Blackwell Dominance in Large Samples*

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

July 27, 2019

Abstract

We study repeated independent Blackwell experiments. Standard examples include drawing multiple samples from a population, or performing a measurement in different locations. In the baseline setting of a binary state of nature, we compare experiments in terms of their informativeness in large samples. Addressing a question due to Blackwell (1951) we show that generically, an experiment is more informative than another in large samples if and only if it has higher Rényi divergences. As an application of our techniques we in addition provide a novel characterization of k^{th}-order stochastic dominance as second-order stochastic dominance of large i.i.d. sums.

1 Introduction

Statistical experiments form a general framework for modeling information: Given a set Θ of parameters, an *experiment* (or *information structure*) 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 a binary hypothesis $\theta \in \{0, 1\}$, based on random samples drawn from one of two experiments P or Q. According to Blackwell's ordering, P is more informative than Q if, for every test performed based on observations produced by Q, there exists another test based on P that has lower probabilities of both

*We would like to thank Kim Border, Laura Doval, Federico Echenique, Tobias Fritz, George Mailath, Massimo Marinacci, Margaret Meyer, Marco Ottaviani and Peter Norman Sørensen for their comments and suggestions.

†Columbia University. Email: xm2230@columbia.edu. Xiaosheng Mu appreciates the hospitality of the Cowles Foundation at Yale University, which hosted him during parts of this research.

‡Caltech. Email: luciano@caltech.edu.

§UC Berkeley. Email: pstrack@berkeley.edu.

¶Caltech. Email: tamuz@caltech.edu. Omer Tamuz was supported by a grant from the Simons Foundation (#419427).

Type-I and Type-II errors (Blackwell and Girshick, 1979). This is a strong notion of informativeness which needs to apply if only *one* sample is produced by each experiment.

In many applications, the information produced by an experiment does not consist of a single observation but of multiple i.i.d. samples. 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 Torgersen (1970) and Azrieli (2014). However, identifying the precise conditions under which this phenomenon can occur 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 first is more informative in terms of Rényi divergences (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 asymptotic informativeness of an experiment.

The result crucially relies on two ingredients. First, the proof uses 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.

By applying the methods we employed to study experiments, in the second part of the paper we establish new laws of large numbers for comparing sums of i.i.d. random variable in terms of stochastic dominance. We say that a random variable X k^{th}-order dominates a random variables Y *in large numbers* if for large n, the sum $X_1 + \cdots + X_n$ of n i.i.d. copies of X dominates the sum of n i.i.d. copies of Y with respect to k-th order stochastic dominance. Stochastic dominance in large numbers is implied by, but strictly weaker than, stochastic dominance between X and Y. We focus on stochastic dominance for its many applications, both in economics and in other fields, as well as for its conceptual simplicity. In the same way the classic law of large numbers provides non-parametric predictions about long-run frequencies, stochastic dominance provides unambiguous choice predictions that are independent of the decision maker's preferences. Dominance in large numbers is a likewise natural criterion for comparing prospects that involve repeated independent risks, such as the return of a long-term investment.

We show that stochastic dominance in large numbers admits a simple characterization in terms of the unanimous rankings of CARA preferences. Theorem 3 shows that all CARA agents rank a gamble X above a gamble Y if and only if X first-order dominates Y in large numbers. Second-order dominance is characterized by the unanimous rankings of risk-averse and the opposite rankings of risk-loving CARA preferences (Theorems 4 and 5). These results shed new light on a classic question, raised first by Samuelson (1963), on the relation between preferences over one-shot gambles and preferences over repeated gambles.

Finally, we provide a novel characterization of higher order stochastic dominance, showing that dominance in the k^{th} order for *some* $k \geq 2$ is equivalent to second-order stochastic dominance in large numbers, which in turn is equivalent to k^{th}-order dominance in large numbers for *every* $k \geq 2$. Thus, perhaps surprisingly, higher-order attitudes over risk (e.g., prudence, temperance, etc.) reduce to simple risk aversion when considering i.i.d. sums of gambles.

The paper is organized as follows. In §2 we provide our main definitions. Section §3 contains the characterization of Blackwell dominance in large samples, while §4 presents a proof sketch. In §5 we study stochastic dominance in large numbers. Finally, we further discuss our results and their relation to the literature in §6.

2 Model

2.1 Statistical Experiments

A state of the world θ can take two possible values, 0 or 1. A *Blackwell-Le Cam experiment* $P = (\Omega, P_0, P_1)$ consists of a sample space Ω and a pair of probability measures (P_0, P_1) defined on a σ-algebra $\mathcal{A}$ of subsets of Ω, with the interpretation that $P_\theta(A)$ is the probability of observing $A \in \mathcal{A}$ in state $\theta \in \{0, 1\}$. To ease the exposition we will suppress the σ-algebra $\mathcal{A}$ from the notation. This framework is commonly encountered in simple hypothesis tests as well as in information economics. In §6 we discuss the case of experiments for more than two states.

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

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

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

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

the n-fold product experiment where n conditionally 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. Fixing a uniform prior simplifies the discussion, but it is without loss of generality. To each experiment $P = (\Omega, P_0, P_1)$ we can associate a Borel probability measure π over $[0,1]$ that represents the distribution over posterior beliefs induced by the experiment. Formally, let $p(\omega)$ be the posterior belief that the state is 1 given the realization ω:

$$p(\omega) = \frac{\ell(\omega)}{1+\ell(\omega)} \quad \text{where} \quad \ell(\omega) = \frac{\mathrm{d}P_1}{\mathrm{d}P_0}(\omega).$$

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

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

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. We say that P is *trivial* if $P_0 = P_1$, and *bounded* if the support of π is strictly included in $(0,1)$.

2.2 The Blackwell Order

We first review the main concepts behind Blackwell's order over experiments (Bohnenblust, Shapley, and Sherman, 1949; Blackwell, 1953). Consider two experiments P and Q and their induced distribution over posterior beliefs denoted by π and τ, respectively. The experiment P *Blackwell dominates* Q, denoted $P \succeq Q$, if

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

for every continuous convex function $v\colon (0,1) \to \mathbb{R}$. We write $P \succ Q$ if $P \succeq Q$ and $Q \not\succeq P$. So, $P \succ Q$ if and only if (1) holds with a strict inequality whenever v is strictly convex.

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

Blackwell's theorem shows that the order $\succeq$ can be equivalently defined by "garbling" operations: Intuitively, $P \succeq Q$ if and only if the outcome of the experiment Q can be generated from the experiment P by compounding the latter with additional noise, without adding further information about the state.[1]

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

[1] Formally, given two experiments $P = (\Omega, P_0, P_1)$ and $Q = (\Xi, Q_0, Q_1)$, $P \succeq Q$ iff there is a measurable

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

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

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

A natural alternative definition would require $P^{\otimes n} \succeq Q^{\otimes n}$ to hold for *some* n, but as we show below the resulting order is equivalent under a mild genericity assumption.

2.3 Rényi Divergence and the Rényi Order

Our main result relates Blackwell dominance in large samples to a well-known notion of informativeness due to Rényi (1961). Given an experiment (Ω, P_0, P_1), a state θ, and parameter $t > 0$, the Rényi t-divergence is defined as

$$R_P^\theta(t) = \frac{1}{t-1} \log \int \left(\frac{\mathrm{d}P_\theta}{\mathrm{d}P_{1-\theta}}(\omega) \right)^{t-1} \mathrm{d}P_\theta \tag{3}$$

when $t \neq 1$, and, to ensure continuity,

$$R_P^\theta(1) = \int_\Omega \log \frac{\mathrm{d}P_\theta}{\mathrm{d}P_{1-\theta}}(\omega)\, \mathrm{d}P_\theta(\omega).$$

Equivalently, $R_P^\theta(1)$ is the Kullback-Leibler divergence between the measures P_θ and $P_{1-\theta}$.

Intuitively, observing a sample realization for which the likelihood-ratio $(\mathrm{d}P_\theta/\mathrm{d}P_{1-\theta})(\omega)$ is high constitutes evidence that favors state θ over $1-\theta$. For instance, in case $t = 2$, a higher value of $R_P^\theta(2)$ describes an experiment that, in expectation, more strongly produces evidence in favor of the state θ when this is the correct state. Varying the parameter t allows to consider different moments for the distribution of likelihood ratios. The Rényi divergence has found applications to statistics and information theory (Liese and Vajda, 2006; Csiszár, 2008), machine learning (Póczos et al., 2012; Krishnamurthy et al., 2014), computer science

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

Fritz (2017), and quantum information (Horodecki et al., 2009; Jensen, 2019). Another well known measure of informativeness is the Hellinger transform (Torgersen, 1991, p.39), which amounts to a monotone transformation of the Rényi divergences of an experiment.

Rényi Order. We say that an experiment P dominates an experiment Q in the *Rényi order* if it holds that

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

for all $\theta \in \{0, 1\}$ and all $t > 0$. The Rényi order is a refinement 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 utility with respect to each decision problem in this family.

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

$$R_P^\theta = R_Q^\theta + R_T^\theta$$

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

3 Characterization

We call two bounded experiments P and Q *generic* if the essential maxima of the log-likelihood ratios $\log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}$ and $\log \frac{\mathrm{d}Q_1}{\mathrm{d}Q_0}$ are different, and if their essential minima are also different.

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.*

It seems difficult to obtain an applicable characterization of large sample dominance without imposing any genericity conditions. In §G in the appendix we discuss the knife-edge case where the maxima or the minima of the supports might be equal, and show that the conclusions of Theorem 1 no longer hold.

3.1 An Example

In this section we introduce a new example of two experiments that are not Blackwell ranked, but are ranked in large samples.

The experiment P appears in Smith and Sørensen (2000). The signal space is the interval $[0,1]$, and the conditional measures P_0 and P_1 are absolutely continuous with densities $f_0(s) = 1$ and $f_1(s) = 1/2 + s$.

Our second experiment Q is binary, with signal space $\{0,1\}$. The conditional measure Q_0 assigns probability $1/2$ to both signals, while the other conditional measure is $Q_1(1) = p$ and $Q_1(0) = 1 - p$.

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

Nevertheless, if we choose $p = 0.63$, then as Figure 1 below suggests, P dominates Q in the Rényi order even though the two experiments are not Blackwell ranked.[2]

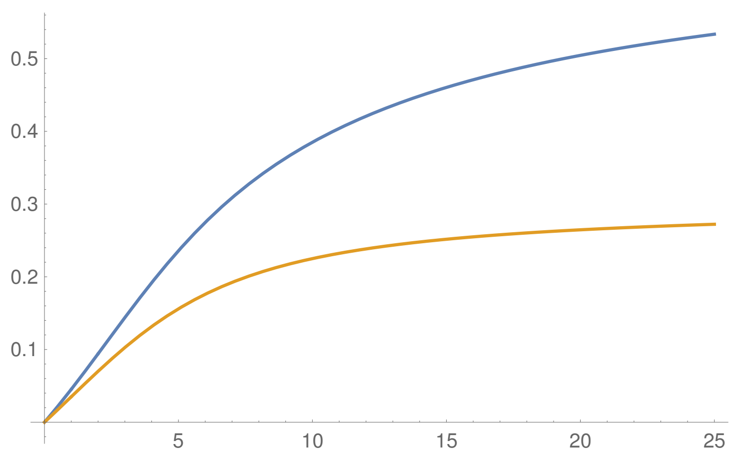

Figure 1: The Rényi divergences R^0_P (blue), and R^0_Q (orange) for $p = 0.63$. The comparison between R^1_P and R^1_Q yields a similar graph.

[2]The Rényi divergences as defined in (3) are computed to be

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

and

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

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

We end this section with a proposition that generalizes the example above, 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.

3.2 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 conditional probabilities of observing each signal realization given the state θ:

P :

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

Q :

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

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

Azrieli (2014) points out that a necessary condition for P to dominate Q in large samples is that the Rényi divergences are ranked at $1/2$, that is $R_P^1(1/2) > R_Q^1(1/2)$.[3] In addition, he conjectures it is also a sufficient condition, and proves it in the special

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

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

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

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.[4]

Proposition 2. *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.*

4 Overview of the Proof and Methods

In this section we illustrate the main ideas and insights behind the proof of Theorem 1. In §4.2 we point out a novel characterization of Blackwell dominance that reduces the comparison of experiments to the comparison in terms of first-order stochastic dominance of an appropriate statistic of the experiments. In §4.3 we use this observation to apply uniform large-deviation techniques. Omitted proofs are deferred to the Appendix.

4.1 Repeated Experiments and Log-Likelihood Ratios

The distribution over posteriors induced by a product experiment $P^{\otimes n}$ can be difficult to analyze directly. A more suitable approach consists in studying the distribution of the induced log-likelihood ratio

$$\log \frac{\mathrm{d}P_\theta}{\mathrm{d}P_{1-\theta}}. \tag{4}$$

As is well known, given a repeated experiment $P^{\otimes n} = (\Omega^n, P^n_0, P^n_1)$, its log-likelihood ratio satisfies, for every realization $\omega = (\omega_1, \ldots, \omega_n)$ in Ω^n,

$$\log \frac{\mathrm{d}P^n_1}{\mathrm{d}P^n_0}(\omega) = \sum_{i=1}^{n} \log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}(\omega_i)$$

and moreover the random variables

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

are i.i.d. under P^n_θ, 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.

[4] We mention that in general, $P^{\otimes n}$ Blackwell dominates $Q^{\otimes n}$ does not imply dominance for *every* sample size $N \geq n$. The case of $\alpha = 0.305$, $\beta = 0.1$ provides a counterexample of this kind, where $P^{\otimes 2}$ Blackwell dominates $Q^{\otimes 2}$, but $P^{\otimes 3}$ does not dominate $Q^{\otimes 3}$. Nonetheless, for generic pairs of experiments P and Q, dominance for some sample size n does imply dominance for all large sample size N. This is further discussed in §6.

4.2 From Blackwell Dominance to First-Order Stochastic Dominance

Our first result is a novel characterization of the Blackwell order expressed in terms of the distributions of the log-likelihood ratios (4). Given two experiments $P = (\Omega, P_0, P_1)$ and $Q = (\Xi, Q_0, Q_1)$ we denote by F_θ and G_θ, respectively, the cumulative distribution function of the log-likelihood ratios conditional on state θ. That is,

$$F_\theta(a) = P_\theta\left(\left\{\log \frac{\mathrm{d}P_\theta}{\mathrm{d}P_{1-\theta}} \leq a\right\}\right) \quad a \in \mathbb{R},\ \theta \in \{0,1\} \tag{5}$$

and G_θ, $\theta \in \{0,1\}$, are defined analogously using Q_θ.

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

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

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

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

$$\tilde{F}(a) = \int_{\mathbb{R}} (1 - H(u-a))\mathrm{d}F_1(u) = F_1(a) + \mathrm{e}^a \int_{(a,\infty)} \mathrm{e}^{-u} \mathrm{d}F_1(u). \tag{6}$$

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

Theorem 2. *Let P and Q be two experiments, and let $\tilde{F}$ and $\tilde{G}$, respectively, be the associated distributions of perfected log-likelihood ratios. Then*

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

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

$$\Lambda_\pi(p) = \int_0^p (p-q)\mathrm{d}\pi(q) \quad \text{and} \quad \Lambda_\tau(p) = \int_0^p (p-q)\mathrm{d}\tau(q) \tag{7}$$

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

We now express (7) in terms of the distributions of log-likelihood ratios F_θ and G_θ. We have

$$\Lambda_\pi(p) = p\left(\frac{1}{2} - \int_{(p,1]} 1\,\mathrm{d}\pi\right) - \int_0^p q\,\mathrm{d}\pi(q). \tag{8}$$

Using the fact that $q\,\mathrm{d}\pi(q) = \frac{1}{2}\,\mathrm{d}\pi_1(q)$ (see (29) for a proof of this fact) we obtain

$$2\Lambda_\pi(p) = p\left(1 - \int_{(p,1]} \frac{1}{q}\,\mathrm{d}\pi_1(q)\right) - \int_0^p \mathrm{d}\pi_1(q).$$

A change of variable from posterior beliefs to log-likelihood ratios, letting $a = \log\frac{p}{1-p}$, implies

$$2\Lambda_\pi(p) = \frac{\mathrm{e}^a}{1+\mathrm{e}^a}\left(1 - \int_{(a,\infty)} \frac{1+\mathrm{e}^u}{\mathrm{e}^u}\,\mathrm{d}F_1(u)\right) - F_1(a). \tag{9}$$

Since

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

(9) leads to

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

where the final equality follows from (6). It then follows that P dominates Q if and only if $\tilde{F}(a) \leq \tilde{G}(a)$ for all $a \in \mathbb{R}$, as desired. □

Intuitively, transferring probability mass from lower to higher values of $\log(\mathrm{d}P_\theta/\mathrm{d}P_{1-\theta})$ leads to an experiment that, conditional on the state being θ, is more likely to shift the decision maker's beliefs towards the correct state. Hence, one might conjecture that Blackwell dominance of the experiments P and Q is related to stochastic dominance of the distributions F_θ and G_θ. However, since the likelihood-ratio $\mathrm{d}P_1/\mathrm{d}P_0$ must satisfy the change of measure identity $\int \frac{\mathrm{d}P_0}{\mathrm{d}P_1}\,\mathrm{d}P_1 = 1$, the distribution F_1 must satisfy

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

Because the function e^{-u} is strictly decreasing and convex, and the same identity must hold for G_1, it is impossible for F_1 to stochastically dominate G_1. Theorem 2 shows that a more useful comparison is between the perfected log-likelihood ratios.[5]

The next lemma simplifies the study of the 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 $\tilde{F}$ and $\tilde{G}$ be the distributions of the perfected log-likelihood ratios. For $b \in \mathbb{R}$, the following hold:*

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

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

[5]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).

4.3 Large Deviations and the Rényi Order

We now illustrate how dominance in the Rényi order translates into properties of the log-likelihood ratios, and provide a sketch of the proof of Theorem 1. In what follows, given a bounded random variable X we denote by $\max[X] = \min\{a : \mathbb{P}[X \leq a] = 1\}$ the essential maximum of X; this is the maximum of the support of its distribution. We denote by $M_X(t) = \mathbb{E}\left[\mathrm{e}^{tX}\right]$ its *moment generating function.*

This proof will use as a crucial ingredient the following uniform large deviation result. A similar statement is proved in Aubrun and Nechita (2008, Lemma 2); we provide an independent proof in the appendix, which additionally provides explicit estimates for the threshold N described below.

Proposition 3. *Let X and Y be bounded random variables that satisfy*

(i). $\max[X] \neq \max[Y]$,

(ii). $\mathbb{E}[X] > \mathbb{E}[Y]$,

(iii). *and $M_X(t) > M_Y(t)$ for all $t > 0$.*

Then there exists N such that for all $n \geq N$ and $a \geq \mathbb{E}[Y]$,

$$\mathbb{P}[X_1 + \cdots + X_n \geq na] \geq \mathbb{P}[Y_1 + \cdots + Y_n \geq na]. \tag{10}$$

The result is based on the following intuition. The fact that $\mathbb{E}[X] > \mathbb{E}[Y]$ guarantees that the dominance condition

$$\mathbb{P}[X_1 + \cdots + X_n \geq na] \geq \mathbb{P}[Y_1 + \cdots + Y_n \geq na]. \tag{11}$$

holds with respect to all values of a that lie between $\mathbb{E}[X]$ and $\mathbb{E}[Y]$. This is established by applying the Berry-Esseen Theorem, a uniform version of the Central Limit Theorem. The main step in the proof Proposition 3 uses large-deviations techniques to obtain lower and upper bounds on the probabilities of the events $\{X_1 + \cdots + X_n \geq na\}$ and $\{Y_1 + \cdots + Y_n \geq na\}$ for the remaining values of a.

Large deviation 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 the probability of the event $\{X_1 + \cdots + X_n \geq na\}$ is low for $a > \mathbb{E}[X]$ and large n. A crucial insight due to Cramér (1938) is that the order of magnitude of the probability of this event is determined by the behavior of the moment generating function $M_X(t)$.

The key difficulty is in obtaining bounds that allow, for a fixed n, a comparison between the two probabilities in (11) over a whole interval of values for a. This requires a careful application of uniform large deviation theorems due to Bahadur and Rao (1960) and Petrov (1965).

Proposition 3 has the following implications for the study of experiments. Consider two bounded, generic experiments $P = (\Omega, P_0, P_1)$ and $Q = (\Xi, Q_0, Q_1)$, with the property that P dominates Q in the Rényi order, and let

$$X = \log \frac{\mathrm{d}P_1}{\mathrm{d}P_0} \quad \text{and} \quad Y = \log \frac{\mathrm{d}Q_1}{\mathrm{d}Q_0}$$

be the corresponding log-likelihood ratios. We are interested in their properties conditional on $\theta = 1$, and so treat X as a random variable defined on the probability space (Ω, P_1), and Y as defined on (Ξ, Q_1), so that their distributions are F_1 and G_1, respectively.

The variables X and Y satisfy all the conditions of Proposition 3. Indeed, it follows immediately from its definition that the Rényi divergence is formally related to the moment generating functions of the log-likelihood ratios. For any $t > 0$,

$$R_P^1(t) = \frac{1}{t-1} \log M_X(t-1) \quad \text{and} \quad R_Q^1(t) = \frac{1}{t-1} \log M_Y(t-1). \tag{12}$$

In addition, $R_P^1(0) = \mathbb{E}[X]$ and $R_Q^1(0) = \mathbb{E}[Y]$. Finally, since the two experiments are generic, the maxima of X and Y are different.

Hence, letting (X_i) and (Y_i) be i.i.d. copies of X and Y, Proposition 3 implies the existence of a large enough N such that for for all $n \geq N$ and $a \geq \mathbb{E}[Y]$,

$$\mathbb{P}[X_1 + \cdots + X_n \geq na] \geq \mathbb{P}[Y_1 + \cdots + Y_n \geq na]. \tag{13}$$

As discussed earlier, $X_1 + \cdots + X_n$ has the same distribution as the log-likelihood ratio $\log(\mathrm{d}P_1^n/\mathrm{d}P_0^n)$ of the repeated experiment. It follows, therefore, that the distribution F_1^{*n} of $\log(\mathrm{d}P_1^n/\mathrm{d}P_0^n)$ and the distribution G_1^{*n} of $\log(\mathrm{d}Q_1^n/\mathrm{d}Q_0^n)$ satisfy

$$F_1^{*n}(na) \leq G_1^{*n}(na) \quad \text{for all} \quad a \geq \mathbb{E}[Y] = D(Q_1 \| Q_0).$$

In turn, Lemma 1 implies that the distributions of the perfected log-likelihood ratios of the two repeated experiments are ranked for all values above $nD(Q_1\|Q_0)$. Note that in applying Proposition 3, we have only used the assumption that $M_X(t) > M_Y(t)$ for $t \geq 0$, or equivalently $R_P^1(t) > R_Q^1(t)$ for $t \geq 1$.

We show the same ranking holds for the range $[-nD(Q_0\|Q_1), nD(Q_1\|Q_0)]$. This follows from the assumption that $R_P^1(t) > R_Q^1(t)$ for $t \in (0,1)$. Finally, to compare the perfected log-likelihood ratios at values below $-nD(Q_0\|Q_1)$, we apply Proposition 3 as above, but to the opposite pair of log-likelihood ratios $\log(\mathrm{d}P_0/\mathrm{d}P_1)$ and $\log(\mathrm{d}Q_0/\mathrm{d}Q_1)$.

5 Stochastic Dominance in Large Numbers

In this section we study and characterize notions of stochastic dominance for sums of i.i.d. random variables. The results are a natural application of the methods developed to study repeated experiments. They provide robust criteria for comparing prospects in decision problems that involve multiple independent risks, as in the case of a physician treating multiple patients, or of an investor managing a large portfolio of assets.

The study of aggregate risks has a long history in economics. Samuelson (1963) asked under what conditions an agent would reject a gamble, but accept n independent copies of it. He deemed inconsistent the behavior of a decision maker who is willing to accept n copies of a lottery but not one, and attributed this choice reversal to a naive interpretation of the law of large numbers. The critical point in Samuelson's argument is that the law of large numbers does not provide information about the odds of rare but large deviations, and it is therefore insufficient as a guide for action. The paper yielded a large literature studying retirement decisions and insurance strategies (see, e.g., Pratt and Zeckhauser, 1987; Kimball, 1993; Gollier, 1996; Benartzi and Thaler, 1999), as well as in behavioral economics (Rabin and Thaler, 2001), and finance (Ross, 1999).

In contrast to the existing literature, our study of repeated gambles makes no assumptions over the decision maker's preferences, beyond monotonicity with respect to stochastic dominance.

5.1 Definition

Recall that a random variable X dominates Y in k^{th}-order stochastic dominance, denoted by $X \geq_k Y$, if $\mathbb{E}[\phi(X)] \geq \mathbb{E}[\phi(Y)]$ for every bounded and k-fold differentiable function ϕ that is increasing and whose first k derivatives alternate in sign. That is, all functions ϕ that satisfy $(-1)^n \phi^{(n)} \leq 0$ for all $n \leq k$.

Definition 2. Let X and Y be random variables, and let (X_i) and (Y_i) be i.i.d. copies of X and Y, respectively. The random variable X is said to k^{th}-order dominate Y *in large numbers* if for all n large enough

$$X_1 + \cdots + X_n \geq_k Y_1 + \cdots + Y_n. \tag{14}$$

It is well-known that stochastic dominance in large numbers is implied by stochastic dominance. In fact if $X \geq_k Y$ then $\sum_{i=1}^n X_i \geq_k \sum_{i=1}^n Y_i$ for any number n of i.i.d. replicas of the two random variable (a proof is provided in Lemma 6 in the appendix). The converse implication, however, is not true.

As a simple example, let X be a lottery that pays 1 or 0 with probability $1/2$, and let Y be distributed uniformly over $[-1/5, 4/5]$. For instance, X might correspond to an Arrow-Debreu security, while Y might correspond to an insurance contract that costs $1/5$ and offers a smoothed distribution of payoff that is uniform on the unit interval. The cumulative distribution functions of X and Y are depicted in Figure 2, from which it is clear that neither first-order dominates the other. In fact, the two distributions are not ranked in terms of second-order stochastic dominance either. To see this, note that Y has higher expected utility than X for the utility function given by $u(x) = x$ for $x \leq 1/5$ and $u(x) = 1/5$ otherwise. It is also clear that Y does not dominate X, since the latter has higher expectation.

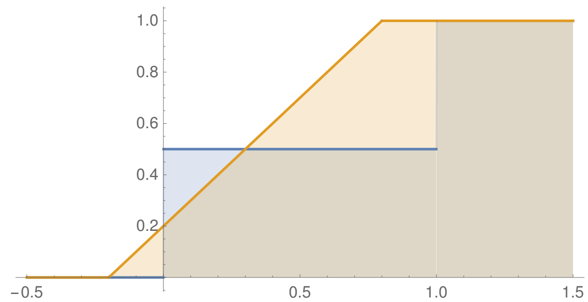

Figure 2: The CDFs of X and Y, in blue and orange, respectively.

Theorem 3 below will show that X first-order dominates Y in large numbers. In this example, it is not difficult to verify that replicating the two gambles makes it possible to rank them in terms of stochastic dominance: Figure 3 shows the cumulative distribution functions of the two sums $X_1 + \cdots + X_n$ and $Y_1 + \cdots + Y_n$ when setting $n = 35$, from which it is apparent that the first sum dominates the second one in terms of first-order stochastic dominance.

5.2 Characterizations

In this section we provide necessary and sufficient conditions for stochastic dominance in large numbers. To each bounded random variable X we associate the function $L_X : \mathbb{R} \to \mathbb{R}$ defined as

$$L_X(t) = \frac{1}{t} \log \mathbb{E}\left[e^{tX}\right] \tag{15}$$

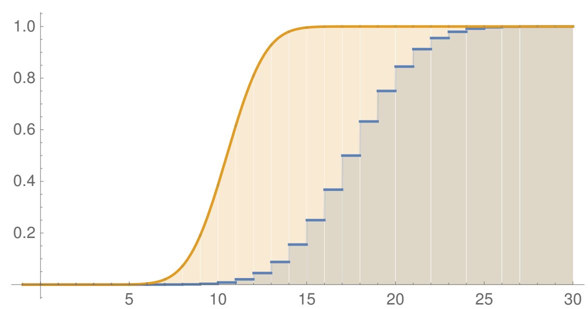

Figure 3: The CDFs of $X_1 + \cdots + X_n$ and $Y_1 + \cdots + Y_n$, for $n = 35$, in blue and orange, respectively.

for all $t \neq 0$, and, to guarantee continuity,

$$L_X(0) = \mathbb{E}[X]. \tag{16}$$

If X is a gamble, then $L_X(t)$ is the certainty equivalent that a decision maker ascribes to X, under expected utility and a utility function u whose coefficient of absolute risk aversion is constant and equal to $-t$.[6] Note that for t positive, such a decision maker is in fact risk-loving; we include these agents for the analysis of *first-order* stochastic dominance.

The quantity L_X is a standard tool in the theory of choice under risk, finance, probability theory, and other fields. Because it amounts to a simple normalization of the moment generating function of X, the certainty equivalent L_X is known or can be easily computed for most families of distributions commonly used in applications.

We similarly impose a mild genericity assumption: Say that a pair (X, Y) is *generic* if $\min[X] \neq \min[Y]$ and $\max[X] \neq \max[Y]$. The next result characterizes first-order dominance in large numbers:

Theorem 3. *Let X and Y be a generic pair of bounded random variables. Then the following are equivalent:*

(i). $L_X(t) > L_Y(t)$ *for all* $t \in \mathbb{R}$,

(ii). *X first-order dominates Y in large numbers.*

The result is an immediate corollary of Proposition 3 introduced in the proof of Theorem 1; see also Aubrun and Nechita (2008, Lemma 2). Figure 4 depicts the certainty equivalents L_X and L_Y for the two gambles introduced in our earlier example. As shown in the figure and can be verified analytically, the certainty equivalent of X lies above that of Y.

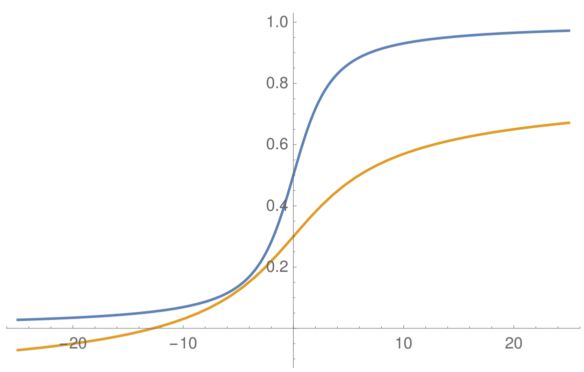

Figure 4: L_X and L_Y in the example of §5.1, in blue and orange, respectively.

[6] Up to an affine transformation, u is of the form $u(x) = e^{tx}$ for t positive, $u(x) = -e^{tx}$ for t negative and $u(x) = x$ for $t = 0$.

We now turn our attention to preferences that display risk aversion. The next theorem characterizes higher-order dominance in large numbers:

Theorem 4. *Let X and Y be a generic pair of bounded random variables such that $\mathbb{E}[X] \neq \mathbb{E}[Y]$. Then the following are equivalent:*

(i). $L_X(t) > L_Y(t)$ *for all* $t \leq 0$.

(ii). X *second-order dominates* Y *in large numbers.*

(iii). X k^{th}*-order dominates* Y *in large numbers, for some* $k \geq 2$.

(iv). $X \geq_k Y$, *for some* $k \geq 2$.

Theorem 4 establishes a sharp equivalence between stochastic dominance in large numbers and an elementary and well-known class of preferences. The correspondence of (i) and (ii) shows that whenever all risk averse CARA agents unanimously prefer X over Y, then, for a large enough number of repetitions, *all* agents with monotone risk-averse preferences will agree on this ranking. Moreover, this condition is both sufficient and necessary.

The equivalence of (ii) and (iii) shows that there is no difference between second and higher order risk attitudes when it comes to large numbers: For every $k \geq 2$, X k^{th}-order dominates Y in large numbers if and only if it second-order dominates it in large numbers. This fact might appear surprising. Higher-order risk attitudes describe increasingly nuanced properties of a decision maker's preferences. Prudence (Kimball, 1990), i.e. the requirement that the third derivative of the decision maker's utility function is positive, and temperance (Kimball, 1991), i.e. the requirement that its fourth derivative is negative, are known to have strong implications for comparative statics in decision problems under risk, including precautionary saving problems and decisions under background risk (Gollier, 2004). Theorem 4 shows that when considering a sum of a sufficiently large number of i.i.d. gambles, the distinction between risk aversion and higher-order risk attitudes collapses.

Finally, the equivalence between (iv) and (ii) establishes a novel characterization of higher-order risk dominance based on the study of compound gambles: for every $k \geq 2$, k-th order stochastic dominance implies second-order stochastic dominance in large numbers, which in turn implies k-th order stochastic dominance for some k.

Next, we recall that Y is a *mean preserving spread* of X if both have the same expectation and X second-order stochastically dominates Y. We define the notion of a mean preserving spread in large numbers analogously to Definition 2, so that for equal mean random variables it coincides with second-order stochastic dominance in large numbers.

Theorem 5. *Let X and Y be a generic pair of bounded random variables such that $\mathbb{E}[X] = \mathbb{E}[Y]$. Then the following are equivalent:*

(i). $\mathrm{Var}(X) < \mathrm{Var}(Y)$, *$L_X(t) > L_Y(t)$ for all $t < 0$, and $L_X(t) < L_Y(t)$ for all $t > 0$.*

(ii). *Y is a mean preserving spread of X in large numbers.*

Hence, when X and Y have the same expected value, X second-order dominates Y in large numbers if and only if X has lower variance and is preferred to Y by any risk-averse CARA agent, while Y is preferred to X by all CARA agents who are risk-loving.

One may wonder about the difference between condition (i) here and condition (i) in Theorem 3. Note that first-order dominance in large numbers is equivalent to $L_X(t) > L_Y(t)$ for all t, whereas Theorem 5 requires $L_X(t)$ to be *smaller* for $t > 0$. There is however no inconsistency, because the assumption $\mathbb{E}[X] = \mathbb{E}[Y]$ in Theorem 5 already rules out the possibility that $X_1 + \cdots + X_n$ can first-order dominate $Y_1 + \cdots + Y_n$. Furthermore, in order for $X_1 + \cdots + X_n$ to second-order dominate $Y_1 + \cdots + Y_n$, the former sum must be a mean-preserving contraction of the latter. This suggests that the right-tail of $X_1 + \cdots + X_n$ should be *less* spread-out, as captured by $L_X(t) < L_Y(t)$ for $t > 0$, unlike in the case of first-order stochastic dominance.

We conclude by observing that stochastic dominance in large numbers can be naturally extended to compare compound i.i.d. returns. Two random returns X and Y can be ranked by requiring that for every n large enough their compounded i.i.d. returns satisfy

$$X_1 \times \cdots \times X_n \geq_k Y_1 \times \cdots \times Y_n. \tag{17}$$

The resulting stochastic order amounts to stochastic dominance in large numbers applied to $\log(X)$ and $\log(Y)$, and is characterized in terms of the certainty equivalents induced by all preferences that display constant *relative* risk aversion.

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. In the statistics literature, Torgersen (1970) provides an early example of two experiments that are not comparable in the Blackwell order, but are comparable in large samples.

Moscarini and Smith (2002) produce 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 such that the payoff achievable from $P^{\otimes n}$ is higher than that from $Q^{\otimes n}$ whenever $n \geq N$. This order is characterized by the *efficiency index* of an experiment, defined, in our notation, as the minimum over $t \in (0,1)$ of the function $(t-1)R_P^0(t)$. While in Moscarini and Smith (2002) the number n of repetitions is allowed to depend on the decision problem, dominance in large samples is a criterion for comparing experiments uniformly over decision problems, and thus is conceptually closer to Blackwell dominance.[7]

Azrieli (2014) shows that the Moscarini-Smith order is a strict refinement of dominance in large samples. Perhaps surprisingly, this conclusion is reversed under a modification of their definition: When extended to consider all decision problems, including problems with infinitely many actions, the Moscarini-Smith order over experiments coincides with dominance in large samples.[8]

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.[9]

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

Motivated by questions in quantum information, Jensen (2019) asks the following question: Given two finitely supported distributions μ and ν, when does the n-fold product $\mu^{\times n} = \mu \times \cdots \times \mu$ majorize $\nu^{\times n}$ for all large n? He shows that for the case that μ and ν have *different* support sizes, the answer is given by the ranking of their Rényi entropies.[10]

[7] Recent work by Hellman and Lehrer (2019) generalizes the Moscarini-Smith order to Markov (rather than i.i.d.) sequences of experiments. An interesting question for future work is whether dominance in large samples admits a similar generalization.

[8] Consider the following extension of the Moscarini-Smith order: say that P dominates Q if for every decision problem (with possibly infinitely many actions) there exists an N such that the expected utility achievable from $P^{\otimes n}$ is higher than that from $Q^{\otimes n}$ whenever $n \geq N$. Each Rényi divergence $R_P^\theta(t)$ corresponds to the indirect utility defined by a decision problem (see the proof of Theorem 1 in the appendix), and for such decision problems the ranking over repeated experiments is independent of the sample size n. We deduce that 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.

[9] To see this, consider the example in §3.2 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.

[10] As discussed above, majorization with different support sizes does not imply Blackwell dominance.

For the case of equal support size, our Theorem 1 implies a similar result, which Jensen (2019, Remark 3.9) conjectures to be true. We prove his conjecture in the appendix.

Fritz (2018) uses an abstract algebraic approach to prove a result that is complementary to Theorem 3. 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. As we mentioned above, a statement similar to Theorem 3 is implied by the proof of Lemma 2 in Aubrun and Nechita (2008), also in the context of majorization and quantum information theory.

Stochastic Orders. Müller and Stoyan (2002) and Shaked and Shanthikumar (2007) are comprehensive sources on stochastic orders. The ordering generated by the functionals of the form $L_X(t)$ for $t > 0$, is known in the literature as the *Laplace Transform Order*, and studied in Reuter and Riedrich (1981), Fishburn (1980), Alzaid et al. (1991) and Caballé and Pomansky (1996), among others.

Hart (2011) proposes two complete stochastic orders that refine second-order stochastic dominance: *wealth-uniform dominance* and *utility-uniform dominance*. He further shows that dominance in these orders is characterized by having a smaller riskiness index/measure given in Aumann and Serrano (2008) and Foster and Hart (2009), respectively. But since these measures of risk are distinct, an open question left by Hart (2011) is whether the two stochastic orders agree on interesting cases beyond second-order stochastic dominance. In §K, we show that the two uniform dominance orders both refine second-order dominance in large numbers.

Experiments for Many States. 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. A natural conjecture is that the ranking of the multidimensional moment generating function of the log-likelihood ratio—which translates to Rényi divergences in the two state case—characterizes this order for any number of states. Unfortunately, our proof technique does not straightforwardly extend to this general case. In particular, we do not know how to extend the reduction of the Blackwell order to first-order stochastic dominance (Theorem 2). The technical difficulty that arises when studying the Blackwell order for more than two states is not new to the literature. As Jewitt (2007) writes, "the problem is the need to deal with a multivariate stochastic dominance relation for a class of functions (convex) for which the set of extremal rays is too complex to be of service."

Indeed, the ranking based on Rényi entropies is distinct from our ranking based on Rényi divergences unless the support sizes are equal. See §L in the appendix for details.

Appendix

A Uniform Large Deviations

We begin by reviewing some standard concepts from large deviations theory. For every bounded random variable X we define $\rho_X : \mathbb{R} \to \mathbb{R}_+$ as

$$\rho_X(a) = \inf_{t \in \mathbb{R}} \mathrm{e}^{-at} M_X(t).$$

where $M_X(t) = \mathbb{E}\left[\mathrm{e}^{tX}\right]$ is the moment generating function of X. We note that $\mathrm{e}^{-at} M_X(t) = M_{X-a}(t)$, hence $\rho_X(a)$ is the infimum of the moment generating function of $X - a$.

We call a random variable *non-degenerate* if its distribution is not a point mass. In this case, as is well known, M_X is strictly log-convex, and if $\min[X] < a < \max[X]$ then $M_{X-a}(t) \to \infty$ as $|t| \to \infty$. It follows that for every a in the range $\min[X] < a < \max[X]$ the minimization problem in the definition of ρ_X, which is equivalent to minimizing the strictly convex function $-at + \log M_X(t)$, has a unique solution. We denote this minimizer by

$$t_X(a) = \operatorname*{argmin}_{t \in \mathbb{R}} \mathrm{e}^{-at} M_X(t).$$

Let $K_X(t) = \log M_X(t)$ denote the "cumulant generating function" of X. The first-order condition gives that $t_X(a)$ solves

$$K_X'(t(a)) = \frac{M_X'(t(a))}{M_X(t(a))} = a.$$

Note that $M_X(0) = 1$ and $M_X'(0) = \mathbb{E}[X]$. So $K_X'(0) = \mathbb{E}[X]$. This, together with the convexity of K_X, shows that $t(a) \geq 0$ if and only if $a \geq \mathbb{E}[X]$.

Finally, for every $\min[X] < a < \max[X]$ we define

$$\sigma_X(a) = \sqrt{\frac{M_X''(t(a))}{M_X(t(a))} - a^2}.$$

Using the above formula for $t(a)$, we also have $\sigma_X(a) = \sqrt{K_X''(t(a))}$ which is strictly positive whenever X is non-degenerate.

We will refer to quantities above as simply $\rho(\cdot)$, $t(\cdot)$ and $\sigma(\cdot)$ whenever X is unambiguously explicit from the context. The following technical lemma relates these functions for a random variable X to the corresponding functions for its negative $-X$; it will allow us to focus on large deviations "on one side" (of the expected value) and quickly deduce analogous results for the other side.

Lemma 2. *Let X be a bounded and non-degenerate random variable. Then $\rho_{-X}(a) = \rho_X(-a)$ for every a. If in addition $\min[X] < a < \max[X]$ then $t_{-X}(a) = -t_X(-a)$ and $\sigma_{-X}(a) = \sigma_X(-a)$.*

Proof. Notice that $M_X(t) = M_{-X}(-t)$. Hence, given a, we have that for every t, $\mathrm{e}^{-at}M_{-X}(t) = \mathrm{e}^{a\cdot(-t)}M_X(-t)$. It follows from this that $\rho_{-X}(a) = \rho_X(-a)$ and $t_{-X}(a) = -t_X(-a)$. $\sigma_{-X}(a) = \sigma_X(-a)$ then follows from the definition. □

The main technical tool of this paper is the following lemma, due, in various forms, to (Bahadur and Rao, 1960, Lemma 2) and to (Petrov, 1965, Theorems 5 and 6). It is a sharp, quantitative large deviations estimate, which will be useful not only for proving our asymptotic results above, but can also be used for estimating the number n of repetitions required to achieve stochastic dominance.

Lemma 3. *Let X be a bounded and non-degenerate random variable and let $b > 0$ satisfy $\mathbb{P}[|X| \leq b/2] = 1$. Let $X_1, X_2, \dots$ be i.i.d. copies of X.*

Then for every $\mathbb{E}[X] \leq a < \max[X]$ and every n, it holds that

$$\mathbb{P}[X_1 + \cdots + X_n \geq a \cdot n] \leq \rho(a)^n. \tag{18}$$

And for every $\mathbb{E}[X] \leq a < \max[X]$ and $n \geq (10b/\sigma(a))^2$ it holds that

$$\mathbb{P}[X_1 + \cdots + X_n \geq a \cdot n] \geq C(a) \cdot \frac{\rho(a)^n}{\sqrt{n}} \tag{19}$$

where

$$C(a) = \frac{\mathrm{e}^{-10t(a)b} \cdot b}{\sigma(a)}.$$

Inequalities similar to (18) and (19) apply to values of a that lie below the expectation of X. Consider the case where $\min[X] < a \leq \mathbb{E}[X]$. Then, by applying the inequality (19) to the random variable $-X$ and using Lemma 2, we obtain that for every $n \geq (10b/\sigma_X(a))^2$,

$$\begin{aligned}\rho_X(a)^n = \rho_{-X}(-a)^n &\geq \mathbb{P}[-X_1 - \cdots - X_n \geq -a \cdot n] = \mathbb{P}[X_1 + \cdots + X_n \leq a \cdot n] \\ &\geq \frac{\mathrm{e}^{-10t_{-X}(-a)b} \cdot b}{\sigma_{-X}(-a)} \cdot \frac{\rho_{-X}(-a)^n}{\sqrt{n}} = \frac{\mathrm{e}^{10t_X(a)b} \cdot b}{\sigma_X(a)} \cdot \frac{\rho_X(a)^n}{\sqrt{n}}.\end{aligned} \tag{20}$$

A corollary of this lemma is a lower estimate that is *uniform* over $a \in [\mathbb{E}[X], \max[X] - \varepsilon]$.

Corollary 1. *In the setting of Lemma 3, let $A = [\underline{a}, \overline{a}] \subset [\mathbb{E}[X], \max[X])$ be a given interval. Then*

$$C_A = \inf_{a \in A} C(a) \quad \text{and} \quad n_A = \sup_{a \in A} (10b/\sigma(a))^2$$

are positive and finite, and hence for every $a \in A$ and every $n \geq n_A$

$$\mathbb{P}[X_1 + \cdots + X_n \geq a \cdot n] \geq C_A \cdot \frac{\rho(a)^n}{\sqrt{n}}. \tag{21}$$

Proof. Since $t(a)$ solves $K_X'(t(a)) = a$ and K_X is strictly convex, $t(a)$ must be strictly increasing in a. It is thus continuous and bounded above on the compact set A. Similarly $\sigma(a)$ is continuous and strictly positive, so it is bounded above and away from zero on A. Thus $C_A > 0$ and $n_A < \infty$. $\square$

The next lemma is a refined version of Lemma 3, applicable to the regime of a that vanishes with n.

Lemma 4. *In the setting of Lemma 3, for every $\mathbb{E}[X] \leq a < \max[X]$ and every n it holds that*

$$\mathbb{P}[X_1 + \cdots + X_n \geq an] \leq \frac{1 + \sqrt{2\pi} \cdot t(a)b}{\sqrt{2\pi} \cdot \sigma(a)t(a)} \cdot \frac{\rho(a)^n}{\sqrt{n}}.$$

And for every $\mathbb{E}[X] \leq a < \max[X]$ and $n \geq 2[\sigma(a)t(a)]^{-2}$ it holds that

$$\mathbb{P}[X_1 + \cdots + X_n \geq an] \geq \frac{1 - 2\sqrt{2\pi} \cdot t(a)b}{2\sqrt{2\pi} \cdot \sigma(a)t(a)} \cdot \frac{\rho(a)^n}{\sqrt{n}}.$$

This, and the previous lemma 3, are proved in the rest of this section.

A.1 Proof of Lemma 3

We follow Bahadur and Rao (1960). For each a such that $\mathbb{E}[X] \leq a < \max[X]$, denote

$$p_n(a) = \mathbb{P}[X_1 + \cdots + X_n \geq an].$$

Let $Y^a = X - a$ and let F_a be its cumulative distribution function. Consider, in addition, a random variable Z^a whose c.d.f. is given by

$$G(z) = \frac{1}{\rho(a)} \cdot \int_{-\infty}^{z} \mathrm{e}^{t(a)\cdot y} \, \mathrm{d}F_a(y).$$

Note that $G(\infty) = 1$ because by definition $M_{Y^a}(t(a)) = \rho(a)$.

More generally, the moment generating function of Z^a is given by

$$M_{Z^a}(r) = \frac{M_{Y^a}(r + t(a))}{\rho(a)} = \frac{M_{Y^a}(r + t(a))}{M_{Y^a}(t(a))}$$

It follows from $M_{Y^a}'(t(a)) = 0$ that $M_{Z^a}'(0) = 0$, hence Z^a has mean 0. Moreover

$$\sigma(a)^2 = M_{Y^a}''(t(a)) = M_{Z^a}''(0) = \mathrm{Var}(Z^a).$$

It is clear that Z^a has the same support as Y^a, which, for the entire range of values of a we consider, is contained in $[-b, b]$. Thus we further have

$$\mathbb{E}\left[|Z^a|^3\right] \leq b \cdot \mathbb{E}\left[(Z^a)^2\right] = b \cdot \sigma(a)^2.$$

Let $Z_1^a, \ldots, Z_n^a$ be i.i.d. copies of Z^a, and define

$$U_n^a = \frac{Z_1^a + \cdots + Z_n^a}{\sqrt{n} \cdot \sigma(a)}.$$

Denote by $H_n^a(z) = \mathbb{P}\left[U_n^a \le z\right]$ the c.d.f. of U_n^a. Then we can apply Lemma 2 in Bahadur and Rao (1960) to obtain[11]

$$p_n(a) = \rho(a)^n \cdot \sqrt{n}\sigma(a)t(a) \cdot \int_0^\infty \mathrm{e}^{-\sqrt{n}\sigma(a)t(a)z} \cdot \left(H_n^a(z) - H_n^a(0)\right) \mathrm{d}z.$$

Clearly, $H_n^a(z) - H_n^a(0) \le 1$ for each z. So $p_n(a) \le \rho(a)^n$, which yields (18), also known as the *Chernoff bound*.

In the other direction, for any $z_0 > 0$ we have

$$\begin{aligned} p_n(a) &\ge \rho(a)^n \cdot \sqrt{n}\sigma(a)t(a) \cdot \int_{z_0}^\infty \mathrm{e}^{-\sqrt{n}\sigma(a)t(a)z} \cdot \left(H_n^a(z_0) - H_n^a(0)\right) \mathrm{d}z \\ &= \rho(a)^n \cdot \mathrm{e}^{-\sqrt{n}\sigma(a)t(a)z_0} \cdot \left(H_n^a(z_0) - H_n^a(0)\right). \end{aligned} \tag{22}$$

By the Berry-Esseen Theorem[12]

$$H_n^a(z_0) - H_n^a(0) \ge \int_0^{z_0} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \,\mathrm{d}x - \frac{\mathbb{E}\left[|Z^a|^3\right]}{\sigma(a)^3\sqrt{n}} \ge \int_0^{z_0} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \,\mathrm{d}x - \frac{b}{\sigma(a)\sqrt{n}}.$$

Note that if $z_0 \le 1$ then the first term on the right hand side is at least $z_0/5$. Hence, if we pick $z_0 = 10b/(\sigma(a)\sqrt{n})$, and let $n_0 = (10b/\sigma(a))^2$, then for all $n \ge n_0$ we have that $z_0 \le 1$ and so the above yields $H_n^a(z_0) - H_n^a(0) \ge b/(\sigma(a)\sqrt{n})$. Hence from (22) it holds for all $n \ge n_0$ that

$$\begin{aligned} p_n(a) &\ge \rho(a)^n \cdot \mathrm{e}^{-\sqrt{n}\sigma(a)t(a)z_0} \cdot \left(H_n^a(z_0) - H_n^a(0)\right) \\ &\ge \rho(a)^n \cdot \mathrm{e}^{-10t(a)b} \cdot \frac{b}{\sigma(a)\sqrt{n}}, \end{aligned}$$

which shows (19).

A.2 Proof of Lemma 4

We initially proceed as in the proof of Lemma 3, arriving at

$$p_n(a) = \rho(a)^n \cdot \sqrt{n}\sigma(a)t(a) \cdot \int_0^\infty \mathrm{e}^{-\sqrt{n}\sigma(a)t(a)z} \cdot \left(H_n^a(z) - H_n^a(0)\right) \mathrm{d}z.$$

Let Φ denote the cdf of a standard Gaussian distribution. By the Berry-Esseen Theorem

$$H_n^a(z) - H_n^a(0) \le \Phi(z) - \Phi(0) + \frac{b}{\sigma(a)\sqrt{n}}.$$

[11] The lemma follows from the definitions and integration by parts. We do not repeat the details.

[12] In fact, to obtain a simpler expression we use some recent improvements in the estimate of the constant in the Berry-Esseen Theorem by Tyurin (2010).

Hence

$$p_n(a) \le \rho(a)^n \cdot \sqrt{n}\sigma(a)t(a) \cdot \int_0^\infty \mathrm{e}^{-\sqrt{n}\sigma(a)t(a)z} \cdot \left(\Phi(z) - \Phi(0) + \frac{b}{\sigma(a)\sqrt{n}}\right) \mathrm{d}z.$$

Let $c = \sqrt{n}\sigma(a)t(a)$. Then integration by parts implies

$$c\int_0^\infty \mathrm{e}^{-cz} \cdot (\Phi(z) - \Phi(0)) \mathrm{d}z = \mathrm{e}^{c^2/2} \cdot \Phi(-c) \tag{23}$$

Standard bounds for Φ assert that

$$\frac{1}{c\sqrt{2\pi}}\left(1 - \frac{1}{c^2}\right) \le \mathrm{e}^{c^2/2} \cdot \Phi(-c) \le \frac{1}{c\sqrt{2\pi}}. \tag{24}$$

We thus obtain from the upper bound and (23) that

$$\begin{aligned} p_n(a) &\le \rho(a)^n \left(\frac{1}{\sqrt{2\pi}\sqrt{n}\sigma(a)t(a)} + \frac{b}{\sigma(a)\sqrt{n}}\right) \\ &= \rho(a)^n \frac{1}{\sqrt{2\pi}\sqrt{n}\sigma(a)t(a)}\left(1 + \sqrt{2\pi}t(a)b\right). \end{aligned}$$

In the other direction, applying Berry-Esseen again, we have

$$H_n^a(z) - H_n^a(0) \ge \Phi(z) - \Phi(0) - \frac{b}{\sigma(a)\sqrt{n}}.$$

For $n \ge 2[\sigma(a)t(a)]^{-2}$, we have $c \ge \sqrt{2}$, and so the lower bound in (24) implies

$$\mathrm{e}^{c^2/2}\Phi(-c) \ge \frac{1}{2\sqrt{2\pi}c}.$$

It follows from this estimate and (23) that

$$\begin{aligned} p_n(a) &\ge \rho(a)^n \left(\frac{1}{2\sqrt{2\pi}\sqrt{n}\sigma(a)t(a)} - \frac{b}{\sigma(a)\sqrt{n}}\right) \\ &= \rho(a)^n \frac{1}{2\sqrt{2\pi}\sqrt{n}\sigma(a)t(a)}\left(1 - 2\sqrt{2\pi}t(a)b\right). \end{aligned}$$

B Proof of Proposition 3 and Theorem 3

It is not difficult to see that Proposition 3 implies Theorem 3. Indeed, to prove Theorem 3 we just need to show one direction, that $L_X(t) > L_Y(t)$ for all t (recall the definition of L_X in (15) and (16)) implies $X_1 + \cdots + X_n$ dominates $Y_1 + \cdots + Y_n$ for large n. By Proposition 3,

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

for every $a \geq \mathbb{E}[Y]$ and $n \geq N$. Moreover, $L_X(t) > L_Y(t)$ for $t \leq 0$ implies that $L_{-Y}(t) > L_{-X}(t)$ for $t \geq 0$. Thus, applying Proposition 3 to the pair $-Y$ and $-X$, we obtain

$$\mathbb{P}[-Y_1 - \cdots - Y_n \geq n\tilde{a}] \geq \mathbb{P}[-X_1 - \cdots - X_n \geq n\tilde{a}]$$

for every $\tilde{a} \geq \mathbb{E}[-X]$ and $n \geq N$. Setting $a = -\tilde{a}$, this is equivalent to

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

for every $a \leq \mathbb{E}[X]$. Thus the inequality holds for all a when n is sufficiently large, and Theorem 3 holds.

To prove Proposition 3, let b be a positive number so that X and Y are supported on $[-b/2, b/2]$. Without loss of generality we assume X and Y are non-degenerate.[13] Moreover, since $L_X(t) > L_Y(t)$ for all $t \geq 0$ (as this is equivalent to conditions (ii) and (iii)), letting $t \to \infty$ yields $\max[X] \geq \max[Y]$. Since they are unequal by assumption, we in fact have $\max[X] > \max[Y]$.

Denote by F^{*n} (respectively G^{*n}) the c.d.f. of the sum of n i.i.d. copies of X (respectively Y). We need to show $1 - F^{*n}(na) \geq 1 - G^{*n}(na)$ for $a \geq \mathbb{E}[Y]$ and n large. We divide the proof into cases.

Case 1: $a > \max[Y]$. In this case $G^{*n}(na) = 1$, and so trivially $1 - F^{*n}(na) \geq 1 - G^{*n}(na)$ for any n.

Case 2: $\mathbb{E}[X] \leq a \leq \max[Y]$. Assume, without loss of generality, that $\max[Y] > \mathbb{E}[X]$. Let $A = [\mathbb{E}[X], \max[Y]]$ and consider C_A, n_A as defined in Corollary 1, applied to the random variable X. When $a \in A$ we have $\mathrm{e}^{-at}M_X(t) > \mathrm{e}^{-at}M_Y(t)$ for every $t > 0$.

Since for $a > \mathbb{E}[X]$ we have $t_X(a) > 0$, this implies

$$\rho_X(a) = M_{X-a}(t_X(a)) > M_{Y-a}(t_X(a)) \geq \rho_Y(a).$$

But even if $a = \mathbb{E}[X]$, it still holds that $\rho_X(a) = 1 = M_{Y-a}(0) > \rho_Y(a)$ since $t_Y(a) > 0$. Thus $\rho_X(a) > \rho_Y(a)$ whenever $a \in A$.

Now, Corollary 1 implies that for all $a \in A$ and $n \geq n_A$,

$$1 - F^{*n}(an) \geq C_A \cdot \frac{\rho_X(a)^n}{\sqrt{n}}, \tag{25}$$

while Lemma 3 implies

$$1 - G^{*n}(an) \leq \rho_Y(a)^n. \tag{26}$$

[13] Otherwise, we can find non-degenerate random variables $\tilde{X}$ and $\tilde{Y}$ with distributions close to X and Y, such that X dominates $\tilde{X}$ and $\tilde{Y}$ dominates Y in first-order stochastic dominance, and that $L_{\tilde{X}}(t) > L_{\tilde{Y}}(t)$ still holds for every $t \geq 0$. The result of Proposition 3 for the pair $\tilde{X}, \tilde{Y}$ implies the corresponding result for the pair X, Y.

As ρ_X and ρ_Y are continuous functions and $\rho_X(a) > \rho_Y(a)$ on A, the ratio ρ_X/ρ_Y is bounded below by $1+\varepsilon$ for some $\varepsilon > 0$.[14]

Hence, for any n such that

$$C_A > \frac{\sqrt{n}}{(1+\varepsilon)^n} \quad \text{and} \quad n \geq n_A$$

it follows from (25) and (26) that $1 - F^{*n}(an) > 1 - G^{*n}(an)$ for all $a \in A$.

Case 3: $\mathbb{E}[Y] \leq a \leq \mathbb{E}[X]$**.** By the Berry-Esseen Theorem there exist constants k_X and k_Y such that for all a,

$$\begin{aligned} \left|F^{*n}(na) - \Phi\left(\sqrt{n} \cdot \frac{a - \mathbb{E}[X]}{\sigma_X}\right)\right| &\leq \frac{k_X}{\sqrt{n}} \\ \left|G^{*n}(na) - \Phi\left(\sqrt{n} \cdot \frac{a - \mathbb{E}[Y]}{\sigma_Y}\right)\right| &\leq \frac{k_Y}{\sqrt{n}}. \end{aligned} \tag{27}$$

where Φ denotes the cdf of a standard Gaussian distribution, and σ_X, σ_Y denote the standard deviations of X and Y.

Fix $a_0 = \frac{1}{2}(\mathbb{E}[X] + \mathbb{E}[Y])$. Since $a_0 > \mathbb{E}[Y]$ there exists an N such that $n \geq N$ implies

$$G^{*n}(na_0) \geq \Phi\left(\sqrt{n} \cdot \frac{a_0 - \mathbb{E}[Y]}{\sigma_Y}\right) - \frac{\kappa_Y}{\sqrt{n}} > 0.99 - \frac{\kappa_Y}{\sqrt{n}} \geq \frac{1}{2} + \frac{\kappa_X}{\sqrt{n}} \geq F^{*n}(n \cdot \mathbb{E}[X]).$$

where the first and the last inequalities follow directly from (27). Similarly, there exists N' such that $n \geq N'$ implies

$$F^{*n}(na_0) \leq \Phi\left(\sqrt{n} \cdot \frac{a_0 - \mathbb{E}[X]}{\sigma_Y}\right) + \frac{\kappa_X}{\sqrt{n}} < 0.01 + \frac{\kappa_X}{\sqrt{n}} \leq \frac{1}{2} - \frac{\kappa_Y}{\sqrt{n}} \leq G^{*n}(n \cdot \mathbb{E}[Y]).$$

Hence for $n \geq \max\{N, N'\}$, if $a_0 \leq a \leq \mathbb{E}[X]$, then

$$G^{*n}(na) \geq G^{*n}(na_0) > F^{*n}(n \cdot \mathbb{E}[X]) \geq F^{*n}(na).$$

Conversely, if $\mathbb{E}[Y] \leq a \leq a_0$ then

$$F^{*n}(na) \leq F^{*n}(na_0) < G^{*n}(n \cdot \mathbb{E}[Y]) \leq G^{*n}(na).$$

Therefore $1 - F^{*n}(na) > 1 - G^{*n}(na)$ holds for all a in this range. Proposition 3 follows.

[14] For $\mathbb{E}[X] \leq a \leq \max[Y]$, and $\rho_Y(a) = 0$ if and only if $a = \max[Y]$ and the distribution of Y has an atom at $\max[Y]$. On the other hand, ρ_X is strictly positive on this interval.

C Preliminaries for Comparison of Experiments

We collect here some useful facts regarding the distributions of log-likelihood ratios induced by an experiment. Let $P = (\Omega, P_0, P_1)$ be an experiment and let

$$\Pi = \frac{\mathrm{d}P_1/\mathrm{d}P_0}{1 + \mathrm{d}P_1/\mathrm{d}P_0}$$

be the random variable corresponding to the posterior probability that θ equals 1. For every $A \subseteq [0,1]$ we have

$$\pi_1(A) = \int_{\Pi \in A} \mathrm{d}P_1 = \int_{\Pi \in A} \frac{\mathrm{d}P_1}{\mathrm{d}P_0}\,\mathrm{d}P_0 = \int_{\Pi \in A} \frac{\Pi}{1-\Pi}\,\mathrm{d}P_0$$

Thus

$$\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_0}(p) = \frac{p}{1-p}. \tag{28}$$

Recall that $\pi = \frac{1}{2}\pi_0 + \frac{1}{2}\pi_1$, so

$$\frac{\mathrm{d}\pi}{\mathrm{d}\pi_1}(p) = \frac{1}{2p} \tag{29}$$

We also observe that for every function ϕ that is integrable with respect to F_1, defined as in (4),

$$\int_{\mathbb{R}} \phi(u)\,\mathrm{d}F_1(u) = \int_{\mathbb{R}} \phi(-v)\mathrm{e}^{-v}\,\mathrm{d}F_0(v). \tag{30}$$

This implies that the moment generating function of F_1

$$M_{F_1}(u) = \int_{-\infty}^{\infty} \mathrm{e}^{tu}\mathrm{d}F_1(u)$$

satisfies

$$M_{F_1}(t) = M_{F_0}(-t-1) \tag{31}$$

Hence, in particular, $M_{F_1}(-1) = 1$.

C.1 Proof of Lemma 1

Given an exponential distribution with support $\mathbb{R}_+$ and cdf $H(x) = 1 - \mathrm{e}^{-x}$ for all $x \geq 0$, $\tilde{F}$ and $\tilde{G}$ can be written as

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

and similarly

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

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

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

$$\tilde{F}(a) = \int_{-a}^{\infty} F_0(v)\mathrm{e}^{-v}\mathrm{d}v \quad \text{and} \quad \tilde{G}(a) = \int_{-a}^{\infty} G_0(v)\mathrm{e}^{-v}\mathrm{d}v. \tag{32}$$

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

We now show how (32) follows from (30). We first observe that by taking $\phi(u) = 1_{(a,\infty)}(u) \cdot \mathrm{e}^{-u}$, (30) implies

$$\tilde{F}(a) = F_1(a) + \mathrm{e}^a \int_{(a,\infty)} \mathrm{e}^{-u}\mathrm{d}F_1(u) = F_1(a) + \mathrm{e}^a F_0((-a)_-) \tag{33}$$

where $F_0((-a)_-)$ denotes the left limit of F_0 evaluated at $-a$. Moreover, taking ϕ to be the indicator function of $(-\infty, a]$ implies

$$F_1(a) = \int_{-a}^{\infty} \mathrm{e}^{-v}\mathrm{d}F_0(v).$$

Integration by parts leads to

$$F_1(a) = \int_{-a}^{\infty} \mathrm{e}^{-v}\mathrm{d}F_0(v) = -\mathrm{e}^a F_0((-a)_-) - \int_{-a}^{\infty} F_0(v)\mathrm{e}^{-v}\mathrm{d}v$$

Hence by (33), we obtain

$$\tilde{F}(a) = \int_{-a}^{\infty} F_0(v)\mathrm{e}^{-v}\mathrm{d}v$$

as desired.

D Proof of Theorem 1

Throughout the proof, we use the notation introduced in §4.1 and §4.2 and further discussed in §C, as well as the notation related to large deviation estimates introduced in §A.

We first show that (i) implies (ii). As discussed in the main text, the comparison of Rényi divergences between two experiments is independent of the number of repetitions. Thus it suffices to show that the Rényi order refines the Blackwell order.

For $t > 1$, the function $v_1(p) = 2p^t(1-p)^{1-t}$ is strictly convex. Thus it is the indirect utility function induced by some decision problem. Moreover, it is straightforward to check that

$$\int_0^1 v_1(p)\,\mathrm{d}\pi(p) = \exp((t-1)R_P^1(t)),$$

which is a monotone transformation of the Rényi divergence. Thus, experiment P yields higher expected payoff in this decision problem (with indirect utility v) than Q only if $R_P^1(t) > R_Q^1(t)$.

For $t \in (0,1)$, we consider the indirect utility function $v_2(p) = -2p^t(1-p)^{1-t}$, which is now convex due to the negative sign. Observe similarly that

$$\int_0^1 v_2(p)\,\mathrm{d}\pi(p) = -\exp((t-1)R_P^1(t))$$

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

For $t = 1$, we consider the indirect utility function $v_3(p) = 2p\log(\frac{p}{1-p})$, which is strictly convex. Since

$$\int_0^1 v_3(p)\,\mathrm{d}\pi(p) = R_P^1(1),$$

P yields higher payoff only if $R_P^1(1) > R_Q^1(1)$.

Summarizing, the above family of decision problems show that P Blackwell-dominates Q only if $R_P^1(t) > R_Q^1(t)$ for all $t > 0$. Since the two states are symmetric, we also have $R_P^0(t) > R_Q^0(t)$ for all $t > 0$. This shows P dominates Q in the Rényi order.

We now show that (ii) implies (i). The assumptions that $R_P^0(1) > R_Q^0(1)$ and that $R_P^1(1) > R_Q^1(1)$ are, in terms of the notation introduced in (5), is equivalent to

$$\mathbb{E}[G_0] < \mathbb{E}[F_0] \quad \text{and} \quad \mathbb{E}[G_1] < \mathbb{E}[F_1],$$

where, with slight abuse of notation, given a cdf H we denote by $\mathbb{E}[H]$ the expectation of a random variable with distribution H.

Let $X, X_1, \ldots, X_n$ be i.i.d. and distributed according to F_1 and let $Y, Y_1, \ldots, Y_n$ be i.i.d. and distributed according to G_1. By (12) the assumption that $R_P^\theta(t) > R_Q^\theta(t)$ for all positive $t \neq 1$ is equivalent to having $M_X(t) > M_Y(t)$ for $t > 0$ and $t < -1$, and $M_X(t) < M_Y(t)$ for $t \in (-1, 0)$. In particular, $L_X(t) > L_Y(t)$ for all $t > -1$.

By Theorem 2, it suffices to show that for n large,

$$X_1 + \cdots + X_n - E \geq_1 Y_1 + \cdots + Y_n - E.$$

where E is an independent, positive, exponential random variable with density e^{-x}. That is, we need to show for n large and all $a \in \mathbb{R}$,

$$\mathbb{P}[X_1 + \cdots + X_n - E \geq na] \geq \mathbb{P}[Y_1 + \cdots + Y_n - E \geq na] \tag{34}$$

We consider a number of cases.

Case 1: $a \geq \mathbb{E}[G_1]$**.** The random variables X and Y satisfy the conditions of Proposition 3. Thus, for every n large enough and every a in this range it holds that

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

Hence $F_1^{*n}(na) \leq G_1^{*n}(na)$, and so the first statement of Lemma 1 applied to F_1^{*n}, G_1^{*n} implies

$$\widetilde{F^{*n}}(na) \leq \widetilde{G^{*n}}(na),$$

which implies (34).

Case 2: $a \leq -\mathbb{E}[G_0]$**.** Here we repeat the argument of the previous case, but applied to F_0 and G_0, instead of F_1 and G_1. The hypothesis that $M_{F_1}(t) < M_{G_1}(t)$ for all $t < -1$ is equivalent, by (31), to $M_{F_0}(t) > M_{G_0}(t)$ for all $t > 0$. Moreover $\mathbb{E}[F_0] > \mathbb{E}[G_0]$, and so the same conditions that applied in the previous case apply here. Thus there exists N such that for $n \geq N$ it holds that

$$F_0^{*n}(na) \leq G_0^{*n}(na),$$

for every $a \geq \mathbb{E}[G_0]$. Hence the second statement of Lemma 1 implies

$$\widetilde{F^{*n}}(na) \leq \widetilde{G^{*n}}(na)$$

for all $a \leq -\mathbb{E}[G_0]$.

Case 3: $-\mathbb{E}[G_0] \leq a \leq \mathbb{E}[G_1]$**.** Here we will still show (as in case 1) that

$$\mathbb{P}[X_1 + \cdots + X_n \geq na] \geq \mathbb{P}[Y_1 + \cdots + Y_n \geq na],$$

which would imply the result via Lemma 1.

Recall that $t_Y(a)$ satisfies $K'_Y(t_Y(a)) = a$. Observe that $K'_Y(0) = \mathbb{E}[G_1] \geq a$ and

$$K'_Y(-1) = K'_{G_1}(-1) = -K'_{G_0}(0) = -\mathbb{E}[G_0] \leq a.$$

Thus by convexity of K_Y, we have $t_Y(a) \in [-1, 0]$. Since $\mathbb{E}[F_1] > \mathbb{E}[G_1]$ and $\mathbb{E}[F_0] > \mathbb{E}[G_0]$, it follows that $t_X(a) \in (-1, 0)$.

Denote $A = [-\mathbb{E}[G_0],\ \mathbb{E}[G_1]]$. By (20), we have for all $n \geq (\frac{10b}{\min_{a \in A} \sigma_Y(a)})^2$ and $a \in A$,

$$\mathbb{P}[X_1 + \cdots + X_n \geq na] \geq 1 - \rho_X(a)^n$$

and

$$\mathbb{P}[Y_1 + \cdots + Y_n \geq na] \leq 1 - \frac{C(a)}{\sqrt{n}} \rho_Y(a)^n,$$

where

$$C(a) = \frac{\mathrm{e}^{10 t_Y(a) b} \cdot b}{\sigma_Y(a)}$$

is strictly positive when $a \in A$.

We now argue that $\rho_X(a) < \rho_Y(a)$ for a in this range. Indeed, since $M_Y(t) > M_X(t)$ for $t \in (0, 1)$, and since (as is true for any distribution of a log-likelihood ratio) $M_X(0) = M_X(-1) = M_Y(0) = M_Y(-1) = 1$, we have

$$\rho_Y(a) = \mathrm{e}^{-a t_Y(a)} \cdot M_Y(t_Y(a)) \geq \mathrm{e}^{-a t_Y(a)} \cdot M_X(t_Y(a)) \geq \mathrm{e}^{-a t_X(a)} \cdot M_X(t_X(a)) = \rho_X(a).$$

But the first inequality holds equal only if $t_Y(a) \in \{-1, 0\}$, in which case the second inequality must be strict, because $t_X(a) = \mathrm{argmin}_t\, \mathrm{e}^{-at} \cdot M_X(t)$ is strictly between -1 and 0.

Therefore $\rho_X(a) < \rho_Y(a)$ for $a \in A$. By continuity,

$$\gamma := \max_{a \in A} \frac{\rho_X(a)}{\rho_Y(a)}$$

is strictly below 1. We therefore conclude that

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

for every n large enough to satisfy

$$\gamma^n < \frac{\min_{a \in A} C(a)}{\sqrt{n}}.$$

This completes the proof.

E 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 mean-preserving contraction of π. We now recall that the Rényi divergences are derived from *strictly* convex indirect utility functions $u(p) = -p^t(1-p)^{1-t}$ for $0 < t < 1$ and $v(p) = p^t(1-p)^{1-t}$ for $t > 1$. Thus, $R_P^\theta(t) > R_Q^\theta(t)$ for all $\theta \in \{0, 1\}$ and $t > 0$.

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

ε is sufficiently small.[15] Since P and Q' also form a generic pair, Theorem 1 shows that P dominates Q' in large samples.

It remains to prove that P does not dominate Q' according to Blackwell. Consider a decision problem where the prior is uniform, the set of actions is $\{0,1\}$, and payoffs are given by $u(\theta=a=0)=p_2$, $u(\theta=a=1)=1-p_2$ and $u(\theta\neq a)=0$. The indirect utility function is $v(p)=\max\{(1-p)p_2,\ p(1-p_2)\}$, which is piece-wise linear on $[0,p_2]$ and $[p_2,1]$ but convex at p_2. Recall that in constructing the garbling from P to Q, those posterior beliefs under P that are below p_2 are "averaged" into the single posterior belief q_1 under Q, and those above p_2 are averaged into the belief q_2. Thus Q achieves the same expected utility in this decision problem as P (despite being a garbling). Nevertheless, observe that Q' achieves higher expected utility in this decision problem than Q.[16] Hence Q' achieves higher expected utility than P, implying that it is not Blackwell dominated.

F Proof of Proposition 2

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

$$\sqrt{\alpha(1-\alpha)} > \sqrt{\beta(\frac{1}{2}-\beta)} + \frac{1}{4}. \tag{35}$$

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

$$(\frac{1}{2}-\beta)^r\beta^{1-r} + (\frac{1}{2}-\beta)^{1-r}\beta^r + \frac{1}{2} < (1-\alpha)^r\alpha^{1-r} + (1-\alpha)^{1-r}\alpha^r, \quad \forall 0<r<1; \tag{36}$$

$$(\frac{1}{2}-\beta)^r\beta^{1-r} + (\frac{1}{2}-\beta)^{1-r}\beta^r + \frac{1}{2} > (1-\alpha)^r\alpha^{1-r} + (1-\alpha)^{1-r}\alpha^r, \quad \forall r<0 \text{ or } r>1; \tag{37}$$

$$\beta\cdot\ln(\frac{\beta}{\frac{1}{2}-\beta}) + (\frac{1}{2}-\beta)\cdot\ln(\frac{\frac{1}{2}-\beta}{\beta}) > \alpha\cdot\ln(\frac{\alpha}{1-\alpha}) + (1-\alpha)\cdot\ln(\frac{1-\alpha}{\alpha}). \tag{38}$$

To prove these, it suffices to consider the α that makes (35) hold with equality.[17] We will show that the above inequalities hold for this particular α, except that (36) holds equal at $r=\frac{1}{2}$. Let us define the following function

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

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

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

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

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

When (35) holds with equality, we have $\Delta(0) = \Delta(\frac{1}{2}) = \Delta(1) = 0$. Thus Δ has roots at 0, 1 as well as a double-root at $\frac{1}{2}$. But since Δ is a weighted sum of 4 exponential functions plus a constant, it has at most 4 roots (counting multiplicity).[18] Hence these are the only roots, and we deduce that the function Δ has constant sign on each of the intervals $(-\infty, 0), (0, \frac{1}{2}), (\frac{1}{2}, 1), (1, \infty)$.

Now observe that since $2\beta < \alpha \leq \frac{1}{2}$, it holds that $\frac{1/2-\beta}{\beta} > \frac{1-\alpha}{\alpha} > 1$. It is then easy to check that $\Delta(r) \to \infty$ as $r \to \infty$. Thus $\Delta(r)$ is strictly positive for $r \in (1, \infty)$. As $\Delta(1) = 0$, its derivative is weakly positive. But recall that we have enumerated the 4 roots of Δ. So Δ cannot have a double-root at $r = 1$, and it follows that $\Delta'(1)$ is strictly positive. Hence (38) holds.

Note that $\Delta'(1) > 0$ and $\Delta(1) = 0$ also implies $\Delta(1 - \varepsilon) < 0$. Thus Δ is negative on $(\frac{1}{2}, 1)$. A symmetric argument shows that Δ is positive on $(-\infty, 0)$ and negative on $(0, \frac{1}{2})$. Hence (36) and (37) both hold, completing the proof.

G Necessity of Genericity Assumption

Here we present examples to show that Theorem 3 and Theorem 1 do not hold without the genericity assumption.

Gambles. The following is an example where $L_X(t) > L_Y(t)$ for all $t \in \mathbb{R}$, but X does not dominate Y in large numbers because $\max[X] = \max[Y]$. Fix any $q \in (0, 1)$, and consider

$$X = \begin{cases} 10, & \text{w.p. } q \\ 2, & \text{w.p. } \frac{1-q}{2} \\ 0, & \text{w.p. } \frac{1-q}{2} \end{cases} \quad ; \qquad Y = \begin{cases} 10, & \text{w.p. } q \\ 1, & \text{w.p. } \frac{2(1-q)}{3} \\ -1, & \text{w.p. } \frac{1-q}{3} \end{cases}$$

Let $\hat{X}$ be the random variable that takes values 2 and 0 with equal probabilities; note that $\hat{X}$ is distributed as X, conditional on $X \neq 10$. Similarly define $\hat{Y}$ to take value 1 w.p. 2/3 and value -1 w.p. $\frac{1}{2}$. It is easy to check that $\hat{X}_1 + \hat{X}_2$ first-order stochastically dominates $\hat{Y}_1 + \hat{Y}_2$. As a result, $L_{\hat{X}}(t) > L_{\hat{Y}}(t)$ for all t. Since

$$M_X(t) = q \cdot \mathrm{e}^{10t} + (1 - q) \cdot M_{\hat{X}}(t),$$

we conclude that $L_X(t) > L_Y(t)$ for all t.

Nonetheless, we now show that X does not dominate Y in large numbers. For each n, consider $\mathbb{P}[X_1 + \cdots + X_n \geq 10n - 9]$. In other for this to happen, either every X_i takes value 10, or all but one X_i equals 10 and the remaining one equals 2. Thus

$$\mathbb{P}[X_1 + \cdots + X_n \geq 10n - 9] = q^n + nq^{n-1} \cdot \frac{1-q}{2}.$$

[18]This follows from Rolle's theorem and an induction argument.

Similarly we have

$$\mathbb{P}\left[Y_1 + \cdots + Y_n \geq 10n - 9\right] = q^n + nq^{n-1} \cdot \frac{2(1-q)}{3}.$$

Since the latter probability is larger, $X_1 + \cdots + X_n$ does not first-order dominate $Y_1 + \cdots + Y_n$.[19]

Nonexistence of a Generator. The above example shows that stochastic dominance in large numbers does not admit a generator. Suppose, by way of contradiction, that V was a family of functions $\phi : \mathbb{R} \to \mathbb{R}$ with the property that for all bounded random variables X and Y, X first-order dominates Y in large numbers if and only if $\mathbb{E}\left[\phi(X)\right] \geq \mathbb{E}\left[\phi(Y)\right]$ for all $\phi \in V$.

Let $X, \hat{X}, Y$ and $\hat{Y}$ be defined as in the previous paragraph. Since $\hat{X}$ dominates $\hat{Y}$ in large numbers, then it must hold that $\mathbb{E}\left[\phi(\hat{X})\right] \geq \mathbb{E}\left[\phi(\hat{Y})\right]$ for all $\phi \in V$. Hence,

$$\mathbb{E}\left[\phi(X)\right] = (1-q)\mathbb{E}\left[\phi(\hat{X})\right] + q\phi(10) = (1-q)\mathbb{E}\left[\phi(\hat{Y})\right] + q\phi(10) \geq \mathbb{E}\left[\phi(Y)\right],$$

implying that X dominates Y, a contradiction.

Experiments. Consider the experiments P and Q described in §3.2. Fix $\alpha = \frac{1}{4}$ and $\beta = \frac{1}{16}$, which satisfy (35). Then by Proposition 2, P dominates Q in large samples.

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

$\tilde{P}$:

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

$\tilde{Q}$:

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

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

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

[19] Related, a slightly modification of this example shows that even if $X_1 + \cdots + X_n \geq_1 Y_1 + \cdots + Y_n$ for all large n, this does not imply that $X_1 + \cdots + X_n \geq_1 X_1 + \cdots + X_{n-1} + Y_n$ for all large n. Indeed, suppose that $Y = 9$ (instead of 10) w.p. q in this example, then Theorem 3 applies and shows that X dominates Y in large numbers, but $\mathbb{P}\left[X_1 + \cdots + X_n \geq 10n - 9\right] < \mathbb{P}\left[X_1 + \cdots + X_{n-1} + Y_n \geq 10n - 9\right]$.

the remaining one is y_1. Thus, in the range $p > \overline{p}$, the posterior belief has the following distribution under $\tilde{Q}^{\otimes n}$:

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

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

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

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

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

That is,

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

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

H Proof of Theorem 4

(i) is equivalent to (ii). If X second-order dominates Y in large numbers, then, by considering risk-averse CARA utility functions, we obtain that $L_X(t) > L_Y(t)$ for all $t < 0$; the strict inequality is because these utility functions are strictly concave. By continuity we thus have $L_X(0) \geq L_Y(0)$, which implies $\mathbb{E}[X] \geq \mathbb{E}[Y]$. Since by assumption they are unequal, we in fact have $\mathbb{E}[X] > \mathbb{E}[Y]$. Hence (ii) implies (i).

To show (i) implies (ii), suppose $L_X(t) > L_Y(t)$ for all $t \leq 0$. As in the proof of Theorem 3, We assume without loss of generality that X and Y are non-degenerate, and denote by F^{*n} (resp. G^{*n}) the c.d.f. of the sum of n i.i.d. copies of X (resp. Y). Furthermore, by shifting X and Y by a constant, we can assume $\mathbb{E}[X] = \mu$ and $\mathbb{E}[Y] = -\mu$ for some positive number μ.

To prove second-order stochastic dominance, we need to show that for n large enough and for every $x \in \mathbb{R}$ it holds that

$$\int_{-\infty}^{x} G^{*n}(t) - F^{*n}(t)\,\mathrm{d}t \geq 0 \tag{39}$$

We again consider a few cases.

Case 1: $x \le 0$**.** In this case Proposition 3 applied to the random variables $-Y$ and $-X$ implies $G^{*n}(x) \ge F^{*n}(x)$ for all $x \le n\mu$. Hence (39) holds too.

Case 2: $x \ge 0$**.** Note that, as can be shown by integration by parts,

$$\int_{-\infty}^{\infty} G^{*n}(t) - F^{*n}(t)\,\mathrm{d}t = n\mathbb{E}[X] - n\mathbb{E}[Y] = 2n\mu$$

Hence

$$\begin{aligned}\int_{-\infty}^{x} G^{*n}(t) - F^{*n}(t)\,\mathrm{d}t &= 2n\mu - \int_{x}^{\infty} G^{*n}(t) - F^{*n}(t)\,\mathrm{d}t \\ &= 2n\mu - \int_{x}^{\infty} (1 - F^{*n}(t)) - (1 - G^{*n}(t))\,\mathrm{d}t \\ &\ge 2n\mu - \int_{0}^{\infty} 1 - F^{*n}(t)\,\mathrm{d}t \end{aligned} \tag{40}$$

Now, again using integration by parts we have that

$$\begin{aligned}\int_{0}^{\infty} 1 - F^{*n}(t)\,\mathrm{d}t &= n\mu + \int_{-\infty}^{0} F^{*n}(t)\,\mathrm{d}t \\ &= n\mu + \int_{n\min[X]}^{0} F^{*n}(t)\,\mathrm{d}t \le n\mu + n \cdot |\min[X]| \cdot F^{*n}(0).\end{aligned}$$

By the Chernoff bound (i.e., (18) in Lemma 3), $F^{*n}(0) \le \rho_{-X}(0)^n$. Since $\rho_{-X}(0) < 1$, for n large enough we have that the above is at most $\frac{3}{2}n\mu$. Applying this estimate to (40) yields

$$\int_{-\infty}^{x} G^{*n}(t) - F^{*n}(t)\,\mathrm{d}t \ge \frac{1}{2}n\mu$$

for all $x \ge 0$, and we have ths shown that (i) is equivalent to (ii).

(ii) is equivalent to (iii). Condition (ii) trivially implies (iii). Conversely, suppose it holds for some n and k that $X_1 + \cdots + X_n \ge_k Y_1 + \cdots + Y_n$. Since the risk-averse CARA utility function $u(x) = e^{-tx}$ has derivatives that alternate signs, by definition of $\ge_k$ we know that each risk-averse CARA agent prefers $X_1 + \cdots + X_n$ to $Y_1 + \cdots + Y_n$. Thus $L_X(t) > L_Y(t)$ for all $t \le 0$. Hence (ii) follows by Theorem 4.

(i) is equivalent to (iv). This follows by applying Theorem 4 in Fishburn (1980), which shows that for bounded random variables X and Y with $\min[X] \ne \min[Y]$ and $\mathbb{E}[X] \ne \mathbb{E}[Y]$, $X \ge_k Y$ for some $k \ge 2$ if and only if $L_X(t) > L_Y(t)$ for all $t < 0$.[20] Since $L_X(t) > L_Y(t)$ for all $t < 0$ holds by Theorem 4.

[20] In the notation used there, $F, G \in P^*$ because X and Y are bounded, the condition $G <_0 F$ is satisfied since $\min[X] > \min[Y]$, and the condition $\mu_F >_L \mu_G$ is satisfied because $\mathbb{E}[X] > \mathbb{E}[Y]$.

I Proof of Theorem 5

We first show (ii) implies (i). Observe that by assumption, $\mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[Y_1 + \cdots + Y_n]$ for each n. Thus $X_1 + \cdots + X_n$ second-order stochastically dominates $Y_1 + \cdots + Y_n$ if and only if the latter is a mean-preserving spread of the former. Thus, for every strictly convex function $\phi : \mathbb{R} \to \mathbb{R}$ (not necessarily increasing), it holds that

$$\mathbb{E}[\phi(X_1 + \cdots + X_n)] < \mathbb{E}[\phi(Y_1 + \cdots + Y_n)].$$

Choosing $\phi(x) = x^2$ and using $\mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[Y_1 + \cdots + Y_n]$, we deduce that $\mathrm{Var}(X_1 + \cdots + X_n) < \mathrm{Var}(Y_1 + \cdots + Y_n)$, and so $\mathrm{Var}(X) < \mathrm{Var}(Y)$. Moreover, choosing $\phi(x) = e^{tx}$ implies that $M_X(t) < M_Y(t)$ for all $t \neq 0$. This is equivalent to $L_X(t) < L_Y(t)$ for all $t > 0$ and $L_X(t) > L_Y(t)$ for all $t < 0$, as we desire to show.

Below we prove that (i) implies (ii). Since $L_X(t) < L_Y(t)$ for all $t > 0$, taking $t \to \infty$ yields $\max[X] \leq \max[Y]$ by continuity. But since X and Y are generic, we in fact have $\max[X] < \max[Y]$. Similarly we have $\min[X] > \min[Y]$. We also assume without loss of generality that $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. Thus X and Y are bounded, zero mean random variables satisfying the following conditions:

(i). $\min[X] > \min[Y]$.

(ii). $\max[X] < \max[Y]$.

(iii). $\mathrm{Var}(X) < \mathrm{Var}(Y)$

(iv). $M_X(t) < M_Y(t)$ for all $t \neq 0$.

As in the proof of Theorem 3, denote by F^{*n} (resp. G^{*n}) the c.d.f. of the sum of n i.i.d. copies of X (resp. Y). Let $b > 0$ be a number such that X and Y are supported on $[-b/2, b/2]$.

To prove (ii) we need to show that for n large enough and for every $x \in \mathbb{R}$ it holds that

$$W(x) := \int_{-\infty}^{x} G^{*n}(y) - F^{*n}(y)\,\mathrm{d}y \geq 0. \tag{41}$$

Since $\mathbb{E}[X] = \mathbb{E}[Y]$, integration by parts shows that $W(x) = 0$ for x sufficiently large. Thus we also have $W(x) = \int_x^{\infty} F^{*n}(y) - G^{*n}(y)\,\mathrm{d}y$. The above inequality reduces to

$$\int_x^{\infty} F^{*n}(t) - G^{*n}(t)\,\mathrm{d}t \geq 0. \tag{42}$$

We will show that (42) holds for all $x \geq 0$. The case of $x \leq 0$ follows by applying the same argument to $-X$ and $-Y$, which also satisfy the above four conditions.

As before, we write $x = na$ and consider a few cases.

Case 1: $\max[X] < a$**.** In this range $F^{*n} = 1$, and hence $F^{*n} \geq G^{*n}$ point-wise.

Case 2: $\varepsilon \leq a \leq \max[X]$**,** with $\varepsilon > 0$ chosen in case 3 below. Note that $L_Y(t) > L_X(t)$ for all $t > 0$, so the random variables Y and X almost satisfy the assumptions of Proposition 3, except that $L_Y(0) = L_X(0)$ (which equals their common expected value). However, since we have $a \geq \varepsilon$, we can follow the analysis in the proof of Proposition 3 and deduce that $\rho_Y(a) > \rho_X(a)$ in this range. The result of Proposition 3 thus gives

$$1 - G^{*n}(na) \geq 1 - F^{*n}(na)$$

for all n large enough (depending on ε) and $a \geq \varepsilon$. The integral in (42) is thus positive in this range.

Case 3: $\sqrt{\frac{1}{2}\operatorname{Var}(X)\frac{\log n}{n}} \leq a \leq \varepsilon$**.** Define $r_X(a) = \log \rho_X(a)$ (and r_Y analogously). It follows from Lemma 4 that

$$1 - F^{*n}(na) \leq \exp\left(n \cdot r_X(a)\right) \cdot \frac{1 + \sqrt{2\pi} t_X(a) b}{\sqrt{2\pi}\sigma_X(a) t_X(a)\sqrt{n}},$$

and that

$$1 - G^{*n}(na) \geq \exp\left(n \cdot r_Y(a)\right) \frac{1 - 2\sqrt{2\pi} t_Y(a) b}{2\sqrt{2\pi}\sigma_Y(a) t_Y(a)\sqrt{n}},$$

provided that

$$n \geq [\sigma_Y(a) t_Y(a)]^{-2}.$$

By Lemma 5 below $r_X'(0) = -t_X(0) = 0$ and

$$r_X''(0) = -t_X'(0) = -\frac{1}{K_X''(0)} = -\frac{1}{\operatorname{Var}(X)}.$$

Hence by Taylor expansion, we can write

$$r_X(a) = r_X(0) + r_X'(0) + \frac{1}{2} r_X''(0) a^2 + O(a^3) = -\frac{1 + O(\varepsilon)}{2\operatorname{Var}(X)}$$

for $0 \leq a \leq \varepsilon$; similarly for $r_Y(a)$.

Note that $t_X(0) = 0$, so $t_X(a) = O(\varepsilon)$. Also, $\sigma_X(0) = \operatorname{Std}(X)$ (the standard deviation of X), which implies $\sigma_X(a) = (1 + O(\varepsilon))\operatorname{Std}(X)$. Moreover, $t_X'(0) = \frac{1}{\operatorname{Var}(X)} > t_Y'(0)$, so $t_X(a) > t_Y(a)$ for $0 \leq a \leq \varepsilon$ whenever ε is sufficiently small. Plugging all of these estimates into the above inequality for F^{*n}, we have

$$1 - F^{*n}(na) \leq \exp\left(-\frac{D(\varepsilon) a^2 n}{2\operatorname{Var}(X)}\right) \frac{1}{2\sqrt{2\pi\operatorname{Var}(X) n} \cdot t_Y(a) D(\varepsilon)}, \tag{43}$$

where $D(\varepsilon) < 1$ is a shorthand for $1 - O(\varepsilon)$, which approaches 1 as $\varepsilon \to 0$.

Similarly,

$$1 - G^{*n}(na) \geq \exp\left(-\frac{a^2 n}{2D(\varepsilon)\operatorname{Var}(Y)}\right) \frac{D(\varepsilon)}{2\sqrt{2\pi \operatorname{Var}(Y) n} \cdot t_Y(a)}, \tag{44}$$

provided that

$$n \geq [\sigma_Y(a) t_Y(a)]^{-2}. \tag{45}$$

Considering the ratio between (43) and (44), we obtain

$$\frac{1 - G^{*n}(na)}{1 - F^{*n}(na)} \geq \exp\left(\frac{1}{2}\left(\frac{D(\varepsilon)}{\operatorname{Var}(X)} - \frac{1}{D(\varepsilon)\operatorname{Var}(Y)}\right) a^2 n\right) \frac{\operatorname{Std}(X)}{2\operatorname{Std}(Y)}.$$

Denote

$$V_{XY} = \frac{1}{4}\left(\frac{1}{\operatorname{Var}(X)} - \frac{1}{\operatorname{Var}(Y)}\right), \tag{46}$$

which is positive since $\operatorname{Var}(X) < \operatorname{Var}(Y)$. We now choose ε small enough so that

$$\frac{1}{2}\left(\frac{D(\varepsilon)}{\operatorname{Var}(X)} - \frac{1}{D(\varepsilon)\operatorname{Var}(Y)}\right) > V_{XY} > 0.$$

For this ε, it thus holds that

$$\frac{1 - G^{*n}(na)}{1 - F^{*n}(na)} \geq \exp\left(V_{XY} a^2 n\right) \frac{\operatorname{Std}(X)}{2\operatorname{Std}(Y)}.$$

Since we are considering the case that $a^2 \geq \frac{1}{2}\operatorname{Var}(X)\frac{\log n}{n}$ we have that

$$\frac{1 - G^{*n}(na)}{1 - F^{*n}(na)} \geq \exp\left(\frac{1}{2} V_{XY} \operatorname{Var}(X) \log n\right) \frac{\operatorname{Std}(X)}{2\operatorname{Std}(Y)}.$$

This is larger than one for n large enough. So we still have $F^{*n}(na) \geq G^{*n}(na)$ point-wise.

Lastly, we need to verify the condition (45). As we noted above, $\sigma_Y(0) = \operatorname{Std}(Y)$, $t_Y(0) = 0$ and $t'_Y(0) = 1/\operatorname{Var}(Y)$. So for ε small enough and all a such that $\frac{1}{2}\operatorname{Var}(X)\frac{\log n}{n} \leq a^2 \leq \varepsilon^2$ we have $\sigma_Y(a) = (1 + O(\varepsilon))\operatorname{Std}(Y)$ and $t_Y(a) = (1 + O(\varepsilon))a/\operatorname{Var}(Y)$. Hence

$$[\sigma_Y(a) t_Y(a)]^{-2} \leq \frac{2\operatorname{Var}(Y)}{a^2} \leq \frac{4\operatorname{Var}(Y)}{\operatorname{Var}(X)} \frac{n}{\log n},$$

And so condition (45) will hold for all n sufficiently large.

Case 4: $\sqrt{\frac{1}{n}} \le a \le \sqrt{\frac{1}{2}\operatorname{Var}(X)\frac{\log n}{n}}$**.** By the Berry-Esseen Theorem

$$F^{*n}(na) \ge \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a\sqrt{n}/\operatorname{Std}(X)} \mathrm{e}^{-x^2/2}\,\mathrm{d}x - \frac{k}{\sqrt{n}}$$

and

$$G^{*n}(na) \le \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a\sqrt{n}/\operatorname{Std}(Y)} \mathrm{e}^{-x^2/2}\,\mathrm{d}x + \frac{k}{\sqrt{n}},$$

where k is a constant depending only on the distribution of X and Y. Hence

$$F^{*n}(na) - G^{*n}(na) \ge \frac{1}{\sqrt{2\pi}} \int_{a\sqrt{n}/\operatorname{Std}(Y)}^{a\sqrt{n}/\operatorname{Std}(X)} \mathrm{e}^{-x^2/2}\,\mathrm{d}x - \frac{2k}{\sqrt{n}}.$$

Since $\mathrm{e}^{-x^2/2}$ is decreasing in this range we can lower bound the integrand by its right limit, yielding

$$F^{*n}(na) - G^{*n}(na) \ge \left(\frac{1}{\operatorname{Std}(X)} - \frac{1}{\operatorname{Std}(Y)}\right) a\sqrt{n} \cdot \mathrm{e}^{-a^2 n/(2\operatorname{Var}(X))} - \frac{2k}{\sqrt{n}}. \tag{47}$$

Applying the assumption $\frac{1}{n} \le a^2 \le \frac{1}{2}\operatorname{Var}(X)\frac{\log n}{n}$ yields

$$F^{*n}(na) - G^{*n}(na) \ge \left(\frac{1}{\operatorname{Std}(X)} - \frac{1}{\operatorname{Std}(Y)}\right) n^{-1/4} - 2k \cdot n^{-1/2}, \tag{48}$$

which is again positive for all n large enough.

Case 5: $0 \le a \le \sqrt{\frac{1}{n}}$**.** Recall that we defined

$$W(x) = \int_x^\infty F^{*n}(y) - G^{*n}(y)\,\mathrm{d}y.$$

From cases 1-4, we have shown that for $y \ge \sqrt{n}$, $F^{*n}(y) - G^{*n}(y) \ge 0$ point-wise. Moreover, from (48) we in fact have

$$F^{*n}(y) - G^{*n}(y) \ge cn^{-1/4}$$

for n large enough and $\sqrt{n} \le y \le \sqrt{\frac{1}{2}\operatorname{Var}(X) n \log n}$, where c is a positive constant independent of n. Integrating this estimate over the range of y to which it applies, we deduce

$$W(\sqrt{n}) \ge \int_{\sqrt{n}}^{\sqrt{\frac{1}{2}\operatorname{Var}(X) n\log n}} F^{*n}(y) - G^{*n}(y)\,\mathrm{d}y = cn^{1/4}.$$

On the other hand, for $y \in [0, \sqrt{n}]$, it follows from 47 that

$$F^{*}(x) - G^{*n}(x) \ge -\frac{2k}{\sqrt{n}}.$$

So for any $x \in [0, \sqrt{n}]$,

$$\begin{aligned}
W(x) &= \int_x^\infty F^{*n}(y) - G^{*n}(y)\,\mathrm{d}t \\
&= \int_x^{\sqrt{n}} F^{*n}(y) - G^{*n}(y)\,\mathrm{d}y + W(\sqrt{n}) \\
&\geq -\frac{2k}{\sqrt{n}}\sqrt{n} + W(\sqrt{n}) \\
&= -2k + cn^{1/4}
\end{aligned}$$

which is positive for n large enough. This completes the proof that $W(x) \geq 0$ for all $x \geq 0$, and the theorem follows.

I.1 Additional Lemma

Lemma 5. *Let X be a bounded, zero mean random variable, and define $r_X(a) = \log \rho_X(a)$. Then $r'_X(0) = -t_X(0) = 0$ and $r''_X(0) = -t'_X(0) = -1/\operatorname{Var}(X)$.*

Proof. We suppress the subscript X in this proof. Observe that $r(a) = \inf_t K(t) - at$. So by the envelope theorem, $r'(a) = -t(a)$. Since $t(\mathbb{E}[X]) = 0$, we deduce $r'(0) = -t(0) = 0$.

Moreover, we have $r''(a) = -t'(a)$. Now recall that $t(a)$ satisfies $K'(t(a)) = a$, and so $t'(a) = \frac{1}{K''(t(a))}$. But from $K(t) = \log \mathbb{E}\left[e^{tX}\right]$ it is easy to deduce $K''(0) = \operatorname{Var}(X)$. Hence $r''_X(0) = -t'_X(0) = -1/\operatorname{Var}(X)$ as desired. □

J Stochastic Dominance implies Dominance in Large Numbers

Lemma 6. *Suppose $X \geq_k Y$ for some $k \geq 1$. Then for each n and i.i.d. replicas $X_1, \ldots, X_n$ of X and $Y_1, \ldots, Y_n$ of Y, it holds that*

$$X_1 + \cdots + X_n \geq_k Y_1 + \cdots + Y_n.$$

Proof. We first show that if $X \geq_k Y$, then $X + Z \geq_k Y + Z$ whenever Z is independent of both X and Y. Indeed, by definition we need to show $\mathbb{E}[u(X+Z)] \geq \mathbb{E}[u(Y+Z)]$ for any u whose first k derivatives have alternating signs. The assumption that $X \geq_k Y$ shows $\mathbb{E}[u(X+z)] \geq \mathbb{E}[u(Y+z)]$ for every realization z, since $u(\cdot + z)$ also has k derivatives that alternate signs. Integrating over z then yields the claim. Repeatedly applying this result, we obtain

$$X_1+\cdots+X_n \geq_k X_1+\cdots+X_{n-1}+Y_n \geq_k X_1+\cdots+X_{n-2}+Y_{n-1}+Y_n \geq_k \cdots \geq_k Y_1+\cdots+Y_n.$$

This proves the lemma. □

K Connection to Other Stochastic Orders

Aumann and Serrano (2008) and Foster and Hart (2009) propose two criteria for measuring the riskiness of a gamble. They focus on random variables X with $\mathbb{E}[X] > 0$ and $\mathbb{P}[X < 0] > 0$. The Aumann-Serrano riskiness index is the unique positive number $R_{AS}(X)$ such that

$$\mathbb{E}\left[e^{-\frac{X}{R_{AS}(X)}}\right] = 1.$$

On the other hand, the Foster-Hart measure of riskiness is the unique positive number $R_{FH}(X)$ such that

$$\mathbb{E}\left[\log\left(1 + \frac{X}{R_{FH}(X)}\right)\right] = 0.$$

Hart (2011) recognizes that these indices induce two complete orderings over gambles that refine second-order stochastic dominance. That is, we can define X to dominate Y if and only if $R_{AS}(X) \le R_{AS}(Y)$ (or $R_{FH}(X) \le R_{FH}(Y)$, respectively). He provides behavioral characterizations of these orders, which are called "uniform-wealth dominance" and "uniform-utility dominance."

In what follows, we show that if $L_X(t) \ge L_Y(t)$ for all $t \le 0$, then X is less risky than Y according to both Aumann-Serrano and Foster-Hart. This, together with one direction of Theorem 4, then proves that the two uniform dominance orders in Hart (2011) both refine our second-order in large numbers dominance order.

To show $R_{AS}(X) \le R_{AS}(Y)$, let a denote $\frac{1}{R_{AS}(X)}$ and b denote $\frac{1}{R_{AS}(Y)}$. By definition we have $M_X(-a) = 1$. But since $L_X(-a) \ge L_Y(-a)$ by assumption, we obtain $M_Y(-a) \ge M_X(-a) = 1$. From $M_Y(-a) \ge 1$, $M_Y(0) = M_Y(-b) = 1$, and the strict convexity of $M_Y(t)$, we can conclude that $b \le a$. Thus

$$R_{AS}(X) = 1/a \le 1/b = R_{AS}(Y).$$

To show $R_{FH}(X) \le R_{FH}(Y)$, we similarly denote $c = \frac{1}{R_{FH}(X)}$ and $d = \frac{1}{R_{FH}(Y)}$. By definition,

$$\mathbb{E}[\log(1 + cX)] = 0 = \mathbb{E}[\log(1 + dY)].$$

Consider the utility function $u(x) = \log(1 + dx)$. Observe that for $x > \frac{-1}{d}$, $u(x)$ has derivatives that alternate signs. By Bernstein's theorem, $u(x)$ can be written as a mixture of linear functions and exponential functions $\{-e^{-tx}\}_{0 \le t \le \infty}$. Since by assumption $M_X(t) \le M_Y(t)$ for all $t \le 0$, we deduce that $\mathbb{E}[u(X)] \ge \mathbb{E}[u(Y)]$.[21] In other words,

$$\mathbb{E}[\log(1 + dX)] \ge \mathbb{E}[\log(1 + dY)] = 0.$$

[21]Note that $L_X(t) \ge L_Y(t)$ for $t \to -\infty$ implies $\min[X] \ge \min[Y]$. Thus whenever $\mathbb{E}[\log(1 + dY)]$ is defined, so is $\mathbb{E}[\log(1 + dX)]$.

Now observe that the function $g(\lambda) = \mathbb{E}\left[\log(1+\lambda X)\right]$ is strictly concave in λ, and $g(0) = g(c) = 0 \leq g(d)$. Hence $d \leq c$. It follows that

$$R_{FH}(X) = 1/c \leq 1/d = R_{FH}(Y).$$

L Proof of a Conjecture Regarding Majorization

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

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

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

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

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

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

$$H_\mu(\alpha) = \log|S| - R_P^1(\alpha). \tag{49}$$

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

$$H_\mu(\alpha) = \log|S| - \frac{\alpha}{1-\alpha} R_P^0(1-\alpha), \tag{50}$$

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

$$H'_\mu(0) = -R_P^0(1) = \log|S| + \frac{1}{|S|}\sum_{s \in S} \log \mu(s). \tag{51}$$

As shown by Torgersen (1985, p. 264), when μ and ν have the same support size, then majorization of ν by μ is equivalent to Blackwell domination of P^ν by P^μ. Thus Jensen's Proposition 3.7, which assumes that the support sizes are different, has no implications for Blackwell dominance. However, our result on Blackwell dominance does have implications for majorization. In particular, the following proposition follows immediately from the application of Theorem 1 to experiments of the form P^μ.

Proposition 4. *Let μ, ν be finitely supported distributions with the same support size (i.e., $H_\mu(0) = H_\nu(0)$), and such that $H_\mu(\infty) \neq H_\nu(\infty)$ and $H_\mu(-\infty) \neq H_\nu(-\infty)$. Then the following are equivalent:*

(i). $H_\mu(\alpha) < H_\nu(\alpha)$ *for all* $\alpha \in (0, \infty]$, $H_\mu(\alpha) > H_\nu(\alpha)$ *for all* $\alpha \in [-\infty, 0)$ *and* $H'_\mu(0) < H'_\nu(0)$.[22]

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

Proof. For notational ease, let P denote P^μ and Q denote P^ν. The assumption $H_\mu(\alpha) < H_\nu(\alpha)$ for all $\alpha > 0$ is equivalent, via (49), to $R^1_P(t) > R^1_Q(t)$ for all $t > 0$, and to $R^0_P(t) > R^0_Q(t)$ for all $t \in (0, 1)$, using $R^0_P(t) = \frac{t}{1-t} R^1_P(1-t)$ for $0 < t < 1$.

On the other hand, $H_\mu(\alpha) > H_\nu(\alpha)$ for all $\alpha < 0$ and $H'_\mu(0) < H'_\nu(0)$ is equivalent, via (50) and (51), to $R^0_P(t) > R^0_Q(t)$ for all $t \geq 1$. So (i) is equivalent to P dominating Q in the Rényi order.

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

References

A. Alzaid, J. S. Kim, and F. Proschan. Laplace ordering and its applications. *Journal of Applied Probability*, 28(1):116–130, 1991.

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

R. J. Aumann and R. Serrano. An economic index of riskiness. *Journal of Political Economy*, 116(5):810–836, 2008.

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

R. R. Bahadur and R. R. Rao. On deviations of the sample mean. *Ann. Math. Statist*, 31 (4):1015–1027, 1960.

S. Benartzi and R. H. Thaler. Risk aversion or myopia? choices in repeated gambles and retirement investments. *Management science*, 45(3):364–381, 1999.

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

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

D. Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, pages 265–272, 1953.

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

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

J. Caballé and A. Pomansky. Mixed risk aversion. *Journal of economic theory*, 71(2): 485–513, 1996.

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

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

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

P. C. Fishburn. Continua of stochastic dominance relations for unbounded probability distributions. *Journal of Mathematical Economics*, 7(3):271–285, 1980.

D. P. Foster and S. Hart. An operational measure of riskiness. *Journal of Political Economy*, 117(5):785–814, 2009.

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

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

C. Gollier. Repeated optional gambles and risk aversion. *Management Science*, 42(11): 1524–1530, 1996.

C. Gollier. *The economics of risk and time.* MIT press, 2004.

S. Hart. Comparing risks by acceptance and rejection. *Journal of Political Economy*, 119 (4):617–638, 2011.

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

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

A. K. Jensen. Asymptotic majorization of finite probability distributions. *IEEE Transactions on Information Theory*, 2019.

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

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

M. S. Kimball. Precautionary saving in the small and in the large. *Econometrica*, 58(1): 53–73, 1990.

M. S. Kimball. Precautionary motives for holding assets. Technical report, National Bureau of Economic Research, 1991.

M. S. Kimball. Standard risk aversion. *Econometrica*, pages 589–611, 1993.

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

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

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

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

A. Müller and D. Stoyan. *Comparison methods for stochastic models and risks*, volume 389. Wiley New York, 2002.

V. V. Petrov. On the probabilities of large deviations for sums of independent random variables. *Theory of Probability & Its Applications*, 10(2):287–298, 1965.

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

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

J. W. Pratt and R. J. Zeckhauser. Proper risk aversion. *Econometrica*, pages 143–154, 1987.

M. Rabin and R. H. Thaler. Anomalies: risk aversion. *Journal of Economic perspectives*, 15(1):219–232, 2001.

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

H. Reuter and T. Riedrich. On maximal sets of functions compatible with a partial ordering for distribution functions. *Mathematische Operationsforschung und Statistik. Series Optimization*, 12(4):597–605, 1981.

S. A. Ross. Adding risks: Samuelson's fallacy of large numbers revisited. *Journal of Financial and Quantitative Analysis*, 34(3):323–339, 1999.

P. A. Samuelson. Risk and uncertainty: A fallacy of large numbers. *Scientia*, 1963.

M. Shaked and J. G. Shanthikumar. *Stochastic orders*. Springer Science & Business Media, 2007.

L. Smith and P. Sørensen. Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398, 2000. ISSN 1468-0262.

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

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

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

I. S. Tyurin. An improvement of upper estimates of the constants in the Lyapunov theorem. *Russian Mathematical Surveys*, 65(3):201–202, 2010.