Econometrica, Vol. 89, No. 1 (January, 2021), 475–506

FROM BLACKWELL DOMINANCE IN LARGE SAMPLES TO RÉNYI DIVERGENCES AND BACK AGAIN

XIAOSHENG MU
Department of Economics, Princeton University

LUCIANO POMATTO
Division of the Humanities and Social Sciences, Caltech

PHILIPP STRACK
Department of Economics, Yale University

OMER TAMUZ
Division of the Humanities and Social Sciences, Caltech

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.

KEYWORDS: Comparison of experiments, stochastic dominance, 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.

As is well known, requiring two experiments to be ranked in the Blackwell order is a demanding condition. Consider the problem of testing $\theta \in \{0, 1\}$ a binary hypothesis, 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

Xiaosheng Mu: xmu@princeton.edu
Luciano Pomatto: luciano@caltech.edu
Philipp Strack: philipp.strack@yale.edu
Omer Tamuz: tamuz@caltech.edu

We are grateful to the coeditor 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. Xiaosheng Mu acknowledges the hospitality of Columbia University and the Cowles Foundation at Yale University, which hosted him during parts of this research. Philipp Strack was supported by a Sloan research fellowship. Omer Tamuz was supported by a grant from the Simons Foundation (#419427), a Sloan research fellowship, and a BSF award (#2018397).

 https://doi.org/10.3982/ECTA17548

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 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

[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.

[2] See, for example, Sawa (1978), White (1982), Critchley, Marriott, and Salmon (1996), Kitamura and Stutzer (1997), Hong and White (2005), Ullah (2002). See Kitamura, Otsu, and Evdokimov (2013) for a recent application of α-divergences, which are a reformulation of Rényi divergences.

principle of noninteraction 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 dataprocessing 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 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 Section 6.

The paper is organized as follows. In Section 2, we provide our main definitions. Section 3 contains the characterization of Blackwell dominance in large samples, with proof deferred to Section 5. In Section 4, we characterize additive divergences. Finally, we further discuss our results and their relation to the literature in Section 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* $\mathsf{P} = (\Omega \mathsf{P}_0 \mathsf{P}_1)$ consists of a sample space Ω, which we assume to be a Polish space, and a pair of Borel probability measures (P_0, P_1) defined over Ω, with the interpretation that $\mathsf{P}_\theta(\mathsf{A})$ is the probability of observing $\mathsf{A} \subseteq \Omega$ in state $\theta \in \{0, 1\}$. This framework is commonly encountered in simple hypothesis tests as well as in information economics. In Section 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 $\mathsf{P} = (\Omega \mathsf{P}_0 \mathsf{P}_1)$ and $\mathsf{Q} = (\Xi \mathsf{Q}_0 \mathsf{Q}_1)$, we can form the *product experiment* $\mathsf{P} \otimes \mathsf{Q}$ given by

$$\mathsf{P} \otimes \mathsf{Q} = (\Omega \times \Xi \mathsf{P}_0 \times \mathsf{Q}_0 \mathsf{P}_1 \times \mathsf{Q}_1)$$

where $\mathsf{P}_\theta \times \mathsf{Q}_\theta$, given, $\theta \in \{0, 1\}$ denotes the product of the two measures. Under the experiment $\mathsf{P} \otimes \mathsf{Q}$ 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 $\mathsf{P} \otimes \mathsf{Q}$ 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 overthat 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 $\omega \in \Omega$:

$$p(\omega) = \frac{dP^1(\omega)}{\int_0^1 dP} \quad [0,1]$$

$$(\omega) \qquad dP(\omega)$$

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 dP_1/dP_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 $P \quad Q$, if

$$\int_0^1 v(p)d\pi(p) \geq \int_0^1 v(p)d\tau(p) \qquad (1)$$

for every convex function $v : (0,1) \quad \mathbb{R}\rightarrow$. Equivalently, $P \quad Q$ if π is a mean-preserving spread of τ. We write $P \quad Q$ if $P \quad Q$ and $Q \quad P$. So, $P \quad Q$ if and only if (1) holds with a strict inequality whenever v is strictly convex, that is, π is a mean-preserving spread of τ and $\pi \quad = \tau$.

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 $A \times \{01\}$ 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, $P \quad Q$ 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 can be equivalently defined by "garbling" operations: Intuitively, P 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

$$P^{\otimes n} \; Q^{\otimes n} \qquad \text{for every } n \geq n_0 \tag{2}$$

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

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 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

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

when t= 1, and, ensuring continuity,

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

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 $\log(\frac{d\mu}{d\nu})$, 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

$$R^{\theta}_{P}(t) = R_t(P_{\theta}\ P_{1-\theta}) \tag{5}$$

Intuitively, observing a sample realization for which the likelihood ratio $dP_{\theta}/dP_{1-\theta}$ is high constitutes evidence that favors state θ over $1 - \theta$. For instance, in the case of $t = 2$, a higher value of $R^{\theta}_{P}(2)$ 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, Xiong, and Schneider (2012), Krishnamurthy, Kandasamy, Poczos, and Wasserman (2014)), computer science (Fritz (2017)), and quantum information (Horodecki, Horodecki, Horodecki, and Horodecki (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^{1}_{P} and R^{0}_{P} of an experiment are related by the identity

$$R^{\frac{1}{P}}(t) = \frac{t}{1-t} R^{0}_{P}(1 - t) \tag{6}$$

Hence the values of $R^{\theta}_{P}(t)$ for $\in [0, 1/2]$ tare determined by the values of $R^{1}_{P}{}^{-\theta}(t)$ on the interval [1/21] . 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 $t > 0$

$$R^{\theta}_{P}(t) > R^{\theta}_{Q}(t)$$

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 Section 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 nth 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 θ = 0 and θ = 1, as there exist pairs of experiments P and Q such that $R^1_P(t) > R^1_Q(t)$ for every t, but $R^0_P(t) < R^0_Q(t)$ for some t.[3]

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 $\log \frac{dP_1}{dP_0}$ and $\log \frac{dQ_1}{dQ_0}$ 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*:

(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 Section 5.7. The complete proof of Theorem 1 appears in Section 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 Section J in the Appendix in the Online Supplementary Material (Mu, Pomatto, Strack, and Tamuz (2021)) for discussion of this and another generalization. On the other hand, the theorem does not remain true if we remove the genericity assumption. In Section 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 nongeneric pair of

[3] 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 $R^1_P(t) > R^1_Q(t)$ for every $t > 0$, while $R^0_P(t) < R^0_Q(t)$ for $t > 2$.

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.

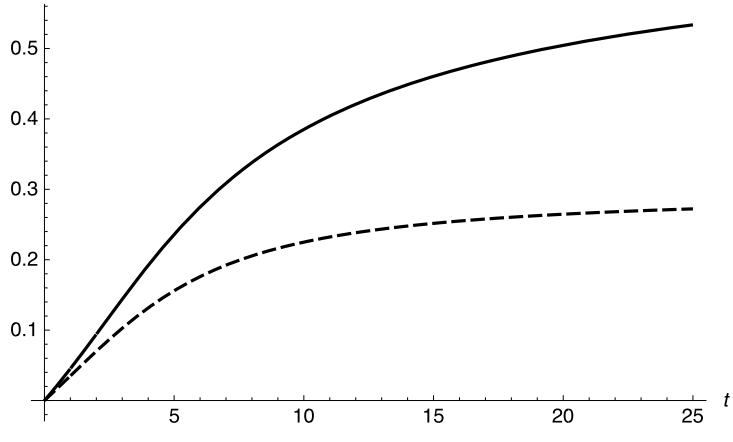
A natural alternative definition of "Blackwell dominance in large samples" would require $P^{\otimes n} \succeq Q^{\otimes n}$ to hold for *some* n, but the resulting order is in fact equivalent under our genericity assumption. This is a consequence of Theorem 1, $P^{\otimes n_0} \succeq Q^{\otimes n_0}$ becausefor any n_0 implies P dominates Q in the Rényi order, which in turn implies P Q for all large n.[4]

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 [01] , 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 {01 } . 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 [01] to {01 } that maps all signal realizations above 1/2 to 1 and all realizations below 1/2 to 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.



$\beta = 0.$

[4] However, it is not true that1 in Example 2 below provides an example where$P^{\otimes n_0}$ $Q^{\otimes n_0}$ for some n_0 implies$P^{\otimes 2}$PBlackwell dominates$^{\otimes n}$ $Q^{\otimes n}$ for all n ≥ n_0Q. The case of$^{\otimes 2}$, but $P^{\otimes 3}\alpha = 0.$does not305,

dominate $Q^{\otimes 3}$.

FIGURE 1.—The Rényi divergences R^0_P (solid line), and R^0_Q (dashed line) for $p = 0.63$ in Example 1. The comparison between R_{1P} and R_{1Q} yields a similar graph.

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

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

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

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

and

$$R_{0Q}(t) = \frac{1}{t-1}\log\left(2^{-t}\cdot\left(p^{1-t} + (1-p)^{1-t}\right)\right); \qquad R_{1Q}(t) = \frac{1}{t-1}\log\left(2^{t-1}\cdot\left(p^{t} + (1-p)^{t}\right)\right)$$

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 $0 \le \beta \le 1/4$ and $0 \le \alpha \le 1/2$. 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, 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 nontrivial 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:

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

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

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

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).

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 nontrivial*, *bounded experiments*. *Then*

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

Furthermore, the dominance ratio P/Q *is always positive*.[5]

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.

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

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

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

where a_n denotes the smallest integer greater than or equal to an.

from the quantum information literature (Duan, Feng, Li, and Ying (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, Strack, and Tamuz (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 $B(\Omega)$ its Borel σ-algebra and by $\Delta(\Omega)$ the collection of Borel probability measures on $B(\Omega)$. Given another Polish space Ξ, a measurable function $f : \Omega \to \Xi$ 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 $E \in B(\Xi)$.

Consider, for each Ω, a map

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

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 $f : \Omega \to \Xi$ it holds that

$$D^{\Xi}\left(f_*(\mu), f_*(\nu)\right) \leq D_{\Omega}(\mu\nu)$$

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.[6] 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 (P_0P_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, $d\mu/d\nu$ 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*

$$D(\mu\nu) = \int_{[1/2,\infty]} R_t(\mu\,\nu)\,dm_0(t) + \int_{[1/2,\infty]} R_t(\nu\,\mu)\,dm_1(t) \tag{7}$$

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.

[6] 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 labeling of the domain.

PROOF SKETCH OF THEOREM 2: The first key idea is to see a bounded pair of probability measures as a bounded experiment (P_0P_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_0P_1) dominates (Q_0Q_1) in the Rényi order, then, by Theorem 1, there exists a number n of repetitions such that $(P_0^nP_1^n)$ dominates $(Q_0^nQ_1^n)$ in the Blackwell order. Hence, by combining Blackwell monotonicity and additivity, we obtain that D must satisfy

$$nD(P^0, P_1) = D\left(P_0^n, P_1^n\right) \geq D\left(Q_0^n, Q_1^n\right) = nD(Q_0Q_1)$$

Hence, D is monotone in the Rényi order.

We deduce from this that D is a monotone functional $F(R^0{}_PR^1{}_P)$ 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. *Q.E.D.*

5. PROOF OF THEOREM 1

The proof of Theorem 1 is organized as follows. In Section 5.1 we first show that the Rényi order is necessary for the large sample order. The remaining subsections demonstrate sufficiency. In Section 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. Section 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.[7] 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

$$\int_0^1 v_1(p)\,d\pi(p) = \int_\Omega \left(\frac{dP_1(\omega)}{dP_0(\omega)}\right)^{t-1} dP_1(\omega) = e^{(t-1)R_p^1(t)} \tag{8}$$

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

$$\pi(p) \quad -\big(\ \pi\ (p) \qquad \pi\ (p)\big) \qquad\qquad 1 \qquad p$$

[7] 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.

$$d = d_1 + d_0 \quad \text{and} \quad d\pi_1(p) = \frac{p}{1}\, d\pi_0(p) \tag{9}$$

$$2$$

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

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

The first equality in (8) then follows from a change of variable from signal realizations ω to posterior beliefs $p = \frac{dP_1(\omega)}{dP_1(\omega)+dP_0(\omega)}$ (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 $2t(1-t) \times p^{t-2}(1-p)^{-1-t}$). Then

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

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 $v_3(p) = 2p\log(\frac{1}{1-p})$, which is strictly convex with a second derivative of $2p^{-1}(1-p)^{-2}$. We have

$$\int_0^1 v_3(p)\, d\pi(p) = \int_0^1 \log\left(\frac{p}{1-p}\right) d\pi_1(p) = \int_\Omega \log\left(\frac{dP_1(\omega)}{dP_0(\omega)}\right) dP_1(\omega) = R^1_P(1)$$

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

$$\log \frac{dP^{\theta}}{dP^{1-\theta}}^{\otimes n} = (\Omega^n, P_0^n, P_1^n)$$

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

$$\log \overline{\quad}_n(\omega) = \log \frac{1}{dP}(\omega_i \frac{dP_{1n}}{dP_{0}} \quad \frac{dP}{dP_0}$$

Moreover, the random variables

$$X_i(\omega) \quad n \qquad = \log \frac{1}{dP_0}(\omega_i) \quad i = 1$$

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,

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

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

$$\tilde{L}_1 = \log \frac{dP^1}{dP_0} - E$$

where E is a random variable that, under P_1, is independent from $\log \frac{dP_1}{dP_0}$ and distributed according to an exponential distribution with support R_+ and cumulative distribution function $1 - e^{-x}$ 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 $a \in R$.

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

$$\tilde{F}_1(a) = \int_{\mathbb{R}} P_1\big(\{-E \le a - u\}\big)\, dF_1(u) = F_1(a) + e^a \int_{(a,\infty)} e^{-u}\, dF_1(u) \tag{11}$$

The next result shows that the Blackwell order over experiments can be reduced to firstorder 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*

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

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

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

must satisfy $\Lambda_\pi(p) \ge \Lambda_\tau(p)$ 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

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

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

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

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

$$2\Lambda_\pi(p) = \frac{e^a}{1+e^a}\left(2 - \int_{(a,\infty)} \frac{1+e^u}{e^u}\, dF_1(u)\right) - F_1(a) \tag{14}$$

Since

$$\int_{(a,\infty)} \frac{1+e^u}{e^u}\, dF_1(u) = \int_{(a,\infty)} e^{-u}\, dF_1(u) + 1 - F_1(a)$$

(14) leads to

$$2\Lambda_\pi(p) = \frac{e^a}{1+e^a}\left(\frac{F_1(a)}{e^a} - \frac{e^a}{1+e^a}\right) + \int_{(a,\infty)} e^{-u}\,dF_1(u) = \frac{e^a}{1+e^a} - \frac{\tilde{F}_1(a)}{1+e^a}$$

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.

Q.E.D.

Intuitively, transferring probability mass from lower to higher values of $\log(dP_\theta/dP_{1-\theta})$ 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 dP_1/dP_0 must satisfy the change of measure identity $\int \frac{dP_0}{dP_1}\,dP_1 = 1$, the distribution F_1 must satisfy

$$\int_{\mathbb{R}} e^{-u}\,dF_1(u) = 1$$

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.[8]

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 $a > E[X]$ 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

[8] 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).

the probability of this event is determined by the *cumulant generating function* of X, defined as

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

for every $t \in R$.

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

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

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

Cramér's theorem establishes that for each threshold $a > E[X]$, 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 $E[X]$ and $E[Y]$. 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 $P[X_1 + \cdots + X_n > na] \geq P[Y_1 + \cdots + Y_n > na]$

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* $a \geq E[Y]$, *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*

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

The condition $K_Y^*(a) - \eta > K_X^*(a + \eta)$ ensures that the rate at which the probability of the events $\{Y_1 + \cdots + Y_n > na\}$ vanish with n is larger by a factor of at least η than the rate of the events $\{X_1 + \cdots + X_n > n(a + \eta)\}$. 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 $\mathsf{P} = (\Omega P_0 P_1)$ and $\mathsf{Q} = (\Xi Q_0 Q_1)$. Denote the corresponding log-likelihood ratios

$$X^\theta = \log \frac{dP^\theta}{dQ_{-\theta}} \quad \text{and} \quad Y^\theta = \log \frac{dQ_\theta}{dP_{1-\theta}}$$

defined over the probability spaces (ΩP_θ) and (ΞQ_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):

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

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:

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

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

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

[12] Throughout the proof, we assume $\theta \neq 0$ Q is a nontrivial experiment, so that $E[Y_\theta]$ being the Kullback–Leibler divergence between Q and $Q_{-\theta}$ is strictly positive. This is without loss, as P clearly dominates Q (in large samples) in case Q is trivial.

between K_{X^θ} and K_{Y^θ} established in (18) and (19) is more complicated, and implies the following ranking of their conjugates:

$$K_{Y^*_\theta}(a) > K_{X^*_\theta}(a) \qquad \text{for } \mathbb{E}[X^\theta] \leq a \leq \max[Y^\theta]$$

$$K_{Y^*_\theta}(a) < K_{X^*_\theta}(a) \qquad \text{for } 0 \leq a \leq \mathbb{E}[Y^\theta]$$

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* $\theta \in \{0, 1\}$:

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

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

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, we need to show that there exists a sample size n_0 such that for all $n \ge n_0$, the perfected loglikelihood 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 nth convolution power of F_θ, which represents the distribution of log-likelihood ratios under the product experiment $P^{\otimes n}$; similarly define $G^*_\theta{}^n$. By Lemma 1, it suffices to show that for $n \ge n_0$ it holds that

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

and

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

Below we show (20); the argument for (21) is identical after relabeling 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 $E[X^1] - \eta > E[Y^1]$.[9]

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

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

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

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

Case 2: $E[X^1] - \eta \le a < \max[Y^1]$. From Lemma 2, we have that

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

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

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 $n \geq 4b^2(1+\eta)\eta^{-2}$. Since $\eta < 1$, it holds for all $n \geq n_0 = 8b^2\eta^{-2}$.

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

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

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

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

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 $K_{-Z}(t) = \log \mathbb{E}[e^{t(-Z)}] = \log \mathbb{E}[e^{(-t)Z}] = K_Z(-t)$, 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 $-a > -\mathbb{E}[Y] - \eta > \mathbb{E}[-X]$. This yields

for all n

≥ 4 1 + −

22) for all $a \geq$

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

b^2 [3]. Hence (n [14] (η)η 22)

holds for ≥ n_0.

This proves (0 and completes the proof of Theorem 1.

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

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 *likelihood ratios* $[-b, b]$ *and* Q *be a generic pair of bounded experiments, with log- taking values in . 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.

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) proposed 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, de- fined, in our notation, as the minimum over $t \in (0, 1)$ of the function $e^{(t-1)R_P(t)}$ (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.[10]

(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) showed 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.[11]

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]

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) asked 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 Section L in the Appendix.

Fritz (2018) used 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, asked a question in the spirit of our dominance ratio, and proved 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 Section K in the Appendix, we identify a set of *necessary conditions* for large sample dominance. These

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

[11] 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

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 (Theorem 3).[12] 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 Section 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 (socalled "Cramér's condition").[13]

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.

APPENDIX A: LARGE DEVIATIONS

$P^{\otimes n}$ is higher than that from $Q^{\otimes n}$ whenever $n \geq n_0$. Each Rényi divergence $R^{\theta}_{P}(t)$ corresponds to the expected payoff in some decision problem (see Section 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 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 Section L in the Appendix for details.

[12] 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).

[13] Although Cramér's result that $\log P[X^1 + \cdots + X^n > na] \sim -n \cdot K^{X*}(a)$ 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.

For every bounded random variable X that is not a constant, we denote by $M_X(t) = \log E[e^{tX}]$ 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 $t \in R$. This solution t is nonnegative if and only if $a \geq E[X]$. In addition, as $K_X(0) = 0$, $K_X^*(a)$ is nonnegative. The function $K_X^*(a)$ is continuous (in fact, analytic) wherever it is finite.

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

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

The next proposition gives a lower bound for this probability.

LEMMA 3: *Let $X_1, \ldots, X_n$ 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*

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

PROOF: We first consider the case where $a \geq E[X] - \eta/2$. 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 nonnegative finite number, since $E[X] \leq a + \eta/2 < \max X$.

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

$$\frac{\hat{\nu}}{\nu}(x) = \frac{e^{tx}}{E\, e^{tX}} = \frac{d}{dK_X(t)} e^{tx}$$

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

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

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

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

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

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

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

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

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

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

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

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,

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

We have thus shown that

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

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

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

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

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

Hence

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

Since K_X^* is nonnegative, we again have

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

This proves the lemma. *Q.E.D.*

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 $n \ge 1$,

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

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

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

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

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

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

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

This proves the desired result (16).

APPENDIX 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

$$\tilde{F}_1(a) = \int_0^\infty F_1(a+u)e^{-u}\,du \qquad \tilde{G}_1(a) = \int_0^\infty G_1(a+u)e^{-u}\,du$$

and likewise

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:

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

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

$$\tilde{F}_1(a) = \int_{-\infty}^a dF_1(u) + e^a \int_a^\infty e^{-u}\,dF_1(u) \tag{24}$$

The key observation is that $dF_1(u) = -e^u\,dF_0(-u)$. Indeed, $dF_1(u)$ is the density under state 1 that the log-likelihood ratio $\log(dP_1/dP_0)$ is equal to u, which is also the density under state 1 that the opposite log-likelihood ratio $\log(dP_0/dP_1)$ 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 $dF_1(u) = -e^u\,dF_0(-u)$ into (24), we have

$$\begin{aligned}\tilde{F}_1(a) &= \int_{-\infty}^a -e^u\,dF_0(-u) + e^a \int_a^\infty -dF_0(-u) \\ &= \int_{-a}^\infty e^{-v}\,dF_0(v) + e^a F_0(-a)\end{aligned}$$

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

APPENDIX 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 $0 < E[Y^\theta] < E[X^\theta]$ and $\max[Y^\theta] < \max[X^\theta]$.

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

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

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 $a = E[X^\theta]$, but in that case the second inequality is strict because a is strictly above the expectation of Y. 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 $_\theta(a+\eta) - K^*_{X^\theta}(a) < \frac{\varepsilon_1}{2}$ 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^*} for all a in this interval. If in addition $\eta < \min\{\varepsilon_2, \varepsilon_2\}$,

then

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

$$X \qquad Y \qquad 2 \qquad 2 \qquad X$$

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

As for the second half, consider a point θ[14] a. $\in [0, \mathbb{E}[Y^\theta]]$ Since $a \leq E[Y^\theta]$ and $a \geq 0 > \min[Y]$, there exists a finite $\tilde{t} \leq 0$ such that $K^*_{Y^\theta}$ $= \tilde{t}a -$ $(a)\, K_{Y\theta}(\tilde{t})$. This $\tilde{t}$ satisfies

$K_{Y\theta}(\tilde{t}) = a.$ θ 1θ

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

$$K_{Y\theta}(t) = K_{Y1-\theta}(-t-1)$$

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 $K_{Y\theta}$ is increasing, we have $\tilde{t} \in (-1, 0]$. Dominance in the Rényi order then implies, by (19),

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

$$K_Y^*(a) = \tilde{t}a - K_{\theta}(\tilde{t}) \le \tilde{t}a - K_{\theta}(\tilde{t}) \le K_{X\theta}^*(a)$$

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. Using continuity as before, any sufficiently small η makes $K_{Y^{*}\theta}(a-\eta) < K_{X^{\theta}}^{*}(a) - \eta \in [0, \mathbb{E}[Y^{\theta}]]$ hold for all a in the slightly bigger interval $[0, \mathbb{E}[Y^{\theta}] + \eta]$. Hence the lemma holds.

APPENDIX 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 $\pi([p_2 p_3]) > \pi(\{p_3\})$.

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 meanpreserving 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, $R^{\theta}_P(t) > R^{\theta}_Q(t)$ for all $\theta \in \{0, 1\}$ and $t > 0$.

We will perturb Q to be a slightly more informative experiment $\tilde{Q}$, such that P still dominates $\tilde{Q}$ 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 $\tilde{Q}$ 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, $R^{\theta}_P(t) > R^{\theta}_{\tilde{Q}}(t)$ still holds when ε is sufficiently small.[15] Since P and $\tilde{Q}$ also form a generic pair, Theorem 1 shows that P dominates $\tilde{Q}$ 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

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

[15] Using the relation between $R^{0}_P(t)$ and $R^{1}_P(1-t)$, it suffices to show $R^{\theta}_P(t) > R^{\theta}_{\tilde{Q}}(t)$ for $\theta \in \{0, 1\}$ and $t \geq 1/2$. Fixing a large T, then by uniform continuity, $R^{\theta}_P(t) > R^{\theta}_Q(t)$ implies $R^{\theta}_P(t) > R^{\theta}_{\tilde{Q}}(t)$ for $t \in [\frac{1}{2}, T]$ 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.

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.[16] Hence Q achieves higher expected utility than P, implying that it is not Blackwell dominated.

APPENDIX E: PROOF OF PROPOSITION 2

It is easily checked that the condition $R_P^1(1/2) > R_Q^1(1/2)$ reduces to

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

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

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

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

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

To prove these, it suffices to consider the α that makes (25) hold with equality.[17] 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:

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

)

[17] It is clear that the inequalities are easier to satisfy when α increases in the range. [25]This follows from Rolle's theorem and an induction argument.

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

When (25) holds with equality, we have $(0) = \Delta(\frac{1}{2}) = \Delta(1) = 0$. Thus has roots at 0, 1 as well as a double-root at $\frac{1}{2}$. But since 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 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 $(r) \to \infty$ as $r \to \infty$. Thus (r) is strictly positive for $r \in (1, \infty)$. As $(1) = 0$, its derivative is weakly positive. But recall that we have enumerated the 4 roots of . So cannot have a double-root at r = 1, and it follows that (1) is strictly positive. Hence (28) holds.

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

APPENDIX F: PROOF OF PROPOSITION 3

Denote $= \inf_{\theta,t} \frac{R_P^{\theta}(t)}{R_Q^{\theta}(t)}$ r. We would like to show that P/Q = r. Let n, m be such that $P^{\otimes n}$ $Q^{\otimes m}$. Then, since ranking of the Rényi divergences is a necessary condition for Blackwell dominance, and by the additivity of Rényi divergences, nfor alland t > 0. Thus any such m/n is bounded above by r, and so P/Q ≤ r. $\theta \in \{0, 1\}$ $\cdot R_P^{\theta}(t) \geq m \cdot R_Q^{\theta}(t)$

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 $\overline{R_Q^{\theta}(t)}$ $R_{\theta}(t)$ ^{n} that $\lim_{t\to\infty}$ ^{P} > m/n implies the pair $P^{\otimes}$ and $Q^{\otimes m}$ is generic. Therefore, by Theorem 1, we have that for some k large enough, $P^{\otimes nk}$ $Q^{\otimes mk}$. Thus P/Q ≥ mk/nk = m/n. Since this holds for every rational m/n that is less than r, we can conclude that P/Q ≥ r. Finally, note that each of the functions R^{θ}_P and R^{θ}_Q are positive, increasing and bounded on $(0, \infty)$. Furthermore, using

$$R_{\theta P}(t) \qquad \overline{R_{1P-\theta}(1-t)\, R_{\theta Q}} = \qquad {}_{1-\theta}(1-t)$$

$$(t) \qquad R_Q$$

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

$$R^{\theta}(t) \qquad R^{\theta}(t)\,(t)$$

Recall that $R^{\theta}_{P}(t)$, $R^{\theta}_{Q}(t)$ are $P/Q = \inf_{\substack{\theta\in\{0,1\}\\ t>0}} \frac{P}{R^{\theta}_{Q}(t)} = \inf_{\substack{\theta\in\{0,1\}\\ t\geq\frac{1}{2}}} \frac{P}{R^{\theta}_{Q}}$ 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

AUBRUN, G., AND I. NECHITA (2008): "Catalytic Majorization and $_{p}$ Norms," *Communications in Mathematical Physics*, 278 (1), 133–144. [496]

AZRIELI, Y. (2014): "Comment on "The Law of Large Demand for Information"," *Econometrica*, 82 (1), 415– 423. [476,479,482,483,495]

BLACKWELL, D. (1951): "Comparison of Experiments," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, 93–102. [475,476,479,495]

_____ (1953): "Equivalent Comparisons of Experiments," *The Annals of Mathematical Statistics*, 24 (2), 265–272. [478]

BLACKWELL, D. A., AND M. A. GIRSHICK (1979): *Theory of Games and Statistical Decisions*. Courier Corporation. [476]

BOHNENBLUST, H. F., L. S. SHAPLEY, AND S. SHERMAN (1949): "Reconnaissance in Game Theory," Unpublished Manuscript, Rand Corporation. [478]

CABRALES, A., O. GOSSNER, AND R. SERRANO (2013): "Entropy and the Value of Information for Investors," *American Economic Review*, 103 (1), 360–377. [496]

CRAMÉR, H. (1938): "Sur un nouveau théoreme-limite de la théorie des probabilités," *Actual. Sci. Ind.*, 736, 5–23. [491]

CRITCHLEY, F., P. MARRIOTT, AND M. SALMON (1996): "On the Differential Geometry of the Wald Test With Nonlinear Restrictions," *Econometrica*, 64 (5), 1213–1222. [476]

CSISZÁR, I. (2008): "Axiomatic Characterizations of Information Measures," *Entropy*, 10 (3), 261–273. [477, 479,485]

DUAN, R., Y. FENG, X. LI, AND M. YING (2005): "Multiple-Copy Entanglement Transformation and Entanglement Catalysis," *Physical Review A*, 71 (4), 042319. [485]

FRITZ, T. (2017): "Resource Convertibility and Ordered Commutative Monoids," *Mathematical Structures in Computer Science*, 27 (6), 850–938. [479,485]

_____ (2018): "A Generalization of Strassen's Positivstellensatz and Its Application to Large Deviation Theory," ArXiv preprint. Available at arXiv:1810.08667v3. [496]

HELLMAN, Z., AND E. LEHRER (2019): "Valuing Information by Repeated Signals," Unpublished Manuscript, Bar-Ilan University. [495]

HONG, Y., AND H. WHITE (2005): "Asymptotic Distribution Theory for Nonparametric Entropy Measures of Serial Dependence," *Econometrica*, 73 (3), 837–901. [476]

HORODECKI, R., P. HORODECKI, M. HORODECKI, AND K. HORODECKI (2009): "Quantum Entanglement," *Reviews of modern physics*, 81 (2), 865. [479]

JENSEN, A. K. (2019): "Asymptotic Majorization of Finite Probability Distributions," *IEEE Transactions on Information Theory*, 65 (12), 8131–8139. [479,496]

JEWITT, I. (2007): "Information Order in Decision and Agency Problems,". Nuffield College. Unpublished Manuscript. [497]

KELLY, J. L. (1956): "A New Interpretation of Information Rate," *IRE Transactions on Information Theory*, 2 (3), 185–189. [496]

KITAMURA, Y., AND M. STUTZER (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65 (4), 861–874. [476]

KITAMURA, Y., T. OTSU, AND K. EVDOKIMOV (2013): "Robustness, Infinitesimal Neighborhoods, and Moment Restrictions," *Econometrica*, 81 (3), 1185–1201. [476]

KRISHNAMURTHY, A., K. KANDASAMY, B. POCZOS, AND L. WASSERMAN (2014): "Nonparametric Estimation of Renyi Divergence and Friends," in *International Conference on Machine Learning*, 919–927. [479]

LIESE, F., AND I. VAJDA (2006): "On Divergences and Informations in Statistics and Information Theory," *IEEE Transactions on Information Theory*, 52 (10), 4394–4412. [479]

LINDLEY, D. V. (1956): "On a Measure of the Information Provided by an Experiment," *The Annals of Mathematical Statistics*, 27 (4), 986–1005. [496]

MOSCARINI, G., AND L. SMITH (2002): "The Law of Large Demand for Information," *Econometrica*, 70 (6), 2351–2366. [477,495]

MU, X., L. POMATTO, P. STRACK, AND O. TAMUZ (2021): "Supplement to 'From Blackwell Dominance in Large Samples to Rényi Divergences and Back Again'," *Econometrica Supplemental Material*, 89, https://doi.org/10.3982/ECTA17548. [481]

PÓCZOS, B., L. XIONG, AND J. SCHNEIDER (2012): "Nonparametric Divergence Estimation With Applications to Machine Learning on Distributions," ArXiv preprint. Available at arXiv:1202.3758. [479]

POMATTO, L., P. STRACK, AND O. TAMUZ (2018): "The Cost of Information," Unpublished Manuscript, California Institute of Technology. [485]

_____ (2020): "Stochastic Dominance Under Independent Noise," *Journal of Political Economy*, 128 (5), 1877–1900. [491]

RÉNYI, A. (1961): "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. [476,477,479,485]

SAWA, T. (1978): "Information Criteria for Discriminating Among Alternative Regression Models," *Econometrica*, 46 (6), 1273–1291. [476]

SMITH, L., AND P. SØRENSEN (2000): "Pathological Outcomes of Observational Learning," *Econometrica*, 68 (2), 371–398. [481]

STEIN, C. (1951): "Notes on the Comparison of Experiments," Unpublished Manuscript, University of Chicago. [476,495]

TORGERSEN, E. (1985): "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*. Springer, 259–310. [496]

_____ (1991): *Comparison of Statistical Experiments*, Vol. 36. Cambridge University Press. [479] TORGERSEN, E. N. (1970): "Comparison of Experiments When the Parameter Space Is Finite," *Probability Theory and Related Fields*, 16 (3), 219–249. [476,479,495]

ULLAH, A. (2002): "Uses of Entropy and Divergence Measures for Evaluating Econometric Approximations and Inference," *Journal of Econometrics*, 107 (1-2), 313–326. [476]

WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50 (1), 1–25. [476]

Co-editor Ulrich K. Müller handled this manuscript.

Manuscript received 5 August, 2019; final version accepted 4 September, 2020; available online 11 September, 2020.