_{X^1}(a) - \eta > K^*_{Y^1}(a - \eta).$$

For any random variable Z, we have $K_{-Z}(t) = \log \mathbb{E}\left[\mathrm{e}^{t(-Z)}\right] = \log \mathbb{E}\left[\mathrm{e}^{(-t)Z}\right] = K_Z(-t)$, and $K^*_{-Z}(a) = \sup_{t\in\mathbb{R}} t \cdot a - K_{-Z}(t) = \sup_{t\in\mathbb{R}}(-t)\cdot(-a) - K_Z(-t) = K^*_Z(-a)$. Therefore

$$K^*_{-X^1}(-a) - \eta > K^*_{-Y^1}(-a + \eta).$$

We can now apply Proposition 5 to the random variables $-Y^1$ and $-X^1$, and the threshold $-a > -\mathbb{E}[Y^1] - \eta > \mathbb{E}[-X^1]$. This yields

$$\mathbb{P}\left[-Y_1^1 - \cdots - Y_n^1 > -na\right] \geq \mathbb{P}\left[-X_1^1 - \cdots - X_n^1 > -na\right]$$

for all $n \geq 4b^2(1+\eta)\eta^{-3}$. Hence (22) holds for $n \geq n_0$.[14]

This proves (22) for all $a \geq 0$ and completes the proof of Theorem 1.

5.7 Number of Samples Required

The proof of Theorem 1 establishes a stronger statement, and in fact provides an explicit bound on the number of repetitions sufficient to achieve large sample dominance.

Theorem 4. *Let P and Q be a generic pair of bounded experiments, with log-likelihood ratios taking values in $[-b, b]$. Assume P dominates Q in the Rényi order, and let $\eta \in (0,1)$ be provided by Lemma 2. Then $P^{\otimes n}$ Blackwell dominates $Q^{\otimes n}$ for all $n \geq n_0 = 8b^2\eta^{-3}$.*

The constant n_0 is decreasing in the parameter η. This fact follows from a logic analogous to the one behind Proposition 5: Larger values of η imply that the probability of unlikely, but very informative, signal realizations decreases at a much slower rate under the experiment $P^{\otimes n}$ than under $Q^{\otimes n}$, as the sample size n becomes large.

While simple, the constant n_0 is far from being tight. For example, our proof of Proposition 5 uses the Chebyshev inequality, which may be improved by a suitable application of the Berry-Esseen Theorem, at the cost of a more complex bound. It remains an open problem to develop more precise estimates.

[14] The comparison $\mathbb{P}\left[X_1^1 + \cdots + X_n^1 < na\right] \leq \mathbb{P}\left[Y_1^1 + \cdots + Y_n^1 < na\right]$ for all a in this range implies the desired result $\mathbb{P}\left[X_1^1 + \cdots + X_n^1 \leq na\right] \leq \mathbb{P}\left[Y_1^1 + \cdots + Y_n^1 \leq na\right]$, by a standard limit argument.

6 Discussion and Related Literature

Comparison of Experiments. Blackwell (1951, p. 101) posed the question of whether dominance of two experiments is equivalent to dominance of their n-fold repetitions. Stein (1951) and Torgersen (1970) provide early examples of two experiments that are not comparable in the Blackwell order, but are comparable in large samples.

Moscarini and Smith (2002) propose an alternative criterion for comparing repeated experiments. According to their notion, an experiment P dominates an experiment Q if for every decision problem with finitely many actions, there exists some n_0 such that the expected payoff achievable from observing $P^{\otimes n}$ is higher than that from observing $Q^{\otimes n}$ whenever $n \geq n_0$. This order is characterized by the *efficiency index* of an experiment, defined, in our notation, as the minimum over $t \in (0,1)$ of the function $\mathrm{e}^{(t-1)R_P^0(t)}$ (where a smaller index means a better experiment). There are two conceptual differences between the order studied in Moscarini and Smith and the large sample order that we characterize:

(i). While in Moscarini and Smith the number n_0 of repetitions is allowed to depend on the decision problem, dominance in large samples is a criterion for comparing experiments uniformly over decision problems, for fixed sample sizes. Thus the large sample order is conceptually closer to Blackwell dominance.[15]

(ii). The order proposed in Moscarini and Smith restricts attention to decision problems with *finitely* many actions, while dominance in the large sample order implies that observing $P^{\otimes n}$ is better that observing $Q^{\otimes n}$ for every decision problem.

Related to (ii), Azrieli (2014) shows that the Moscarini-Smith order is a strict extension of dominance in large samples. Perhaps surprisingly, this conclusion is reversed under a modification of their definition: It follows from our results that when extended to consider all decision problems, including problems with infinitely many actions, the Moscarini-Smith order over experiments (generically) coincides with the large sample order.[16]

Our notion of dominance in large samples is prior-free. In contrast, several authors (Kelly, 1956; Lindley, 1956; Cabrales, Gossner, and Serrano, 2013) have studied a complete ordering of experiments, indexed by the expected reduction of entropy from prior to posterior beliefs (i.e., mutual information between states and signals). We note that unlike Blackwell dominance, dominance in large samples does not guarantee a higher reduction of uncertainty given any prior belief.[17]

[15] Recent work by Hellman and Lehrer (2019) generalizes the Moscarini-Smith order to Markov (rather than i.i.d.) sequences of experiments.

[16] Consider the following variant of the Moscarini-Smith order: Say that P dominates Q if for *every* decision problem (with possibly infinitely many actions) there exists an n_0 such that the expected payoff achievable from $P^{\otimes n}$ is higher than that from $Q^{\otimes n}$ whenever $n \geq n_0$. Each Rényi divergence $R_P^\theta(t)$ corresponds to the expected payoff in some decision problem (see §5.1), and for such decision problems the ranking over repeated experiments is independent of the sample size n. Thus P dominates Q in this order only if P dominates Q in the Rényi order. By Theorem 1, P must then dominate Q in large samples.

[17] To see this, consider Example 2 above with parameters $\alpha = 0.1$ and $\beta = 0$. Then Proposition 2

Majorization and Quantum Information. Our work is related to the study of *majorization* in the quantum information literature. Majorization is a stochastic order commonly defined for distributions on countable sets. For distributions with a given support size, this order is closely related to the Blackwell order. Let $P = (\Omega, P_0, P_1)$ and $Q = (\Xi, Q_0, Q_1)$ be two experiments such that Ω and Ξ are finite and of the same size, and P_0 and Q_0 are the uniform distributions on Ω and Ξ. Then P Blackwell dominates Q if and only if P_1 majorizes Q_1 (see Torgersen, 1985, p. 264). This no longer holds when Ω and Ξ are of different sizes.

Motivated by questions in quantum information, Jensen (2019) asks the following question: Given two finitely supported distributions μ and ν, when does the n-fold product $\mu^{\times n} = \mu \times \cdots \times \mu$ majorize $\nu^{\times n}$ for all large n? He shows that for the case that μ and ν have *different* support sizes, the answer is given by the ranking of their Rényi entropies.[18] For the case of equal support size, Theorem 1 implies a similar result, which Jensen (2019, Remark 3.9) conjectures to be true. We prove his conjecture in §L in the appendix.

Fritz (2018) uses an abstract algebraic approach to prove a result that is complementary to Proposition 5. While Fritz's theorem does not require our genericity condition, the comparison of distributions is stated in terms of a notion of approximate stochastic dominance. A result similar to Proposition 5 (but without the η and the quantitative bound on n) appears as Lemma 2 in Aubrun and Nechita (2008), also in the context of majorization and quantum information theory.

Both Fritz (2018) and Jensen (2019), in their respective settings, ask a question in the spirit of our dominance ratio, and prove results that are similar to Proposition 3.

Experiments for Many States and Unbounded Experiments. Our analysis leaves open a number of questions. The most salient is the extension of Theorem 1, our characterization of dominance in large samples, to experiments with more than two states. In §K in the appendix, we identify a set of *necessary conditions* for large sample dominance. These conditions are expressed in terms of the moment generating function of the log-likelihood ratios—which generalizes the ranking of Rényi divergences in the two state case. While we conjecture this set of conditions to be also sufficient, our proof technique for sufficiency does not straightforwardly extend to more than two states. In particular, we do not know how to extend the reduction of Blackwell dominance to first-order stochastic dominance

ensures that the experiment P dominates Q in large samples. However, given a uniform prior, the residual uncertainty under P is calculated as the expected entropy of posterior beliefs, which is $\frac{1}{2}\log(2) \approx 0.346$. The residual uncertainty under Q is $-\alpha \log \alpha - (1-\alpha)\log(1-\alpha) \approx 0.325$, which is lower.

[18] As discussed above, majorization with different support sizes does not imply Blackwell dominance. Indeed, the ranking based on Rényi entropies is distinct from our ranking based on Rényi divergences unless the support sizes are equal. See §L in the appendix for details.

(Theorem 3).[19] With binary states we have been able to derive this simplification because one-dimensional convex (indirect utility) functions admit an one-parameter family of extremal rays. Going to higher dimensions, the difficulty is that "the extremal rays are too complex to be of service" (Jewitt, 2007).

Another extension for future work is to experiments with unbounded likelihood ratios. As we demonstrate in §J in the appendix, our characterization of the large sample order remains valid if the dominant experiment P is unbounded whereas the dominated experiment Q is bounded. The result also extends, under an additional assumption, to pairs of unbounded experiments whose Rényi divergences are finite. However, we do not know whether and how our result would generalize to the case of infinite Rényi divergences. The technical challenge is that large deviation estimates that are uniform across different thresholds typically require the moment generating function to be finite (so-called "Cramér's condition").[20]

[19] If such a reduction could be obtained, the remaining obstacle would be the characterization of first-order stochastic dominance between large i.i.d. sums of random *vectors*. This would require the development of large deviation estimates in higher dimensions (generalizing Lemma 3 in the appendix).

[20] Although Cramér's result that $\log \mathbb{P}[X_1 + \cdots + X_n > na] \sim -n \cdot K_X^*(a)$ remains true even when $K_X(t)$ can be infinite, as far as we know the proofs of this generalization do not deliver a quantitative lower bound similar to our Lemma 3. As a consequence, Cramér's approximation is not uniform across a.

Appendix

The structure of the appendix follows that of the paper. After reviewing large deviations theory, we complete the proof of Theorem 1 by supplying the proofs of Proposition 5, Lemma 1 and Lemma 2. We then provide proofs for our other results in the order in which they appeared.

A Large Deviations

For every bounded random variable X that is not a constant, we denote by $M_X(t) = \log \mathbb{E}[e^{tX}]$ and $K_X(t) = \log M_X(t)$ the moment and cumulant generating functions of X.

As is well known, M_X and K_X are strictly convex. We denote by

$$K_X^*(a) = \sup_{t \in \mathbb{R}} t \cdot a - K_X(t)$$

the Fenchel conjugate of K_X. For $a \in (\min[X], \max[X])$ the maximization problem has a unique solution, achieved at some $t \in \mathbb{R}$. This solution t is non-negative if and only if $a \geq \mathbb{E}[X]$. In addition, as $K_X(0) = 0$, $K_X^*(a) \geq 0 \cdot a - K_X(0) = 0$ is non-negative. The function $K_X^*(a)$ is continuous (in fact, analytic) wherever it is finite.

The well known Chernoff bound states that if $X, X_1, \ldots, X_n$ are an i.i.d. sequence, then

$$\mathbb{P}\left[X_1 + \cdots + X_n > na\right] \leq e^{-n \cdot K_X^*(a)} \quad \text{for all } a \geq \mathbb{E}[X].$$

The next proposition gives a lower bound for this probability.

Lemma 3. *Let $X, X_1, \ldots, X_n$ be an i.i.d. sequence taking values in $[-b, b]$. For all $\eta > 0$, $a \in [\min[X], \max[X] - \eta)$ and $n \geq 1$, it holds that*

$$\mathbb{P}\left[X_1 + \cdots + X_n > na\right] \geq e^{-n \cdot K_X^*(a+\eta)} \left(1 - \frac{4b^2}{n\eta^2}\right)$$

Proof. We first consider the case where $a \geq \mathbb{E}[X] - \eta/2$. Define t by

$$K_X'(t) = a + \eta/2,$$

so that $K_X^*(a + \eta/2) = (a + \eta/2) \cdot t - K_X(t)$. Such a t is a non-negative finite number, since $\mathbb{E}[X] \leq a + \eta/2 < \max[X]$.

Denote by ν the distribution of X, and let $\hat{X}$ be a real random variable whose distribution $\hat{\nu}$ is given by

$$\frac{d\hat{\nu}}{d\nu}(x) = \frac{e^{tx}}{\mathbb{E}[e^{tX}]} = e^{tx - K_X(t)}.$$

This construction ensures that $\hat{\nu}$ is also a probability measure, so that $\hat{X}$ is a well-defined random variable.

Note that

$$\mathbb{E}[\hat{X}] = \frac{\mathbb{E}[X\mathrm{e}^{tX}]}{\mathbb{E}[\mathrm{e}^{tX}]} = K_X'(t) = a + \eta/2,$$

and that the cumulant generating function of $\hat{X}$ is

$$K_{\hat{X}}(s) = \log \mathbb{E}[\mathrm{e}^{s\hat{X}}] = \log \mathbb{E}[\mathrm{e}^{tX - K_X(t)}\mathrm{e}^{sX}] = K_X(s+t) - K_X(t).$$

Now let $\hat{X}_1, \ldots, \hat{X}_n$ be i.i.d. copies of $\hat{X}$. Denote $S_n = X_1 + \cdots + X_n$ and $\hat{S}_n = \hat{X}_1 + \cdots + \hat{X}_n$. The cumulant generating function of $\hat{S}_n$ is

$$K_{\hat{S}_n}(s) = nK_{\hat{X}}(s) = n(K_X(s+t) - K_X(t)) = K_{S_n}(s+t) - K_{S_n}(t),$$

and so the Radon-Nikodym derivative between the distributions of $\hat{S}_n$ and S_n is $\mathrm{e}^{tx - K_{S_n}(t)} = \mathrm{e}^{tx - nK_X(t)}$. Hence

$$\begin{aligned}
\mathbb{P}\left[S_n > na\right] &= \mathbb{E}[\mathbb{1}_{\{S_n > na\}}] \\
&= \mathbb{E}\left[\mathrm{e}^{-t\hat{S}_n + nK_X(t)}\mathbb{1}_{\left\{\hat{S}_n > na\right\}}\right] \\
&= \mathrm{e}^{nK_X(t)} \cdot \mathbb{E}\left[\mathrm{e}^{-t\hat{S}_n}\mathbb{1}_{\left\{\hat{S}_n > na\right\}}\right].
\end{aligned}$$

The event $\{\hat{S}_n > na\}$ contains the event $\{n(a+\eta) > \hat{S}_n > na\}$, and so

$$\begin{aligned}
\mathbb{P}\left[S_n > na\right] &\geq \mathrm{e}^{nK_X(t)} \cdot \mathbb{E}\left[\mathrm{e}^{-t\hat{S}_n}\mathbb{1}_{\left\{n(a+\eta) > \hat{S}_n > na\right\}}\right] \\
&\geq \mathrm{e}^{nK_X(t) - tn(a+\eta)} \cdot \mathbb{E}\left[\mathbb{1}_{\left\{n(a+\eta) > \hat{S}_n > na\right\}}\right] \\
&= \mathrm{e}^{nK_X(t) - tn(a+\eta)} \cdot \mathbb{P}\left[n(a+\eta) > \hat{S}_n > na\right]
\end{aligned}$$

where the second inequality uses $t \geq 0$ and $\hat{S}_n < n(a+\eta)$ whenever $\mathbb{1}_{\left\{n(a+\eta) > \hat{S}_n > na\right\}} > 0$.

Now, $\hat{S}_n$ has expectation $n\mathbb{E}[\hat{X}] = n(a + \eta/2)$. Its variance is $n\,\mathrm{Var}[\hat{X}] \leq n\mathbb{E}[\hat{X}^2] \leq nb^2$, since $\hat{X}$ has the same support of X by construction. Therefore, by the Chebyshev inequality,

$$\mathbb{P}\left[n(a+\eta) > \hat{S}_n > na\right] = 1 - \mathbb{P}\left[|\hat{S}_n - \mathbb{E}[\hat{S}_n]| \geq n\eta/2\right] \geq 1 - \frac{nb^2}{(n\eta/2)^2} = 1 - \frac{4b^2}{n\eta^2}.$$

We have thus shown that

$$\mathbb{P}\left[S_n > na\right] \geq \mathrm{e}^{-n(t(a+\eta) - K_X(t))}\left(1 - \frac{4b^2}{n\eta^2}\right).$$

Now, by definition $K_X^*(a+\eta) \geq t(a+\eta) - K_X(t)$. Hence we arrive at

$$\mathbb{P}\left[S_n > na\right] \geq \mathrm{e}^{-n \cdot K_X^*(a+\eta)}\left(1 - \frac{4b^2}{n\eta^2}\right).$$

We turn to the case where $a < \mathbb{E}[X] - \eta/2$. In this case, we can directly apply the Chebyshev inequality and obtain

$$\mathbb{P}\left[S_n \le na\right] \le \mathbb{P}\left[S_n - \mathbb{E}[S_n] \le -n\eta/2\right] \le \frac{\mathrm{Var}[S_n]}{(n\eta/2)^2} = \frac{n\,\mathrm{Var}[X]}{(n\eta/2)^2} \le \frac{4b^2}{n\eta^2}.$$

Hence

$$\mathbb{P}\left[S_n > na\right] \ge 1 - \frac{4b^2}{n\eta^2}.$$

Since K_X^* is non-negative, we again have

$$\mathbb{P}\left[S_n > na\right] \ge \mathrm{e}^{-n\cdot K_X^*(a+\eta)}\left(1 - \frac{4b^2}{n\eta^2}\right).$$

This proves the lemma. □

A.1 Proof of Proposition 5

If $a < \min[X]$ then the statement holds since in (16) the LHS is equal to 1. Below we assume $a \ge \min[X]$. By assumption, $K_X^*(a+\eta)$ is finite, and hence $a + \eta < \max[X]$. We can thus apply Lemma 3 to X and conclude that for every $n \ge 1$,

$$\mathbb{P}\left[X_1 + \cdots + X_n > na\right] \ge \mathrm{e}^{-n\cdot K_X^*(a+\eta)}\left(1 - \frac{4b^2}{n\eta^2}\right).$$

By assumption we have that $K_Y^*(a) - \eta \ge K_X^*(a+\eta)$, and so

$$\begin{aligned}\mathbb{P}\left[X_1 + \cdots + X_n > na\right] &\ge \mathrm{e}^{-n\cdot K_Y^*(a)}\mathrm{e}^{n\eta}\left(1 - \frac{4b^2}{n\eta^2}\right)\\ &\ge \mathrm{e}^{-n\cdot K_Y^*(a)}(1+\eta)\left(1 - \frac{4b^2}{n\eta^2}\right)\end{aligned}$$

Hence, for $n \ge 4b^2(1+\eta)\eta^{-3}$,

$$\mathbb{P}\left[X_1 + \cdots + X_n > na\right] \ge \mathrm{e}^{-n\cdot K_Y^*(a)}.$$

On the other hand, since $a \ge \mathbb{E}[Y]$ by assumption, we have the Chernoff bound

$$\mathbb{P}\left[Y_1 + \cdots + Y_n > na\right] \le \mathrm{e}^{-n\cdot K_Y^*(a)}.$$

This proves the desired result (16).

B Proof of Lemma 1

An exponential distribution has probability density function that vanishes for negative u and equals e^{-u} for positive u. Thus $\tilde{F}_1$ and $\tilde{G}_1$ can be written as

$$\tilde{F}_1(a) = \int_0^\infty F_1(a+u)\mathrm{e}^{-u}\,\mathrm{d}u$$

and likewise

$$\tilde{G}_1(a) = \int_0^\infty G_1(a+u)\mathrm{e}^{-u}\,\mathrm{d}u.$$

Consider the first part of the lemma. Suppose $a \geq 0$, then by assumption $F_1(a+u) \leq G_1(a+u)$ for all $u \geq 0$, which implies $\tilde{F}_1(a) \leq \tilde{G}_1(a)$.

For the second part of the lemma, we will establish the following identities:

$$\tilde{F}_1(a) = \int_{-a}^\infty F_0(v)\mathrm{e}^{-v}\,\mathrm{d}v \quad \text{and} \quad \tilde{G}_1(a) = \int_{-a}^\infty G_0(v)\mathrm{e}^{-v}\,\mathrm{d}v. \tag{23}$$

Given this, the result would follow easily: If $F_0(v) \leq G_0(v)$ for all $v \geq 0$, then the above implies $\tilde{F}_1(a) \leq \tilde{G}_1(a)$ for all $a \leq 0$.

To show (23), we recall (11) and write

$$\tilde{F}_1(a) = \int_{-\infty}^a \mathrm{d}F_1(u) + \mathrm{e}^a \int_a^\infty \mathrm{e}^{-u}\,\mathrm{d}F_1(u). \tag{24}$$

The key observation is that $\mathrm{d}F_1(u) = -\mathrm{e}^u\,\mathrm{d}F_0(-u)$. Indeed, $\mathrm{d}F_1(u)$ is the density under state 1 that the log-likelihood ratio $\log(\mathrm{d}P_1/\mathrm{d}P_0)$ is equal to u, which is also the density under state 1 that the opposite log-likelihood ratio $\log(\mathrm{d}P_0/\mathrm{d}P_1)$ is equal to $-u$. By definition of the log-likelihood ratio, this density is scaled by a factor of e^{-u} when we change measure from state 1 to state 0.

Substituting $\mathrm{d}F_1(u) = -\mathrm{e}^u\,\mathrm{d}F_0(-u)$ into (24), we have

$$\tilde{F}_1(a) = \int_{-\infty}^a -\mathrm{e}^u\,\mathrm{d}F_0(-u) + \mathrm{e}^a \int_a^\infty -\,\mathrm{d}F_0(-u) = \int_{-a}^\infty \mathrm{e}^{-v}\,\mathrm{d}F_0(v) + \mathrm{e}^a F_0(-a),$$

where the second equality uses change of variable from u to $v = -u$. Integration by parts then yields (23) and completes the proof.

C Proof of Lemma 2

Fix θ, we will show the result holds for all sufficiently small positive η. Because P dominates Q in the Rényi order, and the pair of experiments is generic, the two log-likelihood ratios satisfy $0 < \mathbb{E}[Y^\theta] < \mathbb{E}[X^\theta]$ and $\max[Y^\theta] < \max[X^\theta]$.

For the first part of the lemma, consider the interval $A = [\mathbb{E}[X^\theta], \max[Y^\theta]]$. If it is empty (i.e., $\mathbb{E}[X^\theta] > \max[Y^\theta]$), the result trivially holds by choosing η small. Otherwise, consider any point $a \in A$. Since a is above the expectation of X^θ,

$$K^*_{X^\theta}(a) = \sup_{t \geq 0} ta - K_{X^\theta}(t).$$

And because $a < \max[X]$ the supremum is achieved at some finite $\hat{t} \geq 0$. Dominance in the Rényi order implies, by (18),

$$K^*_{X^\theta}(a) = \hat{t}a - K_{X^\theta}(\hat{t}) \leq \hat{t}a - K_{Y^\theta}(\hat{t}) \leq K^*_{Y^\theta}(a).$$

The first inequality can only hold equal if $\hat{t} = 0$ and $a = \mathbb{E}[X^\theta]$, but in that case the second inequality is strict because a is strictly above the expectation of Y^θ. Hence $K^*_{Y^\theta}(a) > K^*_{X^\theta}(a)$ for all a in A. Since A is compact and the two Fenchel transforms are continuous, we can find ε_1 positive such that $K^*_{Y^\theta}(a) - \varepsilon_1 > K^*_{X^\theta}(a)$ over all $a \in A$. Choosing positive ε_2 sufficiently small, we in fact have $K^*_{Y^\theta}(a) - \varepsilon_1 > K^*_{X^\theta}(a)$ for all a in the slightly bigger interval $[\mathbb{E}[X^\theta] - \varepsilon_2, \max[Y^\theta]]$. By uniform continuity, any small positive η satisfies $K^*_{X^\theta}(a+\eta) - K^*_{X^\theta}(a) < \frac{\varepsilon_1}{2}$ for all a in this interval. If in addition $\eta < \min\{\frac{\varepsilon_1}{2}, \varepsilon_2\}$, then

$$K^*_{Y^\theta}(a) - \eta > K^*_{Y^\theta}(a) - \varepsilon_1 + \frac{\varepsilon_1}{2} > K^*_{X^\theta}(a) + \frac{\varepsilon_1}{2} > K^*_{X^\theta}(a+\eta)$$

for all $a \in [\mathbb{E}[X^\theta] - \varepsilon_2, \max[Y^\theta]]$, and thus for $a \in [\mathbb{E}[X^\theta] - \eta, \max[Y^\theta]]$. This yields the desired result.

As for the second half, consider a point $a \in [0, \mathbb{E}[Y^\theta]]$. Since $a \leq \mathbb{E}[Y^\theta]$ and $a \geq 0 > \min[Y^\theta]$,[21] there exists a finite $\tilde{t} \leq 0$ such that $K^*_{Y^\theta}(a) = \tilde{t}a - K_{Y^\theta}(\tilde{t})$. This $\tilde{t}$ satisfies $K'_{Y^\theta}(\tilde{t}) = a$.

We now show that $\tilde{t} > -1$. The cumulant generating functions of Y^θ and $Y^{1-\theta}$ satisfy for all $t \in \mathbb{R}$ the relation

$$K_{Y^\theta}(t) = K_{Y^{1-\theta}}(-t-1)$$

and hence $K'_{Y^\theta}(-1) = -K'_{Y^{1-\theta}}(0) = -\mathbb{E}[Y^{1-\theta}] < 0$. Since $K'_{Y^\theta}(\tilde{t}) = a \geq 0$, and K'_{Y^θ} is increasing, we have $\tilde{t} \in (-1, 0]$. Dominance in the Rényi order then implies, by (19),

$$K^*_{Y^\theta}(a) = \tilde{t}a - K_{Y^\theta}(\tilde{t}) \leq \tilde{t}a - K_{X^\theta}(\tilde{t}) \leq K^*_{X^\theta}(a).$$

Similar to before, the first inequality can only hold equal if $\tilde{t} = 0$ and $a = \mathbb{E}[Y^\theta]$, but in that case the second inequality is strict because a is strictly below the expectation of X^θ. Hence $K^*_{Y^\theta}(a) < K^*_{X^\theta}(a)$ for all $a \in [0, \mathbb{E}[Y^\theta]]$. Using continuity as before, any sufficiently small η makes $K^*_{Y^\theta}(a-\eta) < K^*_{X^\theta}(a) - \eta$ hold for all a in the slightly bigger interval $[0, \mathbb{E}[Y^\theta] + \eta]$. Hence the lemma holds.

D Proof of Proposition 1

Let p_1 (resp. p_3) be the essential minimum (resp. maximum) of the distribution π of posterior beliefs induced by P. Since the support of π has at least 3 points, we can find $p_2 \in (p_1, p_3)$ such that $\pi([p_1, p_2]) > \pi(\{p_1\})$ and $\pi([p_2, p_3]) > \pi(\{p_3\})$.

[21] The latter holds because $\max[Y^{1-\theta}] \geq \mathbb{E}[Y^{1-\theta}] > 0$, and by definition $\min[Y^\theta] = -\max[Y^{1-\theta}]$.

We use this p_2 to construct an experiment Q which has signal space $\{0,1\}$, and which is a garbling of P. Specifically, if a signal realization under P leads to posterior belief below p_2, the garbled signal is 0. If the posterior belief under P is above p_2, the garbled signal is 1. Finally, if the posterior belief is exactly p_2, we let the garbled signal be 0 or 1 with equal probabilities.

Since $\pi([p_1,p_2]) > \pi(\{p_1\})$, the signal realization "0" under experiment Q induces a posterior belief that is strictly bigger than p_1, and smaller than p_2. Likewise, the signal realization "1" induces a belief strictly smaller than p_3, and bigger than p_2. Thus P and Q form a generic pair, and the distribution τ of posterior beliefs under Q is a strict mean-preserving contraction of π. We now recall that the Rényi divergences are derived from strictly convex indirect utility functions $u(p) = -p^t(1-p)^{1-t}$ for $0 < t < 1$ and $v(p) = p^t(1-p)^{1-t}$ for $t > 1$. Thus, $R_P^\theta(t) > R_Q^\theta(t)$ for all $\theta \in \{0,1\}$ and $t > 0$.

We will perturb Q to be a slightly more informative experiment Q', such that P still dominates Q' in the Rényi order but not in the Blackwell order. For this, suppose that under Q the posterior belief equals $q_1 \in (p_1,p_2)$ with some probability λ, and equals $q_2 \in (p_2,p_3)$ with remaining probability. Choose any small positive number ε, and let Q' be another binary experiment inducing the posterior belief $q_1 - \varepsilon(1-\lambda)$ with probability λ, and inducing the posterior belief $q_2 + \varepsilon\lambda$ otherwise. Such an experiment exists, because the expected posterior belief is unchanged. By continuity, $R_P^\theta(t) > R_{Q'}^\theta(t)$ still holds when ε is sufficiently small.[22] Since P and Q' also form a generic pair, Theorem 1 shows that P dominates Q' in large samples.

It remains to prove that P does not dominate Q' according to Blackwell. Consider a decision problem where the prior is uniform, the set of actions is $\{0,1\}$, and payoffs are given by $u(\theta = a = 0) = p_2$, $u(\theta = a = 1) = 1 - p_2$ and $u(\theta \neq a) = 0$. The indirect utility function is $v(p) = \max\{(1-p)p_2,\ p(1-p_2)\}$, which is piece-wise linear on $[0,p_2]$ and $[p_2,1]$ but convex at p_2. Recall that in constructing the garbling from P to Q, those posterior beliefs under P that are below p_2 are "averaged" into the single posterior belief q_1 under Q, and those above p_2 are averaged into the belief q_2. Thus Q achieves the same expected utility in this decision problem as P (despite being a garbling). Nevertheless, observe that Q' achieves higher expected utility in this decision problem than Q.[23] Hence Q' achieves higher expected utility than P, implying that it is not Blackwell dominated.

[22] Using the relation between $R_P^0(t)$ and $R_P^1(1-t)$, it suffices to show $R_P^\theta(t) > R_{Q'}^\theta(t)$ for $\theta \in \{0,1\}$ and $t \geq 1/2$. Fixing a large T, then by uniform continuity, $R_P^\theta(t) > R_Q^\theta(t)$ implies $R_P^\theta(t) > R_{Q'}^\theta(t)$ for $t \in [1/2,T]$ when ε is small. This also holds for t large, because as $t \to \infty$ the growth rate of the Rényi divergences are governed by the maximum of likelihood ratios, which is larger under P than under Q'.

[23] Formally, since $q_1 - \varepsilon(1-\lambda) < q_1 < p_2$ and $q_2 + \varepsilon\lambda > q_2 > p_2$, it holds that

$$\lambda \cdot v(q_1 - \varepsilon(1-\lambda)) + (1-\lambda) \cdot v(q_2 + \varepsilon\lambda) > \lambda \cdot v(q_1) + (1-\lambda) \cdot v(q_2).$$

E Proof of Proposition 2

It is easily checked that the condition $R^1_P(1/2) > R^1_Q(1/2)$ reduces to

$$\sqrt{\alpha(1-\alpha)} > \sqrt{\beta(\frac{1}{2}-\beta)} + \frac{1}{4}. \tag{25}$$

Since the experiments form a generic pair, by Theorem 1, we just need to check dominance in the Rényi order. Equivalently, we need to show

$$(\frac{1}{2}-\beta)^r\beta^{1-r} + (\frac{1}{2}-\beta)^{1-r}\beta^r + \frac{1}{2} < (1-\alpha)^r\alpha^{1-r} + (1-\alpha)^{1-r}\alpha^r, \quad \forall 0 < r < 1; \tag{26}$$

$$(\frac{1}{2}-\beta)^r\beta^{1-r} + (\frac{1}{2}-\beta)^{1-r}\beta^r + \frac{1}{2} > (1-\alpha)^r\alpha^{1-r} + (1-\alpha)^{1-r}\alpha^r, \quad \forall r < 0 \text{ or } r > 1; \tag{27}$$

$$\beta \cdot \ln(\frac{\beta}{\frac{1}{2}-\beta}) + (\frac{1}{2}-\beta) \cdot \ln(\frac{\frac{1}{2}-\beta}{\beta}) > \alpha \cdot \ln(\frac{\alpha}{1-\alpha}) + (1-\alpha) \cdot \ln(\frac{1-\alpha}{\alpha}). \tag{28}$$

To prove these, it suffices to consider the α that makes (25) hold with equality.[24] We will show that the above inequalities hold for this particular α, except that (26) holds equal at $r = \frac{1}{2}$. Let us define the following function

$$\Delta(r) := (\frac{1}{2}-\beta)^r\beta^{1-r} + (\frac{1}{2}-\beta)^{1-r}\beta^r + \frac{1}{2} - (1-\alpha)^r\alpha^{1-r} - (1-\alpha)^{1-r}\alpha^r.$$

When (25) holds with equality, we have $\Delta(0) = \Delta(\frac{1}{2}) = \Delta(1) = 0$. Thus Δ has roots at 0, 1 as well as a double-root at $\frac{1}{2}$. But since Δ is a weighted sum of 4 exponential functions plus a constant, it has at most 4 roots (counting multiplicity).[25] Hence these are the only roots, and we deduce that the function Δ has constant sign on each of the intervals $(-\infty, 0), (0, \frac{1}{2}), (\frac{1}{2}, 1), (1, \infty)$.

Now observe that since $2\beta < \alpha \leq \frac{1}{2}$, it holds that $\frac{1/2-\beta}{\beta} > \frac{1-\alpha}{\alpha} > 1$. It is then easy to check that $\Delta(r) \to \infty$ as $r \to \infty$. Thus $\Delta(r)$ is strictly positive for $r \in (1, \infty)$. As $\Delta(1) = 0$, its derivative is weakly positive. But recall that we have enumerated the 4 roots of Δ. So Δ cannot have a double-root at $r = 1$, and it follows that $\Delta'(1)$ is strictly positive. Hence (28) holds.

Note that $\Delta'(1) > 0$ and $\Delta(1) = 0$ also implies $\Delta(1-\varepsilon) < 0$. Thus Δ is negative on $(\frac{1}{2}, 1)$. A symmetric argument shows that Δ is positive on $(-\infty, 0)$ and negative on $(0, \frac{1}{2})$. Hence (26) and (27) both hold, completing the proof.

F Proof of Proposition 3

Denote $r = \inf_{\theta,t} \frac{R^\theta_P(t)}{R^\theta_Q(t)}$. We would like to show that $P/Q = r$. Let n, m be such that $P^{\otimes n} \succeq Q^{\otimes m}$. Then, since ranking of the Rényi divergences is a necessary condition for

[24] It is clear that the inequalities are easier to satisfy when α increases in the range $[0, \frac{1}{2}]$.

[25] This follows from Rolle's Theorem and an induction argument.

Blackwell dominance, and by the additivity of Rényi divergences, $n \cdot R_P^\theta(t) \geq m \cdot R_Q^\theta(t)$ for all $\theta \in \{0, 1\}$ and $t > 0$. Thus any such m/n is bounded above by r, and so $P/Q \leq r$.

In the other direction, take any rational number $m/n < r$. Then, again by the additivity of the Rényi divergences, $P^{\otimes n}$ dominates $Q^{\otimes m}$ in the Rényi order. Furthermore, the fact that $\lim_{t\to\infty} \frac{R_P^\theta(t)}{R_Q^\theta(t)} > m/n$ implies the pair $P^{\otimes n}$ and $Q^{\otimes m}$ is generic. Therefore, by Theorem 1, we have that for some k large enough, $P^{\otimes nk} \succeq Q^{\otimes mk}$. Thus $P/Q \geq mk/nk = m/n$. Since this holds for every rational m/n that is less than r, we can conclude that $P/Q \geq r$. Finally, note that each of the functions R_P^θ and R_Q^θ are positive, increasing and bounded on $(0, \infty)$. Furthermore, using

$$\frac{R_P^\theta(t)}{R_Q^\theta(t)} = \frac{R_P^{1-\theta}(1-t)}{R_Q^{1-\theta}(1-t)},$$

for $t \in (0, 1)$, we can rewrite

$$P/Q = \inf_{\substack{\theta\in\{0,1\},\\ t>0}} \frac{R_P^\theta(t)}{R_Q^\theta(t)} = \inf_{\substack{\theta\in\{0,1\},\\ t\geq\frac{1}{2}}} \frac{R_P^\theta(t)}{R_Q^\theta(t)}.$$

Recall that $R_P^\theta(t), R_Q^\theta(t)$ are positive, continuous in t and approach $\max[X^\theta]$ and $\max[Y^\theta]$ as $t \to \infty$. Thus a compactness argument shows that P/Q is always positive.

References

C. D. Aliprantis and K. Border. *Infinite dimensional analysis*. Springer, 2006.

G. Aubrun and I. Nechita. Catalytic majorization and ℓ_p norms. *Communications in Mathematical Physics*, 278(1):133–144, 2008.

Y. Azrieli. Comment on "the law of large demand for information". *Econometrica*, 82(1): 415–423, 2014.

D. Blackwell. Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 93–102. University of California Press, 1951.

D. Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2):265–272, 1953.

D. A. Blackwell and M. A. Girshick. *Theory of games and statistical decisions*. Courier Corporation, 1979.

V. I. Bogachev. *Measure theory*, volume 1. Springer Science & Business Media, 2007.

H. F. Bohnenblust, L. S. Shapley, and S. Sherman. Reconnaissance in game theory. 1949.

A. Cabrales, O. Gossner, and R. Serrano. Entropy and the value of information for investors. *American Economic Review*, 103(1):360–377, 2013.

H. Cramér. Sur un nouveau théoreme-limite de la théorie des probabilités. *Actual. Sci. Ind.*, 736:5–23, 1938.

F. Critchley, P. Marriott, and M. Salmon. On the differential geometry of the wald test with nonlinear restrictions. *Econometrica*, 64(5):1213–1222, 1996.

I. Csiszár. Axiomatic characterizations of information measures. *Entropy*, 10(3):261–273, 2008.

R. Duan, Y. Feng, X. Li, and M. Ying. Multiple-copy entanglement transformation and entanglement catalysis. *Physical Review A*, 71(4):042319, 2005.

T. Fritz. Resource convertibility and ordered commutative monoids. *Mathematical Structures in Computer Science*, 27(6):850–938, 2017.

T. Fritz. A generalization of Strassen's Positivstellensatz and its application to large deviation theory. *arXiv preprint arXiv:1810.08667v3*, 2018.

Z. Hellman and E. Lehrer. Valuing information by repeated markov signals. Working Paper, 2019.

Y. Hong and H. White. Asymptotic distribution theory for nonparametric entropy measures of serial dependence. *Econometrica*, 73(3):837–901, 2005.

R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Reviews of modern physics*, 81(2):865, 2009.

A. K. Jensen. Asymptotic majorization of finite probability distributions. *IEEE Transactions on Information Theory*, 65(12):8131–8139, 2019.

I. Jewitt. Information order in decision and agency problems. Technical report, Nuffield College, 2007.

L. Kantorovich. On the moment problem for a finite interval. In *Dokl. Akad. Nauk SSSR*, volume 14, pages 531–537, 1937.

J. L. Kelly. A new interpretation of information rate. *IRE Transactions on Information Theory*, 2(3):185–189, 1956.

Y. Kitamura and M. Stutzer. An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, 65(4):861–874, 1997.

Y. Kitamura, T. Otsu, and K. Evdokimov. Robustness, infinitesimal neighborhoods, and moment restrictions. *Econometrica*, 81(3):1185–1201, 2013.

A. Krishnamurthy, K. Kandasamy, B. Poczos, and L. Wasserman. Nonparametric estimation of renyi divergence and friends. In *International Conference on Machine Learning*, pages 919–927, 2014.

F. Liese and I. Vajda. On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory*, 52(10):4394–4412, 2006.

D. V. Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, 1956.

G. Moscarini and L. Smith. The law of large demand for information. *Econometrica*, 70 (6):2351–2366, 2002.

B. Póczos, L. Xiong, and J. Schneider. Nonparametric divergence estimation with applications to machine learning on distributions. *arXiv preprint arXiv:1202.3758*, 2012.

L. Pomatto, P. Strack, and O. Tamuz. The cost of information. *arXiv preprint arXiv:1812.04211*, 2018.

L. Pomatto, P. Strack, and O. Tamuz. Stochastic dominance under independent noise. *arXiv preprint arXiv:1807.06927*, 2019.

A. Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California, 1961.

T. Sawa. Information criteria for discriminating among alternative regression models. *Econometrica*, 46(6):1273–1291, 1978.

L. Smith and P. Sørensen. Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398, 2000.

C. Stein. Notes on the comparison of experiments. *University of Chicago*, 1951.

E. Torgersen. Majorization and approximate majorization for families of measures, applications to local comparison of experiments and the theory of majorization of vectors in r n (schur convexity). In *Linear Statistical Inference*, pages 259–310. Springer, 1985.

E. Torgersen. *Comparison of statistical experiments*, volume 36. Cambridge University Press, 1991.

E. N. Torgersen. Comparison of experiments when the parameter space is finite. *Probability Theory and Related Fields*, 16(3):219–249, 1970.

A. Ullah. Uses of entropy and divergence measures for evaluating econometric approximations and inference. *Journal of Econometrics*, 107(1-2):313–326, 2002.

H. White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1): 1–25, 1982.

Online Appendix

G Proof of Proposition 4

That (i) implies (ii) follows from the fact that Rényi divergences are monotone in the Blackwell order, and additive with respect to independent experiments.

To show (ii) implies (i), we introduce some notation. Given two experiments $P = (\Omega, P_0, P_1)$ and $Q = (\Xi, Q_0, Q_1)$, for each $\alpha \in [0,1]$ we denote by $\alpha P + (1-\alpha)Q = (\Psi, M_0, M_1)$ the mixed experiment where the sample space is the disjoint union $\Psi = \Omega \sqcup \Xi$ endowed with the corresponding σ-algebra, and the measures M_0, M_1 satisfy for every measurable $E \subseteq \Psi$

$$M_\theta(E) = \alpha P_\theta(E \cap \Omega) + (1-\alpha) Q_\theta(E \cap \Xi).$$

Intuitively, the mixed experiment corresponds to a randomized experiment where P is carried out with probability α and Q with probability $1-\alpha$. The mixture operation and the product operation satisfy $(\alpha P + (1-\alpha)Q) \otimes R = \alpha(P \otimes R) + (1-\alpha)(Q \otimes R)$.

Now suppose P dominates Q in the Rényi order, then by Theorem 1, P dominates Q in the large sample order. The next lemma concludes the proof.

Lemma 4. *Let P, Q be bounded experiments such that P dominates Q in the large sample order. Then there exists a bounded experiment R such that $P \otimes R$ Blackwell dominates $Q \otimes R$.*

This lemma replicates a more general statement that appears in Duan et al. (2005); Fritz (2017).

Proof of Lemma 4. Assume $P^{\otimes n} \succeq Q^{\otimes n}$. Let

$$R = \frac{1}{n}\left(Q^{\otimes n} + P \otimes Q^{\otimes(n-1)} + P^{\otimes 2} \otimes Q^{\otimes(n-2)} + \cdots + P^{\otimes(n-2)} \otimes Q^{\otimes 2} + P^{\otimes(n-1)} \otimes Q\right).$$

Then

$$\begin{aligned}
P \otimes R &= P \otimes \frac{1}{n}\left(Q^{\otimes n} + P \otimes Q^{\otimes(n-1)} + \cdots + P^{\otimes(n-2)} \otimes Q^{\otimes 2} + P^{\otimes(n-1)} \otimes Q\right) \\
&= \frac{1}{n}\left(P \otimes Q^{\otimes n} + P^{\otimes 2} \otimes Q^{\otimes(n-1)} + \cdots + P^{\otimes(n-1)} \otimes Q^{\otimes 2} + P^{\otimes n} \otimes Q\right) \\
&\succeq \frac{1}{n}\left(P \otimes Q^{\otimes n} + P^{\otimes 2} \otimes Q^{\otimes(n-1)} + \cdots + P^{\otimes(n-1)} \otimes Q^{\otimes 2} + Q^{\otimes(n+1)}\right) \\
&= Q \otimes \frac{1}{n}\left(Q^{\otimes n} + P \otimes Q^{\otimes(n-1)} + \cdots + P^{\otimes(n-1)} \otimes Q\right) \\
&= Q \otimes R,
\end{aligned}$$

where the middle step uses the assumption $P^{\otimes n} \succeq Q^{\otimes n}$, so that $P^{\otimes n} \otimes Q \succeq Q^{\otimes(n+1)}$. □

H Proof of Theorem 2

Throughout this section, we denote by D an additive divergence that satisfies the data-processing inequality and is finite on bounded experiments.

Lemma 5. *If a bounded experiment $P = (\Omega, P_0, P_1)$ dominates another bounded experiment $Q = (\Xi, Q_0, Q_1)$ in the Blackwell order, then $D(P_0, P_1) \geq D(Q_0, Q_1)$.*

Proof. By Blackwell's Theorem there exists a measurable function $\sigma\colon \Omega \to \Delta(\Xi)$ such that $Q_\theta(A) = \int \sigma(\omega)(A)\,\mathrm{d}P_\theta(\omega)$ for every measurable $A \subseteq \Xi$ and every θ. Let λ be the Lebesgue measure on $[0,1]$. Since Ω and Ξ are Polish spaces, there exists a measurable function $f\colon \Omega \times [0,1] \to \Xi$ such that for every $\omega \in \Omega$, $\sigma(\omega) = f(\omega, \cdot)_*(\lambda)$, where $f(\omega,\cdot)_*(\lambda)$ is the push-forward of λ induced by the function $f(\omega,\cdot)$ (see, for example, Proposition 10.7.6 in Bogachev, 2007). Hence,

$$Q_\theta(A) = \int \lambda(\{t \in [0,1] : f(\omega,t) \in A\})\,\mathrm{d}P_\theta(\omega) = f_*(P_\theta \times \lambda)(A)$$

where now $f_*(P_\theta \times \lambda)$ is the pushforward of $P_\theta \times \lambda$ induced by f. Being a divergence, D satisfies $D(\lambda,\lambda) = 0$. Moreover, by additivity, $D(P_0 \times \lambda, P_1 \times \lambda) = D(P_0, P_1)$. The data processing inequality then implies $D(P_0, P_1) = D(P_0 \times \lambda, P_1 \times \lambda) \geq D(Q_0, Q_1)$. □

Lemma 6. *If the bounded experiments $P = (P_0, P_1)$ and $Q = (Q_0, Q_1)$ satisfy $R^\theta_P(t) \geq R^\theta_Q(t)$ for every $t > 0$ and $\theta \in \{0,1\}$, then $D(P_0, P_1) \geq D(Q_0, Q_1)$.*

Proof. Suppose first that the strict inequality $R^\theta_P(t) > R^\theta_Q(t)$ holds for every $t > 0$, including at the limit $t = \infty$ (corresponding to the genericity assumption in the main text). Then, by Theorem 1 there exists n such that $P^{\otimes n}$ dominates $Q^{\otimes n}$ in the Blackwell order. Hence, by applying the previous lemma and by additivity, we obtain

$$nD(P_0, P_1) = D(P_0^n, P_1^n) \geq D(Q_0^n, Q_1^n) = nD(Q_0, Q_1).$$

More generally, suppose we only have the weak inequality $R^\theta_P(t) \geq R^\theta_Q(t)$ for $t > 0$. Fix a bounded and non-trivial experiment $S = (S_0, S_1)$. Then, for every $k \in \mathbb{N}$ we have

$$R^\theta_{P^{\otimes k} \otimes S}(t) = kR^\theta_P(t) + R^\theta_S(t) > kR^\theta_Q(t) = R^\theta_{Q^{\otimes k}}(t)$$

for every $t \in (0, \infty]$ and $\theta \in \{0,1\}$. Given what we just proved, it follows that

$$D(P_0^k \times S_0, P_1^k \times S_1) \geq D(Q_0^k, Q_1^k).$$

By additivity, $D(P_0, P_1) + \frac{1}{k} D(S_0, S_1) \geq D(Q_0, Q_1)$. Since this holds for every k and $D(S_0, S_1)$ is finite, the proof is concluded. □

Let $\overline{\mathbb{R}} = [-\infty, \infty]$ be the extended real line. Given a bounded experiment P we define the function $H_P \colon \overline{\mathbb{R}} \to \mathbb{R}$ as

$$H_P(t) = \begin{cases} R_P^1(t) & \text{if } \ t \geq 1/2 \\ R_P^0(1-t) & \text{if } \ t \leq 1/2 \end{cases}$$

Recall that the Rényi divergences of an experiment P satisfy the relation $(1-t)R_P^1(t) = tR_P^0(1-t)$. This implies that the function H_P is well defined, continuous, and bounded. It is a convenient representation of the Rényi divergences that retains the main properties of the latter, and has the advantage of being strictly positive whenever P is nontrivial. Since $H_P(t)$ is continuous and has a compact domain, it is furthermore bounded away from 0. The functional $P \mapsto H_P$ satisfies two additional properties. An experiment P dominates an experiment Q in the Rényi order if and only if $H_P(t) > H_Q(t)$ for every t. Moreover, the functional is additive: $H_{P \otimes Q}(t) = H_P(t) + H_Q(t)$ for every t.

Thus, to prove Theorem 2 it suffices to show that under the hypotheses of the theorem there exists a finite measure m on $\overline{\mathbb{R}}$ such that for every bounded pair of measures P_0, P_1

$$D(P_0, P_1) = \int_{\overline{\mathbb{R}}} H_P(t) \, \mathrm{d}m(t)$$

where P is the experiment (P_0, P_1). The theorem's conclusion (7) follows easily from this by setting $\mathrm{d}m_0(t) = -\mathrm{d}m(1-t)$ and $\mathrm{d}m_1(t) = \mathrm{d}m(t)$ for $t \geq \frac{1}{2}$.

Let $C(\overline{\mathbb{R}})$ be the space of continuous functions defined over the compact set $\overline{\mathbb{R}}$. Each function H_P belongs to $C(\overline{\mathbb{R}})$. Consider the set

$$\mathcal{H} = \{H_P : P \text{ is a bounded experiment}\} \subseteq C(\overline{\mathbb{R}}).$$

By Lemma 6, if $H_P = H_Q$ then $D(P_0, P_1) = D(Q_0, Q_1)$. Thus there exists a map $F \colon \mathcal{H} \to \mathbb{R}$ such that $D(P_0, P_1) = F(H_P)$.

By Lemma 6 the functional F is monotone. It is moreover additive: Given two experiments P and Q, the additivity of D and the additivity of $P \mapsto H_P$ imply

$$\begin{aligned} F(H_P) + F(H_Q) &= D(P_0, P_1) + D(Q_0, Q_1) \\ &= D(P_0 \times Q_0, P_1 \times Q_1) \\ &= F(H_{P \otimes Q}) \\ &= F(H_P + H_Q). \end{aligned}$$

Next, we define $\mathrm{cone}_{\mathbb{Q}}(\mathcal{H}) = \{\sum_{i=1}^n \alpha_i H_{P^i} : \alpha_i \in \mathbb{Q}_+, P^i \text{ is a bounded experiment}\}$ to be the rational cone generated by $\mathcal{H}$, where coefficients (α_i) are positive rational numbers. Similarly define

$$\mathrm{cone}(\mathcal{H}) = \left\{ \sum_{i=1}^n \alpha_i H_{P^i} : \alpha_i \in \mathbb{R}_+, P^i \text{ is a bounded experiment} \right\}$$

to be the cone generated by $\mathcal{H}$, where coefficients can be all positive numbers. Below we extend the functional F from $\mathcal{H}$ to $\text{cone}_{\mathbb{Q}}(\mathcal{H})$ and then to $\text{cone}(\mathcal{H})$.

Because $P \mapsto H_P$ is additive, $\mathcal{H}$ is itself closed under addition. This implies

$$\text{cone}_{\mathbb{Q}}(\mathcal{H}) = \bigcup_{n \geq 1} \frac{1}{n} \mathcal{H}.$$

Define $G\colon \text{cone}_{\mathbb{Q}}(\mathcal{H}) \to \mathbb{R}$ as $G(\frac{1}{n} H_P) = \frac{1}{n} F(H_P)$. The functional G is well-defined: If $\frac{1}{n} H_P = \frac{1}{m} H_Q$ then $H_{P^{\otimes m}} = m H_P = n H_Q = H_{Q^{\otimes n}}$, which implies $m F(H_P) = n F(H_Q)$ by the additivity of F. Similarly, G inherits the monotonicity and additivity of F on the larger domain $\text{cone}_{\mathbb{Q}}(\mathcal{H})$.

We now show G is a Lipschitz functional, where we endow the space $C(\overline{\mathbb{R}})$ with the sup norm. Let S_0 be a nontrivial experiment, so that $H_{S_0}(t)$ is positive and in fact bounded away from 0 for every t. By letting $S = S_0^{\otimes k}$ for large k, we obtain that $H_S(t) > 1$ for every t. Given two functions $f, \hat{f} \in \text{cone}_{\mathbb{Q}}(\mathcal{H})$, we have the pointwise comparison

$$f(t) \leq \hat{f}(t) + \|f - \hat{f}\| \times H_S(t).$$

Let $r > \|f - \hat{f}\|$ be a rational number. The additivity and the monotonicity of G imply

$$G(f) \leq G(\hat{f} + r H_S) = G(\hat{f}) + r G(H_S).$$

Symmetrically $G(\hat{f}) \leq G(f + r H_S) = G(f) + r G(H_S)$, so that $|G(f) - G(\hat{f})| \leq r G(H_S)$. By taking the limit $r \to \|f - \hat{f}\|$ we obtain that G is Lipschitz with Lipschitz constant $G(H_S) < \infty$, i.e.

$$|G(f) - G(\hat{f})| \leq \|f - \hat{f}\| \cdot G(H_S).$$

Thus G can be extended to a Lipschitz functional $\overline{G}$ defined on the closure of $\text{cone}_{\mathbb{Q}}(\mathcal{H})$, which contains $\text{cone}(\mathcal{H})$.

We now verify that $\overline{G}$ is still monotone on $\text{cone}(\mathcal{H})$. Let $f \geq \hat{f}$ be two functions in $\text{cone}(\mathcal{H})$, and take any two sequences $\{\frac{1}{p_n} H_{P_n}\}$ and $\{\frac{1}{q_n} H_{Q_n}\}$ in $\text{cone}_{\mathbb{Q}}(\mathcal{H})$ that converge to f and $\hat{f}$ as $n \to \infty$. For any positive integer m, convergence in the sup-norm implies $\frac{1}{p_n} H_{P_n} \geq f - \frac{1}{2m} H_S$ for all large n, where S is the experiment with $H_S > 1$ everywhere. Similarly $\frac{1}{q_n} H_{Q_n} \leq \hat{f} + \frac{1}{2m} H_S$. Since $f \geq \hat{f}$, we thus have $\frac{1}{p_n} H_{P_n} \geq \frac{1}{q_n} H_{Q_n} - \frac{1}{m} H_S$ for all large n. By monotonicity and additivity of G, $G(\frac{1}{p_n} H_{P_n}) \geq G(\frac{1}{q_n} H_{Q_n}) - \frac{1}{m} G(H_S)$, which implies $\overline{G}(f) \geq \overline{G}(\hat{f}) - \frac{1}{m} G(H_S)$ by taking $n \to \infty$. As m is arbitrary, we have shown that $\overline{G}$ is monotonic.

We show $\overline{G}$ is additive and satisfies $\overline{G}(af + b\hat{f}) = a\overline{G}(f) + b\overline{G}(\hat{f})$ for any functions $f, \hat{f} \in \text{cone}(\mathcal{H})$ and $a, b \in \mathbb{R}_+$. To show this, first suppose a, b are rational numbers. Consider $\{\frac{1}{p_n} H_{P_n}\} \to f$ and $\{\frac{1}{q_n} H_{Q_n}\} \to \hat{f}$ as above, where f need not be bigger than $\hat{f}$.

Then the sequence of functions $\{\frac{a}{p_n}H_{P_n} + \frac{b}{q_n}H_{Q_n}\} \in \text{cone}_{\mathbb{Q}}(\mathcal{H})$ converges to $af + b\hat{f}$. It follows that

$$\begin{aligned}\overline{G}(af + b\hat{f}) &= \lim_{n\to\infty} G(\frac{a}{p_n}H_{P_n} + \frac{b}{q_n}H_{Q_n}) \\ &= a \cdot \lim_{n\to\infty} G(\frac{1}{p_n}H_{P_n}) + b \cdot \lim_{n\to\infty} G(\frac{1}{q_n}H_{Q_n}) = a \cdot \overline{G}(f) + b \cdot \overline{G}(\hat{f}).\end{aligned}$$

If a, b are real numbers, we can deduce the same result by the Lipschitz property of $\overline{G}$.

Consider next $V = \text{cone}(\mathcal{H}) - \text{cone}(\mathcal{H})$, which is vector subspace of $C(\overline{\mathbb{R}})$. $\overline{G}$ can be further extended to a functional $I\colon V \to \mathbb{R}$, defined as

$$I(M_1 - M_2) = \overline{G}(M_1) - \overline{G}(M_2)$$

for all $M_1, M_2 \in \text{cone}(\mathcal{H})$. The functional I is well defined and linear because $\overline{G}$ is affine. Moreover, by monotonicity of $\overline{G}$, $I(f) \geq 0$ for any non-negative function $f \in V$.

The following theorem, a generalization of the Hahn-Banach Theorem (see, e.g., Theorem 8.32 in Aliprantis and Border, 2006), shows that I can be further extended to a positive linear functional on the entire space $C(\overline{\mathbb{R}})$:

Theorem 5 (Kantorovich (1937))**.** *Let V be a vector subspace of $C(\overline{\mathbb{R}})$ with the property that for every $f \in C(\overline{\mathbb{R}})$ there exists a function $g \in V$ such that $g \geq f$. Then every positive linear functional on V extends to a positive linear functional on $C(\overline{\mathbb{R}})$.*

The "majorization" condition $g \geq f$ is satisfied because every function in $C(\overline{\mathbb{R}})$ is bounded by some n, and V contains the function nH_S which takes values greater than n everywhere.

To summarize, we have obtained a positive linear functional J defined on $C(\overline{\mathbb{R}})$ that extends the original functional $F(H_P) = D(P_0, P_1)$. By the Riesz Representation Theorem for positive linear functionals over spaces of continuous functions on compact sets, we conclude that $J(f) = \int_{\overline{\mathbb{R}}} f(t)\,\mathrm{d}m(t)$ for some finite measure m. Hence $D(P_0, P_1) = F(H_P) = J(H_P)$ is an integral of the Rényi divergences of P, completing the proof of Theorem 2.

I Necessity of the Genericity Assumption

Here we present examples to show that Theorem 1 does not hold without the genericity assumption.

Consider the experiments P and Q described in Example 2 in §3.1. Fix $\alpha = \frac{1}{4}$ and $\beta = \frac{1}{16}$, which satisfy (25). Then by Proposition 2, P dominates Q in large samples.

We will perturb these two experiments by adding another signal realization (to each experiment) which strongly indicates the true state is 1. The perturbed conditional probabilities are given below:

$\tilde{P}$:

θ	x_0	x_1	x_2	x_3
0	ε	$\frac{1}{16}$	$\frac{1}{2}$	$\frac{7}{16}-\varepsilon$
1	100ε	$\frac{7}{16}$	$\frac{1}{2}$	$\frac{1}{16}-100\varepsilon$

$\tilde{Q}$:

θ	y_0	y_1	y_2
0	ε	$\frac{1}{4}$	$\frac{3}{4}-\varepsilon$
1	100ε	$\frac{3}{4}$	$\frac{1}{4}-100\varepsilon$

If ε is a small positive number, then by continuity $\tilde{P}$ still dominates $\tilde{Q}$ in the Rényi order. Nonetheless, we show below that $\tilde{P}^{\otimes n}$ does not Blackwell dominate $\tilde{Q}^{\otimes n}$ for any n and $\varepsilon > 0$.

To do this, let $\overline{p} := \frac{100^{n-1}}{100^{n-1}+1}$ be a threshold belief. We will show that a decision maker whose indirect utility function is $(p-\overline{p})^+$ strictly prefers $\tilde{Q}^{\otimes n}$ to $\tilde{P}^{\otimes n}$. Indeed, it suffices to focus on posterior beliefs $p > \overline{p}$; that is, the likelihood ratio should exceed 100^{n-1}. Under $\tilde{Q}^{\otimes n}$, this can only happen if every signal realization is y_0, or all but one signal is y_0 and the remaining one is y_1. Thus, in the range $p > \overline{p}$, the posterior belief has the following distribution under $\tilde{Q}^{\otimes n}$:

$$p = \begin{cases} \frac{100^n}{100^n+1} & \text{w.p. } \frac{1}{2}(100^n+1)\varepsilon^n \\ \frac{3\cdot 100^{n-1}}{3\cdot 100^{n-1}+1} & \text{w.p. } \frac{n}{8}(3\cdot 100^{n-1}+1)\varepsilon^{n-1} \end{cases}$$

Similarly, under $\tilde{P}^{\otimes n}$ the relevant posterior distribution is

$$p = \begin{cases} \frac{100^n}{100^n+1} & \text{w.p. } \frac{1}{2}(100^n+1)\varepsilon^n \\ \frac{7\cdot 100^{n-1}}{7\cdot 100^{n-1}+1} & \text{w.p. } \frac{n}{32}(7\cdot 100^{n-1}+1)\varepsilon^{n-1} \end{cases}$$

Recall that the indirect utility function is $(p-\overline{p})^+$. So $\tilde{Q}^{\otimes n}$ yields higher expected payoff than $\tilde{P}^{\otimes n}$ if and only if

$$\frac{n}{8}(3\cdot 100^{n-1}+1)\varepsilon^{n-1}\cdot\left(\frac{3\cdot 100^{n-1}}{3\cdot 100^{n-1}+1}-\overline{p}\right) > \frac{n}{32}(7\cdot 100^{n-1}+1)\varepsilon^{n-1}\cdot\left(\frac{7\cdot 100^{n-1}}{7\cdot 100^{n-1}+1}-\overline{p}\right).$$

That is,

$$4(3\cdot 100^{n-1}+1)\cdot\left(\frac{3\cdot 100^{n-1}}{3\cdot 100^{n-1}+1}-\frac{100^{n-1}}{100^{n-1}+1}\right) > (7\cdot 100^{n-1}+1)\cdot\left(\frac{7\cdot 100^{n-1}}{7\cdot 100^{n-1}+1}-\frac{100^{n-1}}{100^{n-1}+1}\right).$$

The LHS is computed to be $\frac{8\cdot 100^{n-1}}{100^{n-1}+1}$, while the RHS is $\frac{6\cdot 100^{n-1}}{100^{n-1}+1}$. Hence the above inequality holds, and it follows that $\tilde{P}^{\otimes n}$ does not Blackwell dominate $\tilde{Q}^{\otimes n}$.

J Generalization to Unbounded Experiments

In this section we present two generalizations of Theorem 1 to experiments that may have unbounded likelihood ratios. Note that the Rényi divergences for an unbounded experiment can still be defined by (3), (4) and (5), so long as we allow these divergences to take the value $+\infty$.

The first result shows that Theorem 1 hold without change so long as the dominated experiment Q is bounded.

Theorem 6. *For a generic pair of experiments P and Q where Q is bounded, the following are equivalent:*

(i). *P dominates Q in large samples.*

(ii). *P dominates Q in the Rényi order.*

To interpret the statement, "generic" means (as in the main text) that $\log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}$ has different essential maximum and minimum from $\log \frac{\mathrm{d}Q_1}{\mathrm{d}Q_0}$. In the current setting P may be unbounded, so that its log-likelihood ratio may have essential maximum $+\infty$ and/or minimum $-\infty$. In those cases the the genericity assumption is automatically satisfied.

We also reiterate that dominance in the Rényi order means the Rényi divergences of P and Q are ranked as $R_P^\theta(t) > R_Q^\theta(t)$ for all $t > 0$ and $\theta \in \{0, 1\}$. Since Q is by assumption bounded, $R_Q^\theta(t)$ is always finite. Thus the requirement in (ii) is that $R_P^\theta(t)$ is either a bigger finite number, or it is $+\infty$.

Our second result in this section deals with pairs of experiments where both P and Q may be unbounded, but they still have finite Rényi divergences. To state the result, we need to generalize the notion of genericity as follows: Say P and Q form a *generic* pair, if for both $\theta = 0$ and $\theta = 1$,

$$\liminf_{t\to\infty} |R_P^\theta(t) - R_Q^\theta(t)| > 0. \tag{29}$$

Note that when P and Q are bounded, $R_P^\theta(t) \to \max[X^\theta]$ and $R_Q^\theta(t) \to \max[Y^\theta]$ as $t \to \infty$. So in this special case the genericity assumption reduces to the one we introduced in the main text.

The following result shows that under one extra assumption, Theorem 1 once again extends.

Theorem 7. *Suppose P and Q are a generic pair of (possibly unbounded) experiments with finite Rényi divergences. Let $(X^\theta), (Y^\theta)$ be the corresponding log-likelihood ratios, and suppose further that their cumulant generating functions satisfy $\sup_{t\in\mathbb{R}} K''_{X^\theta}(t) < \infty$ and $\sup_{t\in\mathbb{R}} K''_{Y^\theta}(t) < \infty$.*[26] *Then the following are equivalent:*

[26] Since $K_{X^0}(t) = K_{X^1}(-1-t)$, it suffices to check the assumptions $\sup_{t\in\mathbb{R}} K''_{X^\theta}(t) < \infty$ and $\sup_{t\in\mathbb{R}} K''_{Y^\theta}(t) < \infty$ for one of the two states.

(i). *P dominates Q in large samples.*

(ii). *P dominates Q in the Rényi order.*

We note that if a random variable X is bounded between $-b$ and b, then its Rényi divergences are finite, and $K''_X(t) \leq b^2$ for every t.[27] Thus Theorem 7 is another strict generalization of Theorem 1 beyond bounded experiments.

More generally, the following is a sufficient condition for Theorem 7 to apply. Roughly speaking, we require the log-likelihood ratios X^θ, Y^θ to have *tails decaying faster than some Gaussian distribution.*

Lemma 7. *Let X be a random variable whose distribution admits a density $h(x)$ that is positive and twice continuously differentiable. Suppose there exists $\epsilon > 0$ and $M > 0$ such that the following holds:*

$$\frac{\partial^2 \log h(x)}{\partial x^2} \leq -\epsilon \quad \textit{for all} \quad |x| > M.$$

Then the cumulant generating function $K_X(t)$ is finite for every t, and $\sup_{t\in\mathbb{R}} K''_X(t) < \infty$.

Note that $\frac{\partial^2 \log h(x)}{\partial x^2} \leq -\epsilon$ implies the standard assumption that the density h is (strictly) log-concave. The requirement that the same ϵ works for all large x makes our assumption stronger, and in particular rules out densities such as $h_1(x) = c_1 \cdot e^{-\lambda_1|x|}$ or $h_2(x) = c_2 \cdot e^{-\lambda_2|x|^{1.99}}$.[28] Nonetheless, any Gaussian density h satisfies the assumption regardless of how big the variance is, and so does any other density that decays faster at infinity. Hence Theorem 7 is applicable to a broad class of unbounded experiments.

Below we prove Theorem 6, Theorem 7 and Lemma 7 in turn.

J.1 Proof of Theorem 6

That (i) implies (ii) follows from the same argument as in §5.1. To prove (ii) implies (i), the idea is to garble P into a bounded experiment $\tilde{P}$ that still has higher Rényi divergences than Q. By Theorem 1, $\tilde{P}^{\otimes n}$ Blackwell dominates $Q^{\otimes n}$ for all large n. But since P Blackwell dominates $\tilde{P}$, $P^{\otimes n}$ also Blackwell dominates $\tilde{P}^{\otimes n}$. Therefore, by transitivity, we would be able to conclude that $P^{\otimes n}$ Blackwell dominates $Q^{\otimes n}$ for all large n.

To construct such a $\tilde{P}$, we first note that by taking $t \to \infty$, $R^1_P(t) > R^1_Q(t)$ implies $\max[X^1] \geq \max[Y^1]$ where X^1 and Y^1 are the log-likelihood ratios. Similarly $\max[X^0] \geq$

[27] The latter follows by showing $K''_X(t)$ to be the variance of some random variable $\hat{X}$ that shares the same support as X. See Proposition 6 and its proof.

[28] It is easy to see that the random variable with density $h_1(x)$ does not have finite Rényi divergences everywhere. It can also be shown that the random variable with density $h_2(x)$ has a cumulant generating function with $K''_X(t) \to \infty$ as $t \to \infty$. Thus, it seems difficult to substantially weaken the condition in Lemma 7 while maintaining the same result.

$\max[Y^0]$. By the genericity assumption, both comparisons are in fact strict. We can thus find a pair of positive numbers $b_1 \in (\max[Y^1], \max[X^1])$ and $b_0 \in (\max[Y^0], \max[X^0]) = (-\min[Y^1], -\min[X^1])$. These numbers will be fixed throughout.

Now take any positive number $B \geq \max\{b_1, b_0\}$. We construct a garbling of P, denoted P_B, as follows: All signal realizations under P that induce a log-likelihood ratio $\log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}$ greater than B (if any) are garbled into a single signal $\overline{s}$, and similarly all realizations with log-likelihood ratio less than $-B$ are garbled into another signal $\underline{s}$. The remaining signal realizations under P (with log-likelihood ratio in $[-B, B]$) are unchanged under P_B. It is easy to see that not only is P_B a garbling of P, but more generally P_B is a garbling of $P_{B'}$ whenever $B' > B$. Thus, as B increases, the experiment P_B becomes more informative in the Blackwell sense.

Let $R^\theta_{P_B}(t)$ denote the Rényi divergences of P_B. Since the Rényi order extends the Blackwell order, we know that as B increases, $R^\theta_{P_B}(t)$ also increases for each θ and t, with an upper bound of $R^\theta_P(t)$. In fact, we can show that for fixed θ and t,

$$\lim_{B\to\infty} R^\theta_{P_B}(t) = R^\theta_P(t).$$

The proof is technical and deferred to later. Assuming this, we next show that for sufficiently large B, $R^\theta_{P_B}(t) > R^\theta_Q(t)$ holds for *all* $t \geq 1/2$ (thus for all $t > 0$, by (6)). This will prove P_B as the desired garbling $\tilde{P}$ that dominates Q in the Rényi order, which will complete the proof of the theorem.[29]

To this end, fix $\theta = 1$, and define for each B a set

$$T_B = \{t \geq 1/2 : R^1_{P_B}(t) \leq R^1_Q(t)\}.$$

By continuity of the Rényi divergences, T_B is a closed set. Moreover, as $t \to \infty$ we have $R^1_{P_B}(t) \to \max[X^1_B]$, where X^1_B is the log-likelihood ratio of state 1 to state 0, distributed under the experiment P_B and true state 1. By the assumption $B \geq b_1$ and the construction of P_B, we have that

$$\mathbb{P}[X^1_B \geq b_1] = \mathbb{P}[X^1 \geq b_1],$$

which is positive because $b_1 < \max[X^1]$. Thus $\max[X^1_B] \geq b_1$. It follows that

$$\lim_{t\to\infty} R^1_{P_B}(t) \geq b_1 > \max[Y^1] = \lim_{t\to\infty} R^1_Q(t).$$

Hence $R^1_{P_B}(t) > R^1_Q(t)$ for all large t and T_B is a bounded set.

We have shown that each T_B is compact set. Note also that because $R^1_{P_B}(t)$ increases in B, the set T_B shrinks as B increases. Therefore, by the finite intersection property, either there exists some t that belongs to every T_B, or T_B is the empty set for all large B.

[29] Note that $B \geq \max\{b_1, b_2\}$ ensures P_B and Q is a generic pair, so we can apply Theorem 1 to deduce $P_B^{\otimes n} \succeq Q^{\otimes n}$ for large n. Therefore $P^{\otimes n} \succeq P_B^{\otimes n} \succeq Q^{\otimes n}$.

The former is impossible because $R^1_{P_B}(t) \leq R^1_Q(t)$ for all B would imply $R^1_P(t) \leq R^1_Q(t)$ in the limit, contradicting the assumption in (ii).

We thus conclude that T_B must be empty for all large B. In other words, when B is large $R^1_{P_B}(t) > R^1_Q(t)$ holds for all $t \geq \frac{1}{2}$. A symmetric argument shows that $R^0_{P_B}(t) > R^0_Q(t)$ holds for all $t \geq \frac{1}{2}$, completing the proof.

It remains to show $\lim_{B\to\infty} R^\theta_{P_B}(t) = R^\theta_P(t)$. We again fix $\theta = 1$ for easier exposition. Consider the following three cases:

Case 1: $t > 1$**.** We recall that $R^1_{P_B}(t) = \frac{1}{t-1} \log \mathbb{E}[\mathrm{e}^{(t-1)X^1_B}]$. So we need to show

$$\lim_{B\to\infty} \mathbb{E}[\mathrm{e}^{(t-1)X^1_B}] = \mathbb{E}[\mathrm{e}^{(t-1)X^1}].$$

Since $R^1_{P_B}(t) \leq R^1_P(t)$ for each B, the LHS above is weakly smaller than the RHS. On the other hand, by construction X^1_B coincides with X^1 conditional on being in the interval $[-B, B]$. As the exponential function is always positive, we have

$$\begin{aligned}\mathbb{E}[\mathrm{e}^{(t-1)X^1_B}] &\geq \mathbb{P}\left[|X^1_B| \leq B\right] \cdot \mathbb{E}[\mathrm{e}^{(t-1)X^1_B} \;\mid\; |X^1_B| \leq B] \\ &= \mathbb{P}\left[|X^1| \leq B\right] \cdot \mathbb{E}[\mathrm{e}^{(t-1)X^1} \;\mid\; |X^1| \leq B].\end{aligned}$$

Taking the limit as $B \to \infty$, we obtain $\lim_{B\to\infty} \mathbb{E}[\mathrm{e}^{(t-1)X^1_B}] \geq \mathbb{E}[\mathrm{e}^{(t-1)X^1}]$, which proves they are equal.

Case 2: $t = 1$**.** Here we have $R^1_{P_B}(1) = \mathbb{E}[X^1_B]$. So we need to show

$$\lim_{B\to\infty} \mathbb{E}[X^1_B] = \mathbb{E}[X^1].$$

Once again we already know the LHS is weakly smaller, so it suffices to show the opposite inequality. By construction, X^1_B coincides with X^1 on the interval $[-B, B]$. Other than this part, there is probability $\mathbb{P}[X^1 > B]$ that signal $\overline{s}$ occurs under the experiment P_B; when this happens we also have $X^1_B > B$, which contributes a positive amount to $\mathbb{E}[X^1_B]$.

With remaining probability $\mathbb{P}[X^1 < -B]$, the signal $\underline{s}$ occurs, and the induced log-likelihood ratio X^1_B is at least $\log \mathbb{P}[X^1 < -B]$ (since this event occurs with probability at most one under state 0). Here the contribution to $\mathbb{E}[X^1_B]$ can be negative, but is no less than $\mathbb{P}[X^1 < -B] \cdot \log \mathbb{P}[X^1 < -B]$.

Summarizing, for each B we have

$$\mathbb{E}[X^1_B] \geq \mathbb{P}\left[|X^1| \leq B\right] \cdot \mathbb{E}[X^1 \;\mid\; |X^1| \leq B] \quad + \quad \mathbb{P}[X^1 < -B] \cdot \log \mathbb{P}[X^1 < -B].$$

Taking the limit as $B \to \infty$, the first summand on the RHS converges to $\mathbb{E}[X^1]$. In addition, the second summand vanishes because $\mathbb{P}[X^1 < -B] \to 0$ and $\lim_{x\to 0} x \log x = 0$. We thus obtain $\lim_{B\to\infty} \mathbb{E}[X^1_B] \geq \mathbb{E}[X^1]$ as desired.

Case 3: $t \in (0,1)$**.** In this case we will again show

$$\lim_{B\to\infty} \mathbb{E}[\mathrm{e}^{(t-1)X_B^1}] = \mathbb{E}[\mathrm{e}^{(t-1)X^1}].$$

Since $R^1_{P_B}(t) \le R^1_P(t)$, and $R^1_{P_B}(t) = \frac{1}{t-1}\log \mathbb{E}[\mathrm{e}^{(t-1)X_B^1}]$, the negative factor $\frac{1}{t-1}$ implies that the LHS above is now weakly *bigger* than the RHS.

To prove it is smaller, we proceed as in Case 2. With probability $\mathbb{P}[X^1 > B]$ the signal $\overline{s}$ occurs, and the induced log-likelihood ratio X_B^1 is *at least* $\log \mathbb{P}[X^1 > B]$. As $t-1$ is negative here, the contribution of this part to $\mathbb{E}[\mathrm{e}^{(t-1)X_B^1}]$ is *at most*

$$\mathbb{P}[X^1 > B] \cdot \mathbb{E}[\mathrm{e}^{(t-1)\log \mathbb{P}[X^1>B]}] = (\mathbb{P}[X^1 > B])^t.$$

Similarly the contribution of the signal $\underline{s}$ is at most $(\mathbb{P}[X^1 < -B])^t$. We thus have

$$\mathbb{E}[\mathrm{e}^{(t-1)X_B^1}] \le \mathbb{P}\left[|X^1| \le B\right] \cdot \mathbb{E}[\mathrm{e}^{(t-1)X^1} \;\mid\; |X^1| \le B] \;+\; (\mathbb{P}[X^1 > B])^t \;+\; (\mathbb{P}[X^1 < -B])^t.$$

As $B \to \infty$, both $(\mathbb{P}[X^1 > B])^t$ and $(\mathbb{P}[X^1 < -B])^t$ vanish since $t > 0$. We therefore conclude $\lim_{B\to\infty} \mathbb{E}[\mathrm{e}^{(t-1)X_B^1}] \le \mathbb{E}[\mathrm{e}^{(t-1)X^1}]$, completing the whole proof.

J.2 Proof of Theorem 7

We only need to prove (ii) implies (i). Here we will follow the arguments in §5.6 and make necessary modifications. Since Lemma 1 remains valid, it suffices to prove (22), i.e.,

$$\mathbb{P}\left[X_1^1 + \cdots + X_n^1 \le na\right] \le \mathbb{P}\left[Y_1^1 + \cdots + Y_n^1 \le na\right], \qquad \text{for all } a \ge 0.$$

The analysis of the four cases in §5.6 relies on Lemma 2 and Proposition 5. We will show later that Lemma 2 continues to hold even if P and Q are unbounded (but have finite Rényi divergences). On the other hand, Proposition 5 cannot hold as stated, but we do have the following modified version where b^2 is replaced by $\sup_{t\in\mathbb{R}} K_X''(t)$:

Proposition 6. *Let X and Y be random variables with finite cumulant generating functions $K_X(t)$ and $K_Y(t)$. Further let $X_1, \ldots, X_n$, $Y_1, \ldots, Y_n$ be i.i.d. copies of X and Y respectively. Suppose $a \ge \mathbb{E}[Y]$, and $\eta > 0$ satisfies $K_Y^*(a) - \eta > K_X^*(a+\eta)$. Then for all $n \ge 4(1+\eta)\eta^{-3} \cdot \sup_{t\in\mathbb{R}} K_X''(t)$, it holds that*

$$\mathbb{P}\left[X_1 + \cdots + X_n > na\right] \ge \mathbb{P}\left[Y_1 + \cdots + Y_n > na\right].$$

Using Lemma 2 and Proposition 6, we can replicate the results in Cases 1, 2 and 4 in §5.6. Specifically, let $M = \max\{\sup_{t\in\mathbb{R}} K_{X^1}''(t), \sup_{t\in\mathbb{R}} K_{Y^1}''(t)\}$, then for all $n \ge 4M(1+\eta)\eta^{-3}$ the inequality $\mathbb{P}\left[X_1^1 + \cdots + X_n^1 \le na\right] \le \mathbb{P}\left[Y_1^1 + \cdots + Y_n^1 \le na\right]$ holds for values of a outside of the interval $(\mathbb{E}[Y] + \eta, \mathbb{E}[X] - \eta)$ in Case 3.

Turning to $a \in (\mathbb{E}[Y]+\eta, \mathbb{E}[X]-\eta)$, we can still use the Chebyshev inequality to deduce

$$\mathbb{P}\left[X_1^1 + \cdots + X_n^1 \leq na\right] \leq \frac{\mathrm{Var}[X^1]}{n\eta^2} = \frac{K''_{X^1}(0)}{n\eta^2} \leq \frac{M}{n\eta^2}.$$

Similarly we also have

$$\mathbb{P}\left[Y_1^1 + \cdots + Y_n^1 \leq na\right] \geq 1 - \frac{\mathrm{Var}[Y^1]}{n\eta^2} \geq 1 - \frac{M}{n\eta^2}.$$

Thus $\mathbb{P}\left[X_1^1 + \cdots + X_n^1 \leq na\right] \leq \mathbb{P}\left[Y_1^1 + \cdots + Y_n^1 \leq na\right]$ holds for all $n \geq 2M\eta^{-2}$, and hence for all $n \geq 4M(1+\eta)\eta^{-3}$. This then implies that $P^{\otimes n}$ Blackwell dominates $Q^{\otimes n}$ for all $n \geq 4M(1+\eta)\eta^{-3}$.

Below we supply the proofs for Lemma 2 (for unbounded experiments) and Proposition 6.

Proof of Lemma 2 for unbounded experiments. We note that the second part $K^*_{Y^\theta}(a-\eta) < K^*_{X^\theta}(a) - \eta$ continues to hold. This is because, by the same argument as in the case of bounded experiments, $K^*_{Y^\theta}(a) < K^*_{X^\theta}(a)$ holds for all a in the *compact* interval $[0, \mathbb{E}[Y^\theta]]$. Thus by (uniform) continuity, we can "squeeze in" a small positive η without changing the inequality.

The first part of Lemma 2 also holds so long as $\max[Y^\theta]$ is finite, in which case the range of a under consideration is again compact. If instead $\max[Y^\theta] = \infty$, we use a new argument that takes advantage of the genericity assumption. Note that by assumption, $R^\theta_P(t) - R^\theta_Q(t)$ is positive for each θ and t. Given this, the genericity assumption (29) further implies this difference is bounded away from zero as $t \to \infty$. That is, there exists small $\epsilon > 0$ and large $T > 1$ such that

$$R^\theta_P(t) - R^\theta_Q(t) > \epsilon \quad \text{for all} \quad \theta \in \{0,1\},\ t > T.$$

Since $K^\theta_X(t) = tR^\theta_P(t+1)$, we deduce

$$K^\theta_X(t) - K^\theta_Y(t) > \epsilon t > \frac{\epsilon}{2}(t+1) \quad \text{for all} \quad \theta \in \{0,1\},\ t > T. \tag{30}$$

We can now prove the first part of Lemma 2. Define $\delta > 0$ by $K'_{X^\theta}(T) = \mathbb{E}[X^\theta] + \delta$. The original proof of Lemma 2 yields that for all sufficiently small $\eta > 0$,

$$K^*_{Y^\theta}(a) - \eta > K^*_{X^\theta}(a+\eta) \quad \text{holds for} \quad \mathbb{E}[X^\theta] - \eta \leq a \leq \mathbb{E}[X^\theta] + \delta.$$

Note that $\mathbb{E}[X^\theta] + \delta$ is finite, so the range of a considered above is compact, enabling us to use the original argument. We claim that by choosing $\eta < \epsilon/2$, where ϵ is defined earlier, the same inequality holds even if a is bigger than $\mathbb{E}[X^\theta] + \delta$. For this define $\hat{t}$ by

$K'_{X^\theta}(\hat{t}) = a + \eta$, then $\hat{t} > T$ by the convexity of K_X. Therefore, by (30),

$$\begin{aligned} K^*_{X^\theta}(a+\eta) &= \hat{t}(a+\eta) - K_{X^\theta}(\hat{t}) \\ &< \hat{t}(a+\eta) - K_{Y^\theta}(\hat{t}) - \frac{\epsilon}{2}(\hat{t}+1) \\ &< \hat{t}(a+\eta) - K_{Y^\theta}(\hat{t}) - \eta(\hat{t}+1) \\ &= \hat{t}a - K_{Y^\theta}(\hat{t}) - \eta \\ &\le K^*_{Y^\theta}(a) - \eta. \end{aligned}$$

This completes the proof of Lemma 2 for unbounded experiments. □

Proof of Proposition 6. Following the original proof of Proposition 5, we just need to show a modified version of Lemma 3 (with $\sup_{t\in\mathbb{R}} K''_X(t)$ replacing b^2):

$$\mathbb{P}\left[X_1 + \cdots + X_n > na\right] \ge \mathrm{e}^{-n\cdot K^*_X(a+\eta)} \left(1 - \frac{4\cdot \sup_{t\in\mathbb{R}} K''_X(t)}{n\eta^2}\right).$$

This follows the same proof as in §A, except that in applying the Chebyshev inequality, we now use

$$\mathrm{Var}[\hat{S}_n] = n\,\mathrm{Var}[\hat{X}] = n\cdot K''_X(t) \le n\cdot \sup_{\hat{t}\in\mathbb{R}} K''_X(\hat{t})$$

instead of $\mathrm{Var}[\hat{S}_n] \le nb^2$. The key equality $\mathrm{Var}[\hat{X}] = K''_X(t)$ holds because

$$\mathrm{Var}[\hat{X}] = \mathbb{E}[\hat{X}^2] - \mathbb{E}[\hat{X}]^2 = \frac{\mathbb{E}[X^2\mathrm{e}^{tX}]}{\mathbb{E}[\mathrm{e}^{tX}]} - \left(\frac{\mathbb{E}[X\mathrm{e}^{tX}]}{\mathbb{E}[\mathrm{e}^{tX}]}\right)^2 = K''_X(t).$$

Hence the result. □

J.3 Proof of Lemma 7

We first prove K_X is everywhere finite, i.e., $\log\mathbb{E}[\mathrm{e}^{tX}]$ is finite for every t. Using the density $h(x)$, we can write

$$\mathbb{E}[\mathrm{e}^{tX}] = \int_{-\infty}^{\infty} h(x)\mathrm{e}^{tx}\,\mathrm{d}x = \int_{-\infty}^{\infty} \mathrm{e}^{tx+l(x)}\,\mathrm{d}x,$$

where we define $\ell(x) = \log h(x)$. Since by assumption $\ell''(x) \le -\epsilon$ for $|x| > M$, it is easy to show $\ell(x) \le -\frac{\epsilon}{4}x^2$ as $|x| \to \infty$. Hence the above integral is finite.

To prove K''_X is bounded, we begin with the formula

$$K''_X(t) = \frac{\mathbb{E}[X^2\mathrm{e}^{tX}]\cdot\mathbb{E}[\mathrm{e}^{tX}] - \mathbb{E}[X\mathrm{e}^{tX}]^2}{\mathbb{E}[\mathrm{e}^{tX}]^2}.$$

Let X_1, X_2 be i.i.d. copies of X. Then the denominator above is $\mathbb{E}[\mathrm{e}^{tX_1}] \cdot \mathbb{E}[\mathrm{e}^{tX_2}] = \mathbb{E}[\mathrm{e}^{t(X_1+X_2)}]$. The numerator can be rewritten as

$$\begin{aligned}
&\mathbb{E}[X_1^2 \mathrm{e}^{tX_1}] \cdot \mathbb{E}[\mathrm{e}^{tX_2}] - \mathbb{E}[X_1 \mathrm{e}^{tX_1}] \cdot \mathbb{E}[X_2 \mathrm{e}^{tX_2}] \\
=&\mathbb{E}[(X_1^2 - X_1X_2) \cdot \mathrm{e}^{t(X_1+X_2)}] \\
=&\mathbb{E}[\frac{X_1^2 - X_1X_2 + X_2^2 - X_1X_2}{2} \cdot \mathrm{e}^{t(X_1+X_2)}] \\
=&\mathbb{E}[\frac{(X_1 - X_2)^2}{2} \cdot \mathrm{e}^{t(X_1+X_2)}],
\end{aligned}$$

where the penultimate step uses the symmetry between X_1 and X_2. Define

$$D(s) = \mathbb{E}[(X_1 - X_2)^2 \mid X_1 + X_2 = s].$$

Then we have shown that

$$K_X''(t) = \frac{\frac{1}{2}\mathbb{E}[D(X_1 + X_2) \cdot \mathrm{e}^{t(X_1+X_2)}]}{\mathbb{E}[\mathrm{e}^{t(X_1+X_2)}]}.$$

Thus, in order to show K_X'' is bounded, it suffices to show $D(s)$ is bounded as s varies.

Recall that by assumption $\ell''(x) \le -\epsilon$ for $|x| > M$. We will show (with proof deferred to later) there exists $S > 2M$, such that

$$\ell'(x) - \ell'(s - x) \le -\frac{\epsilon}{2}(2x - s) \quad \text{for all} \quad s > S,\ x > \frac{s}{2}. \tag{31}$$

Note that (31) in particular implies $\ell'(x) - \ell'(s - x) \le -1$ for $x > \frac{s}{2} + C$, with $C = \epsilon^{-1}$. Given this, we can show $D(s)$ is bounded.

Without loss consider $s \ge 0$. We use the density $h(x)$ to write

$$D(s) = \frac{\int_{-\infty}^{\infty} h(x)h(s - x)(2x - s)^2 \,\mathrm{d}x}{\int_{-\infty}^{\infty} h(x)h(s - x) \,\mathrm{d}x} = \frac{\int_{s/2}^{\infty} h(x)h(s - x)(2x - s)^2 \,\mathrm{d}x}{\int_{s/2}^{\infty} h(x)h(s - x) \,\mathrm{d}x} \tag{32}$$

Since $D(s)$ is continuous, it suffices to prove it is bounded when $s > S$, where S is given earlier. We now break the integral in (32) into two parts, with cutoff $s/2 + 2C$:

$$D(s) = \frac{\int_{s/2}^{s/2+2C} h(x)h(s - x)(2x - s)^2 \,\mathrm{d}x}{\int_{s/2}^{\infty} h(x)h(s - x) \,\mathrm{d}x} \quad + \quad \frac{\int_{s/2+2C}^{\infty} h(x)h(s - x)(2x - s)^2 \,\mathrm{d}x}{\int_{s/2}^{\infty} h(x)h(s - x) \,\mathrm{d}x}.$$

The first term is bounded by $16C^2$, which is the maximum value of $(2x - s)^2$ for $x \in [s/2, s/2 + 2C]$. To bound the second term, we rewrite it as

$$\int_{s/2+2C}^{\infty} \frac{\mathrm{e}^{l(x)+l(s-x)}}{\int_{s/2}^{\infty} \mathrm{e}^{l(y)+l(s-y)} \,\mathrm{d}y} \cdot (2x - s)^2 \,\mathrm{d}x. \tag{33}$$

As $l'(y) - l'(s-y) \le -1$ for $y \ge s/2 + C$, we have $l(y) + l(s-y) \ge x - y + l(x) + l(s-x)$ for all $x \ge y \ge s/2 + C$. Thus

$$\int_{s/2}^{\infty} \mathrm{e}^{l(y)+l(s-y)}\,\mathrm{d}y \ge \int_{s/2+C}^{x} \mathrm{e}^{l(y)+l(s-y)}\,\mathrm{d}y \ge \int_{s/2+C}^{x} \mathrm{e}^{x-y+l(x)+l(s-x)}\,\mathrm{d}y = (\mathrm{e}^{x-s/2-C}-1)\mathrm{e}^{l(x)+l(s-x)}.$$

Plugging back into (33), the second term contributing to $D(s)$ is bounded above by

$$\int_{s/2+2C}^{\infty} \frac{1}{\mathrm{e}^{x-s/2-C}-1} \cdot (2x-s)^2\,\mathrm{d}x = \int_{C}^{\infty} \frac{1}{\mathrm{e}^{u}-1} \cdot (2u+2C)^2\,\mathrm{d}u,$$

where we used change of variable from x to $u = x - s/2 - C$. Since the RHS is a finite constant independent of s, we conclude that $D(s)$ is bounded even as $s \to \infty$.

It remains to prove (31). We write the difference on the LHS as $\int_{s-x}^{x} \ell''(u)\,\mathrm{d}u$. If $s - x > M$, the result follows from the fact that $\ell''(u) \le -\epsilon \le -\frac{\epsilon}{2}$ for every u in the range of integration. Suppose instead that $s - x \le M$, thus $x \ge s - M$. In this case because $\ell''(u)$ can only be positive for $u \in [-M, M]$, we have

$$\begin{aligned}
\int_{s-x}^{x} \ell''(u)\,\mathrm{d}u &\le -\epsilon(2x-s-2M) + \int_{-M}^{M} |\ell''(u)|\,\mathrm{d}u \\
&= -\epsilon(x-s/2) - \epsilon(x-s/2-2M) + \int_{-M}^{M} |\ell''(u)|\,\mathrm{d}u \\
&\le -\epsilon(x-s/2) - \epsilon(s/2-3M) + \int_{-M}^{M} |\ell''(u)|\,\mathrm{d}u \\
&\le -\epsilon(x-s/2).
\end{aligned}$$

The penultimate inequality uses $x \ge s - M$, whereas the last inequality holds when s is sufficiently large (since $\int_{-M}^{M} |\ell''(u)|\,\mathrm{d}u$ is finite by the assumption that h is positive and twice continuously differentiable). This completes the proof.

K Necessary Condition for Large Sample Dominance with Many States

In this section we show that the Rényi order can be generalized to more than two states to yield a general necessary condition for large sample dominance. Consider $k+1$ states $\theta \in \{0, 1, \ldots, k\}$ and two experiments $P = (\Omega, (P_\theta))$, $Q = (\Xi, (Q_\theta))$ revealing information about these states. Conditioning on $\theta = 0$, we consider the moment generating function of the log-likelihood ratio vector $(\frac{\mathrm{d}P_0}{\mathrm{d}P_1}, \ldots, \frac{\mathrm{d}P_0}{\mathrm{d}P_k})$, given by

$$M_{X^0}(t) = \int_{\Omega} \mathrm{e}^{\sum_{j=1}^{k} t_j \log \frac{\mathrm{d}P_0(\omega)}{\mathrm{d}P_j(\omega)}}\,\mathrm{d}P_0(\omega) \tag{34}$$

with $t = (t_1, \ldots, t_k) \in \mathbb{R}^k$. Similarly define $M_{Y^0}(t)$ for the experiment Q.

By the same argument as in §5.1 (see the derivation of (8)), $M_{X^0}(t)$ *would be* the ex-ante expected payoff from observing P, in a decision problem with uniform prior and indirect utility function

$$v(p) = (k+1)p_0^{1+t_1+\cdots+t_k} \cdot p_1^{-t_1} \cdots p_k^{-t_k},$$

where $p = (p_0, p_1, \ldots, p_k)$ represents the belief about the $k+1$ states. If the function $v(p)$ were convex in p, then it is indeed an indirect utility function. Blackwell dominance of P over Q then requires $M_{X^0}(t) \geq M_{Y^0}(t)$. Since the moment generating function is raised to the n-th power when n i.i.d. samples are drawn, we would be able to conclude that $M_{X^0}(t) \geq M_{Y^0}(t)$ also has to hold if P dominates Q in large samples. If instead $v(p)$ were concave, then $-v(p)$ is an indirect utility function, leading to the reverse ranking between the moment generating functions.

We can characterize those parameters $t = (t_1, \ldots, t_k)$ that make the function $v(p)$ globally convex/concave. To make the result easy to state, we make the variables symmetric and consider a function of the form

$$v(p) = (k+1)p_0^{\alpha_0} \cdot p_1^{\alpha_1} \cdots p_k^{\alpha_k}$$

with $\alpha_0 + \alpha_1 + \cdots + \alpha_k = 1$.

Lemma 8. *Consider the function $v(p)$ defined above, over the domain $p \in int(\Delta^k)$. Suppose $\alpha_0 + \alpha_1 + \cdots + \alpha_k = 1$ and $\alpha_0 > 0$. Then $v(p)$ is convex in p if and only if $\alpha_1, \ldots, \alpha_k$ are all non-positive. Conversely, $v(p)$ is concave in p if and only if $\alpha_1, \ldots, \alpha_k$ are non-negative. Moreover, the convexity/concavity is strict when $\alpha_1, \ldots, \alpha_k$ are strictly negative/positive.*

The proof of this lemma is deferred to the end of the section. Note that unlike the case of two states, there are situations where $v(p)$ is neither convex nor concave.

By rewriting $\alpha_j = -t_j$ for $1 \leq j \leq k$, we obtain the following necessary condition for Blackwell dominance in large samples. Say the experiments P and Q form a generic pair, if for every pair of states $i \neq j$, the maximum and minimum of $\log \frac{\mathrm{d}P_i}{\mathrm{d}P_j}$ differ from those of $\log \frac{\mathrm{d}Q_i}{\mathrm{d}Q_j}$.

Proposition 7. *Suppose P and Q are a generic pair of bounded experiments for $k+1$ states. If P Blackwell dominates Q in large samples, then the following conditions hold:*[30]

(i). *For all $t \in \mathbb{R}^k_+ \backslash \{\mathbf{0}\}$, $M_{X^0}(t) > M_{Y^0}(t)$ and symmetrically $M_{X^i}(t) > M_{Y^i}(t)$ if we define the moment generating functions for true state i analogously to (34);*

[30] We exclude $t = \{\mathbf{0}\}$ from the conditions because $M_X(\mathbf{0}) = M_Y(\mathbf{0}) = 1$ always holds.

(ii). *For all* $t \in \mathbb{R}^k_- \backslash \{\mathbf{0}\}$ *such that* $\sum_{j=1}^k t_j > -1$, $M_{X^0}(t) < M_{Y^0}(t)$ *and symmetrically* $M_{X^i}(t) < M_{Y^i}(t)$ *for* $1 \leq i \leq k$*;*

(iii). *For every pair of states* $i \neq j$*, the Kullback-Leibler divergence between* P_i *and* P_j *exceeds the divergence between* Q_i *and* Q_j*:*

$$\int_\Omega \log \frac{\mathrm{d}P_i(\omega)}{\mathrm{d}P_j(\omega)} \, \mathrm{d}P_i(\omega) > \int_\Xi \log \frac{\mathrm{d}Q_i(\xi)}{\mathrm{d}Q_j(\xi)} \, \mathrm{d}Q_i(\xi).$$

To understand Proposition 7, note from (34) that when t_j are *all* positive, a bigger value of $M_{X^0}(t)$ indicates higher likelihood ratios $\frac{\mathrm{d}P_0}{\mathrm{d}P_j}$ between state 0 and *every* other state j, when state 0 is the true state. It is intuitive that in this case $M_{X^0}(t) > M_{Y^0}(t)$ corresponds to P being (on average) a more informative experiment than Q.[31] This is the content of part (i), which generalizes the comparison of Rényi divergences $R^\theta_P(t) > R^\theta_Q(t)$ in the two state case, for $t > 1$.

Conversely, part (ii) says that when t_j are all negative (subject to the extra condition $\sum_j t_j > -1$), informativeness is captured by the reverse ranking $M_{X^0}(t) < M_{Y^0}(t)$. In this case, the smaller value of $M_{X^0}(t)$ actually indicates higher likelihood ratios $\frac{\mathrm{d}P_0}{\mathrm{d}P_j}$ under true state 0. This part generalizes the comparison $R^\theta_P(t) > R^\theta_Q(t)$ for $t \in (0, 1)$.

Finally, part (iii) directly imposes the Rényi comparison $R^\theta_P(1) > R^\theta_Q(1)$ when it is applied to every pair of states.

We conjecture that the set of necessary conditions identified in Proposition 7 are also sufficient for large sample Blackwell dominance; see §6 for discussion of the difficulties.

Below we supply the proof of Lemma 8:

Proof of Lemma 8. The Hessian matrix of $v(\cdot)$ at p is computed as

$$Hess_v(p) = v(p) \times \begin{pmatrix} \frac{\alpha_0(\alpha_0-1)}{p_0^2} & \frac{\alpha_0\alpha_1}{p_0p_1} & \dots \\ \frac{\alpha_0\alpha_1}{p_0p_1} & \frac{\alpha_1(\alpha_1-1)}{p_1^2} & \dots \\ \dots & \dots & \dots \end{pmatrix}.$$

For any direction $(x_0, x_1, \dots, x_k)$, the directional second derivative of $v(\cdot)$ at p is thus

$$(x_0, x_1, \dots) \cdot \begin{pmatrix} \frac{\alpha_0(\alpha_0-1)}{p_0^2} & \frac{\alpha_0\alpha_1}{p_0p_1} & \dots \\ \frac{\alpha_0\alpha_1}{p_0p_1} & \frac{\alpha_1(\alpha_1-1)}{p_1^2} & \dots \\ \dots & \dots & \dots \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ x_1 \\ \dots \end{pmatrix} = \left(\sum_{i=0}^k \frac{\alpha_i x_i}{p_i} \right)^2 - \sum_{i=0}^k \frac{\alpha_i x_i^2}{p_i^2}, \quad (35)$$

[31] To prove the strict inequality $M_{X^0}(t) > M_{Y^0}(t)$, suppose that $t_1, \dots, t_l$ are positive whereas $t_{l+1}, \dots, t_k$ are zero, for some $1 \leq l \leq k$. Let $\tilde{P} = (\Omega, (P_0, \dots, P_l))$ be the restriction of the experiment P to the first $l+1$ states; similarly define $\tilde{Q}$. Then $P^{\otimes n} \succeq Q^{\otimes n}$ implies $\tilde{P}^{\otimes n} \succeq \tilde{Q}^{\otimes n}$, which must in fact be a strict comparison by the genericity assumption. Therefore, as the indirect utility function $\tilde{v}(p_0, \dots, p_l) = (k+1)p_0^{1+t_1+\dots+t_l} \cdot p_1^{-t_1} \cdots p_k^{-t_l}$ is *strictly* convex on the smaller belief space Δ^l (Lemma 8), the ex-ante expected payoff $M_{X^0}(t)$ must be strictly higher than $M_{Y^0}(t)$.

where for simplicity we have ignored the positive factor $v(p)$ as it does not affect the sign.

We first use this to show that if $\alpha_1 > 0$ (or any $\alpha_j > 0$), then the function $v(p)$ is *not* convex for $p \in int(\Delta^k)$. Indeed, consider the direction $(1, -1, 0, 0, \ldots, 0)$, which maintains $p \in int(\Delta^k)$. The directional second derivative can be computed as

$$\frac{\alpha_0(\alpha_0 - 1)}{p_0^2} - \frac{2\alpha_0\alpha_1}{p_0 p_1} + \frac{\alpha_1(\alpha_1 - 1)}{p_1^2}.$$

Suppose $p_0 = \alpha_0 x$, $p_1 = \alpha_1 x$ for some small positive number x, and $p_2, p_3, \ldots$ are arbitrary. Then the above second derivative simplifies to $-\frac{(\alpha_0+\alpha_1)}{\alpha_0\alpha_1 x^2} < 0$. Thus $v(p)$ is not convex along this direction.

Suppose instead $\alpha_1, \ldots, \alpha_k \leq 0$, we will show $v(p)$ is convex. For this it suffices to show the RHS of (35) is non-negative. Indeed, by the Cauchy-Schwartz inequality,

$$\left(\left(\sum_{i=0}^{k} \frac{\alpha_i x_i}{p_i}\right)^2 + \frac{-\alpha_1 x_1^2}{p_1^2} + \cdots + \frac{-\alpha_k x_k^2}{p_k^2}\right) \cdot (1 + (-\alpha_1) + \cdots + (-\alpha_k))$$
$$\geq \left(\sum_{i=0}^{k} \frac{\alpha_i x_i}{p_i} + \frac{-\alpha_1 x_1}{p_1} + \cdots + \frac{-\alpha_k x_k}{p_k}\right)^2 = \left(\frac{\alpha_0 x_0}{p_0}\right)^2.$$

Using $\alpha_0 + \alpha_1 + \cdots + \alpha_k = 1$ to simplify, this exactly implies $\left(\sum_{i=0}^k \frac{\alpha_i x_i}{p_i}\right)^2 \geq \sum_{i=0}^k \frac{\alpha_i x_i^2}{p_i^2}$ as desired. In fact, $v(p)$ is convex for all $p \gg 0$, including $p \in int(\Delta^k)$.

Moreover, if $\alpha_1, \ldots, \alpha_k$ are *strictly* negative, then the equality condition of the Cauchy-Schwartz inequality above requires $\sum_{i=0}^k \frac{\alpha_i x_i}{p_i} = \frac{x_1}{p_1} = \cdots = \frac{x_k}{p_k}$, which in turn implies that $x_0, x_1, \ldots, x_k$ have the same sign (under the assumption $\alpha_0 > 0 > \alpha_1, \ldots, \alpha_k$). Thus, for any direction $(x_0, x_1, \ldots, x_k)$ with $x_0 + x_1 + \cdots + x_k = 0$, the directional second derivative of v is strictly positive. So v is strictly convex for $p \in int(\Delta^k)$.

Next, we will show that if $\alpha_1 < 0$ (or any $\alpha_j < 0$), then the function $v(p)$ is *not* concave for $p \in int(\Delta^k)$. For this we again consider the second derivative along the direction $(1, -1, 0, 0, \ldots, 0)$, which is $\frac{\alpha_0(\alpha_0-1)}{p_0^2} - \frac{2\alpha_0\alpha_1}{p_0 p_1} + \frac{\alpha_1(\alpha_1-1)}{p_1^2}$. As $\alpha_1 < 0$, we have $\alpha_1(\alpha_1 - 1) > 0$. Thus for p_0 close to 1 and p_1 close to 0, the above second derivative is positive and $v(p)$ is not concave along this direction.

Finally, we show that if $\alpha_1, \ldots, \alpha_k \geq 0$, then the function $v(p)$ is concave. By the Cauchy-Schwartz inequality,

$$\left(\sum_{i=0}^{k} \frac{\alpha_i x_i^2}{p_i^2}\right) \cdot \left(\sum_{i=0}^{k} \alpha_i\right) \geq \left(\sum_{i=0}^{k} \frac{\alpha_i x_i}{p_i}\right)^2.$$

Since $\sum_{i=0}^k \alpha_i = 1$, this implies the RHS of (35) is non-positive. Hence v has non-positive directional second derivatives and must be globally concave.

Moreover, if $\alpha_1, \ldots, \alpha_k$ are strictly positive, then the equality condition of the Cauchy-Schwartz inequality requires $\frac{x_0}{p_0} = \frac{x_1}{p_1} = \cdots = \frac{x_k}{p_k}$, which in turn requires $x_0, x_1, \ldots, x_k$ to have the same sign. By the same argument as above, we conclude that in this case v is strictly concave for $p \in int(\Delta^k)$. □

L Proof of a Conjecture Regarding Majorization

Jensen (2019) studies the majorization order on finitely supported distributions. Given two such distributions μ and ν, μ is said to *majorize* ν if for every $n \geq 1$ it holds that the sum of the largest n probabilities in μ is greater than or equal to the sum of the n largest probabilities in ν. The Rényi entropy of a distribution μ defined on a finite set S is given by

$$H_\mu(\alpha) = \frac{1}{1-\alpha} \log \left(\sum_{s \in S} \mu(s)^\alpha \right),$$

for $\alpha \in [0, \infty) \setminus \{1\}$. As with our definition of Rényi divergences, this definition is extended to $\alpha = 1$ by continuity to equal the Shannon entropy, and extended to $\alpha = \infty$ to equal $-\log \max_s \mu(s)$. Hence H_μ is defined on $[0, \infty]$.

Note that $H_\mu(0)$ is the size of the support of μ. In his Proposition 3.7, Jensen shows that if $H_\mu(\alpha) < H_\nu(\alpha)$ for all $\alpha \in [0, \infty]$ then the n-fold product $\mu^{\times n}$ majorizes $\nu^{\times n}$.

Commenting on his Proposition 3.7, Jensen writes "The author cautiously conjectures that ... the requirement of a sharp inequality at 0 could be replaced by a similar condition regarding the α-Rényi entropies for negative α."

To understand this statement in terms of the nomenclature and notation of our paper, we identify each distribution μ whose support is a finite set S with the experiment $P^\mu = (S, P_1, P_0)$, where $P_1 = \mu$ and P_0 is the uniform distribution on S. There is a simple connection between the Rényi entropy of μ and the Rényi divergence of P^μ. For $\alpha \geq 0$,

$$H_\mu(\alpha) = \log |S| - R_P^1(\alpha). \tag{36}$$

As Jensen suggests, $H_\mu(\alpha)$ for negative α is also important, as it relates to R_P^0. For $\alpha \leq 0$,

$$H_\mu(\alpha) = \log |S| - \frac{\alpha}{1-\alpha} R_P^0(1-\alpha), \tag{37}$$

which extends to $\alpha = -\infty$ to equal $-\log \min_s \mu(s)$. Moreover, note that

$$H'_\mu(0) = -R_P^0(1) = \log |S| + \frac{1}{|S|} \sum_{s \in S} \log \mu(s). \tag{38}$$

As shown by Torgersen (1985, p. 264), when μ and ν have the same support size, then majorization of ν by μ is equivalent to Blackwell dominance of P^μ over P^ν. Thus Jensen's Proposition 3.7, which assumes that the support sizes are different, has no implications for

Blackwell dominance. However, our result on Blackwell dominance does have implications for majorization. In particular, the following proposition follows immediately from the application of Theorem 1 to experiments of the form P^μ.

Proposition 8. *Let μ, ν be finitely supported distributions with the same support size (i.e., $H_\mu(0) = H_\nu(0)$), and such that $H_\mu(\infty) \neq H_\nu(\infty)$ and $H_\mu(-\infty) \neq H_\nu(-\infty)$. Then the following are equivalent:*

(i). *$H_\mu(\alpha) < H_\nu(\alpha)$ for all $\alpha \in (0, \infty]$, $H_\mu(\alpha) > H_\nu(\alpha)$ for all $\alpha \in [-\infty, 0)$ and $H'_\mu(0) < H'_\nu(0)$.*[32]

(ii). *There exists an n_0 such that $\mu^{\times n}$ majorizes $\nu^{\times n}$ for every $n \geq n_0$.*

Proof. For notational ease, let P denote P^μ and Q denote P^ν. The assumption $H_\mu(\alpha) < H_\nu(\alpha)$ for all $\alpha > 0$ is equivalent, via (36), to $R^1_P(t) > R^1_Q(t)$ for all $t > 0$, and to $R^0_P(t) > R^0_Q(t)$ for all $t \in (0, 1)$, using $R^0_P(t) = \frac{t}{1-t} R^1_P(1-t)$ for $0 < t < 1$.

On the other hand, $H_\mu(\alpha) > H_\nu(\alpha)$ for all $\alpha < 0$ and $H'_\mu(0) < H'_\nu(0)$ is equivalent, via (37) and (38), to $R^0_P(t) > R^0_Q(t)$ for all $t \geq 1$. So (i) is equivalent to P dominating Q in the Rényi order.

Finally, the assumptions that $H_\mu(\infty) \neq H_\nu(\infty)$ and $H_\mu(-\infty) \neq H_\nu(-\infty)$ translate into $\max_s \mu(s) \neq \max_s \nu(s)$ and $\min_s \mu(s) \neq \min_s \nu(s)$, which are in turn equivalent to requiring that P and Q be a generic pair. Therefore, by Theorem 1, (i) is equivalent to $P^{\otimes n}$ Blackwell dominates $Q^{\otimes n}$ for every large n. It follows from Torgersen (1985) that (i) is equivalent to (ii). □

[32] This last condition is necessary for majorization, but it was not recognized in the original conjecture of Jensen (2019).