"Dice"-sion Making under Uncertainty:
When Can a Random Decision Reduce Risk?

Erick Delage[1], Daniel Kuhn[2], and Wolfram Wiesemann[3]

[1]Department of Decision Sciences, HEC Montréal, Canada
[2]Risk Analytics and Optimization Chair, École Polytechnique Fédérale de Lausanne, Switzerland
[3]Imperial College Business School, Imperial College London, United Kingdom

May 4, 2018

Abstract

Stochastic programming and distributionally robust optimization seek *deterministic* decisions that optimize a risk measure, possibly in view of the most adverse distribution in an ambiguity set. We investigate under which circumstances such deterministic decisions are strictly outperformed by *random* decisions which depend on a randomization device producing uniformly distributed samples that are independent of all uncertain factors affecting the decision problem. We find that in the absence of distributional ambiguity, deterministic decisions are optimal if both the risk measure and the feasible region are convex, or alternatively if the risk measure is mixture-quasiconcave. We show that several risk measures, such as mean (semi-)deviation and mean (semi-)moment measures, fail to be mixture-quasiconcave and can therefore give rise to problems in which the decision maker benefits from randomization. Under distributional ambiguity, on the other hand, we show that for any ambiguity averse risk measure satisfying a mild continuity property we can construct a decision problem in which a randomized decision strictly outperforms all deterministic decisions.

Keywords: stochastic programming; risk measures; distributionally robust optimization; ambiguity aversion; randomized decisions

1 Introduction

Would you choose your life partner by spinning the fortune wheel? Would you decide on a company merger by tossing a coin? Replacing sound reasoning with the flip of a coin appears to violate the very fundamentals of rational decision-making, namely that actions should be assessed in view of their desirability towards a goal, and that eliminating uncertainty is always desirable. Nevertheless, random decision-taking, or *flipism*,[1] has a long history. The Kantu' farmers in Borneo selected their farming sites through a ritualized form of birdwatching, the Naskapi hunters in Canada chose their hunting sites by reading the cracks on the scapula of a caribou, and the Azande people of central Africa took difficult decisions by poisoning a chicken and seeing whether it survived.[2] Public officials in ancient Athens were frequently chosen by lottery, and the nobles of Renaissance Venice selected their head of state through a partially randomized process so that no group could impose its will without an overwhelming majority or exceptional luck. More recently, British Columbia and Ontario have used randomly selected panels of Canadian citizens to propose changes in electoral regulations, and the International Skating Union adopted a random voting system in figure-skating competitions. Many more examples of randomized decision-making in politics, military and society have been compiled by Stone (2011). We conclude that randomization—either in its explicit form or disguised as a divine dispensation—has survived and thrived as a quick, cheap and fair decision aid that avoids unintentional biases and deliberate interferences in the decision-making process.

In this paper we explore whether randomization can help a rational decision maker who has all the resources to determine an optimal solution, and who is not subjected to the strategic reactions of other agents. To this end, we assume that the decision maker has access to a randomization device which produces a uniformly distributed sample from the interval $[0, 1]$ that is independent of all uncertain factors impacting the decision problem, and we investigate under which circumstances the decision maker can benefit from taking decisions that depend on the outcome of this randomization device. We illustrate this seemingly counterintuitive idea with an example.

Example 1 (Project Selection). *A manager must implement one out of five candidate projects.*

___

[1]Flipism refers to a pseudophilosophy under which all decisions are made by flipping a coin. The name originates from the Disney comic "Flip Decision" (Walt Disney Comics & Stories 149, Vol. 13, No. 5, 1953).

[2]After visiting the Azande in the 1920s, the British anthropologist E. E. Evans-Pritchard adopted this custom and concluded: "I found this as satisfactory a way of running my home and affairs as any other I know of".

	Low NPV	High NPV	Mean	M/SV
Project 1	0.45	0.59	0.52	−0.52
Project 2	0.04	1.79	0.92	−0.53
Project 3	−0.05	3.50	1.73	−0.15
Project 4	−0.10	5.39	2.65	1.13
Project 5	−0.51	6.62	3.06	3.30

Table 1. Low and high NPV scenarios (in $1,000,000$), as well means and mean semi-variances ('M/SV') of the five projects from Example 1.
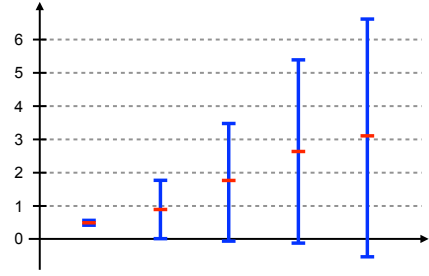
Figure 1. Low and high NPV scenarios.

The projects' net present values (NPVs) follow independent Bernoulli distributions that assign equal probabilities $p_i^L = p_i^H = 0.5$ to the low and high NPV scenarios from Table 1, see also Figure 1. Note that projects with higher indices display a higher expected NPV as well as a higher dispersion, as indicated by the increasing spread between the associated NPV scenarios. If the manager aims to maximize the expected NPV, then she selects project 5. The manager cannot improve upon this pure (i.e., deterministic) choice by making a random choice as long as the NPVs are independent of the randomization device. If the manager is risk-averse and uses the mean semi-variance $\rho(X) = -\mathbb{E}[X] + \mathbb{E}[\mathbb{E}[X] - X]_+^2$ to measure the risk of an uncertain position X, on the other hand, then the best pure choice is to select project 2 (with risk -0.53), followed by projects 1 (with risk -0.52) and 3 (with risk -0.15). Interestingly, the manager could reduce her risk exposure to -0.57 by randomly selecting one of the individually suboptimal projects 1 and 3 with equal probability. In fact, this randomized decision results in an expected NPV of

$$\mathbb{E}[X] = \frac{1}{2}\left(\frac{1}{2}\cdot 0.45 + \frac{1}{2}\cdot 0.59\right) + \frac{1}{2}\left(-\frac{1}{2}\cdot 0.05 + \frac{1}{2}\cdot 3.50\right) = 1.23$$

and a semi-variance of

$$\mathbb{E}[\mathbb{E}[X] - X]_+^2 = \frac{1}{2}\left(\frac{1}{2}\cdot[1.225 - 0.45]_+^2 + \frac{1}{2}\cdot[1.225 - 0.59]_+^2\right)$$
$$+ \frac{1}{2}\left(\frac{1}{2}\cdot[1.225 + 0.05]_+^2 + \frac{1}{2}\cdot[1.225 - 3.50]_+^2\right) = 0.66.$$

One can show that the globally optimal strategy randomizes over the projects 1 and 4 with probabilities 81% *and* 19%, *respectively, and that it yields a mean semi-variance of* -0.69.

The example highlights that the potential benefits of randomization are, at least in part, determined by the choice of risk measure. In the remainder of the paper, we will call a risk measure *randomization-receptive* if a decision maker employing the risk measure can benefit from randomization, and we call the risk measure *randomization-proof* otherwise. We will see that randomization-receptiveness is intimately related to the mixture-quasiconcavity of the considered risk measure.

Traditionally, decision theory assumes that the probability distribution governing the stochastic problem parameters is known precisely. This often contradicts managerial practice, where this distribution can neither be deduced from theoretical models nor be approximated to a satisfactory degree from historical obserations. For such settings, the distributionally robust optimization literature suggests to construct an ambiguity set that contains all plausible distributions, and to optimize a risk measure in view of the worst (*i.e.*, most adverse) distribution from within the ambiguity set (Delage and Ye 2010, Wiesemann et al. 2014b). This approach enjoys strong justification from decision theory, where it has been argued that decision makers exhibit a low tolerance towards distributional ambiguity (Epstein 1999, Gilboa and Schmeidler 1989). In such cases, randomization can reduce, and sometimes even eliminate, the effects of ambiguity.

Example 2 (The Ellsberg Urn Game). *Consider an urn that contains an unknown number of red and blue balls in equal proportion. A player is asked to name one of the two colors and to draw a random ball from the urn. If the chosen ball has the stated color, the player incurs a penalty of* $1; *otherwise, the player is rewarded* $1. *One readily verifies that the player receives an expected reward of* $0 *under either choice ('red' or 'blue'). Assume now that the same game is played with an urn that contains red and blue balls, but neither the number of balls nor the proportions of their colors are known. Since the distribution of colors is completely ambiguous, the worst scenario is that all balls in the urn have the color named by the player, in which case the penalty of* $1 *arises with certainty. Thus, an ambiguity-averse player is indifferent between any pure strategy (*i.e., *naming red or blue) and paying a fixed amount of* $1. *Assume instead that the player randomly names 'red' (or 'blue') with probability* p *(or* $1 - p$*), and that the (unknown) probability of the drawn ball being red (or blue) is* q *(or* $1 - q$*). In that case, an ambiguity-averse player would then be confronted with*

the following optimization problem:

$$\underset{p\in[0,1]}{maximize} \ \underset{q\in[0,1]}{\min} \ [pq + (1-p)(1-q)] \cdot (-\$1) + [p(1-q) + (1-p)q] \cdot \$1 \,.$$

The inner minimization problem has the parametric optimal solution $q^\star = 1$ if $p > 1/2$; $q^\star \in [0,1]$ if $p = 1/2$; $q^\star = 0$ if $p < 1/2$, which results in an objective value of $-2|p - 1/2|$. Under the optimal choice $p^\star = 1/2$ (i.e., picking a color based on the throw of a fair coin), the player can completely suppress the ambiguity and receive the same expected reward of $\$0$ as in the first urn game where the proportions of colors are known, and the player should consider both games as equivalent.

We will show that under distributional ambiguity, every risk measure satisfying a mild regularity condition is randomization-receptive. We illustrate this with our earlier project selection example.

Example 3 (Project Selection cont'd)**.** *We again consider the project selection problem from Example 1. Assume now that the probabilities p_i^L and p_i^H corresponding to the low and high NPV scenarios of the i-th project are ambiguous, and that they are only known to satisfy*

$$p_i^L = \frac{1}{2} + \min\left\{0.3\mu_i, \frac{1}{2}\right\} z_i \quad and \quad p_i^H = \frac{1}{2} - \min\left\{0.3\mu_i, \frac{1}{2}\right\} z_i \tag{1}$$

for some $z_i \in [-1, +1]$ with $\sum_{i=1}^5 |z_i| \leq 1$, where μ_i denotes project i's expected NPV under the nominal Bernoulli distribution from Table 1. Thus, the possible deviations of the probabilities (p_i^L, p_i^H) from their nominal values $(1/2, 1/2)$ are proportional to the project's expected NPV under the nominal distribution. Assume further that the manager maximizes the worst-case expected NPV over all distributions satisfying (1). In this case, the optimal pure strategy selects project 5, resulting in a worst-case expected value of 0.50, whereas the optimal randomized strategy chooses projects 3, 4 and 5 with probabilities 47%, 30% and 23%, respectively, yielding an almost three times as large worst-case expected value of 1.48.

In this paper, we take a first step towards exploring the implications of randomization in decision problems affected by risk (stochastic programming) and ambiguity (distributionally robust optimization). More specifically, the contributions of this paper can be summarized as follows.

1. We show that in the absence of distributional ambiguity, there is no benefit in randomization when *(i)* both the risk measure and the set of feasible decisions are convex, or if *(ii)* the risk measure is mixture-quasiconcave, irrespective of the structure of the feasible set.

2. We show that although many risk measures are mixture-quasiconcave and thus randomization-proof, several convex risk measures, such as mean (semi-)deviation and mean (semi-)moment measures, can benefit from randomization if the set of feasible decisions is nonconvex.

3. We develop a formal model for decision-making under risk and ambiguity, and we define the notion of an ambiguity averse risk measure which offers an axiomatic justification of the popular distributionally robust optimization paradigm.

4. We show that for any ambiguity averse risk measure satisfying a mild continuity property there exists a decision problem with a nonconvex feasible set in which a randomized strategy strictly dominates all pure strategies.

Our findings imply that a large number of stochastic and distributionally robust optimization problems in classical combinatorial optimization (Cheng et al. 2014, 2016) as well as applications in healthcare (Meng et al. 2015, Mittal et al. 2014), energy (Jiang et al. 2016, Yu 2007), finance (Kellerer et al. 2000) and supply chain management (Chan et al. 2015, Gounaris et al. 2013) may benefit from randomization.

The remainder of the paper is structured as follows. Section 2 briefly surveys the literature on randomization in decision theory and related fields. Sections 3–5 discuss randomized decisions in the absence of distributional ambiguity. After a theoretical characterization of the decision problems that can benefit from randomization in Section 3, Sections 4 and 5 delineate the classes of randomization-receptive and randomization-proof risk measures. Section 6 extends the concept of randomized decision-making to problems with distributional ambiguity. We finally discuss the question of time consistency in Section 7 and conclude in Section 8.

2 A Survey of Relevant Literature

Decision-making under *risk* (where the probability distribution of the outcomes is known) and *ambiguity* (where the probability distribution of the outcomes is itself unknown) has a long history that dates back to the early work of Keynes (1921), Knight (1921) and others. Within this broad domain, our work is most closely related to the fields of stochastic programming and (distributionally) robust optimization, which model, analyze and numerically solve decision problems affected by risk and ambiguity. In these problems, risk is quantified by a suitably chosen risk measure, such

as the expected value, the variance or the (conditional) value-at-risk, and ambiguity is accounted for by evaluating the risk measure over the most adverse distribution from within an ambiguity set. We refer to Ben-Tal et al. (2009), Birge and Louveaux (2000) and Shapiro et al. (2009) for surveys of the extensive stochastic programming and distributionally robust optimization literature. We emphasize that in contrast to our paper, both literature streams have to date solely focused on the choice of pure decisions. A notable exception is the recent research on distributionally robust Markov decision processes, which has shown that randomized strategies can increase the worst-case expected reward in certain classes of dynamic decision problems where the underlying stochastic process is itself only partially known (Tallec 2007, Wiesemann et al. 2014a).

Randomized strategies have a long history in game theory. Nash (1951) has shown that a large class of non-cooperative, simultaneous-move games have an equilibrium if the players are allowed to randomize between their strategies. Moreover, it has been shown that ambiguity-averse players may have additional incentives to randomize (Klibanoff 1996, Lo 1996). More recently, it has been observed that in Stackelberg leader-follower games, the leader can benefit from randomized strategies whenever the follower is oblivious, that is, rather than observing the actually implemented decision, the follower can only observe the probabilities with which different decisions are selected. Stackelberg games with randomized decisions have numerous applications in security games, see An et al. (2016), Bertsimas et al. (2016), Korzhyk et al. (2011) and Mastin et al. (2014).

Randomized decision-making has also received significant attention in the wider economic literature, which offers explanations that broadly fall into three categories (Agranov and Ortoleva 2017): *(i) random utility models*, where decision makers randomize over time as their preferences evolve (see, *e.g.*, Gul and Pesendorfer 2006); *(ii) bounded rationality models*, where agents take the best pure decision over a randomly formed consideration set (see, *e.g.*, Manzini and Mariotti 2014); and *(iii) deliberate randomization models*, where randomized strategies are deliberately taken as they outperform the available pure choices. Since our paper most closely aligns to the third model class, we restrict our review to these models.

Randomization-seeking preferences have been studied under different decision-theoretic frameworks (Eichberger et al. 2016, Eichberger and Kelsey 1996, Saito 2015). While there seems to be a general consensus that decision makers should not benefit from *ex ante randomization*, where the outcome of the randomization device is observed before taking the decision, in some frameworks

decision makers can benefit from *ex post randomization*, where they commit themselves here-and-now to a strategy that depends on the subsequently observed outcome of the randomization device (Eichberger et al. 2016, Saito 2012). In these frameworks, the preference for randomized strategies is envisaged by the axioms, for example through the comonotonic independence axiom in Choquet expected utility theory (Schmeidler 1989) and the uncertainty aversion axiom in maximin expected utility theory (Gilboa and Schmeidler 1989). In contrast, we focus on risk measures that are used in finance, stochastic programming and distributionally robust optimization, and we investigate under which conditions a randomized strategy can outperform all pure choices. It is perhaps surprising that the benefits of randomization have not yet been explored in those literature streams.

On the empirical side, the economics literature remains inconclusive about the prevalence of and the motivation underlying randomized decision-making. A number of lab and field experiments provide evidence that agents deliberately randomize in practice (see, *e.g.*, Sopher and Narramore 2000 and Agranov and Ortoleva 2017), which indicates that models of randomized decision-making may capture an important feature of human behavior. Reasons for randomization include, among others, the desire to minimize the anticipated regret when taking difficult decisions (Dwenger et al. 2016), the presence of incomplete preferences (Cettolin and Riedl 2016) or social aspects that trigger fairness concerns (Kircher et al. 2013). Other studies conclude that agents randomize mistakenly rather than intentionally (Hey and Carbone 1995, Rubinstein 2002), while others even observe a dispreference for randomization (DeJarnette et al. 2015, Dominiak and Schnedler 2011).

Randomized decisions also play a crucial role in the design of algorithms for, among others, data manipulation, computational geometry, number theory and control theory (Campi 2010, Motwani and Raghavan 1995). In those fields, the performance of a deterministic algorithm is typically measured by the worst-case runtime over all instances of a fixed size. Thus, the optimal design of a deterministic algorithm can be viewed as a Stackelberg game between a leader that specifies an algorithm and a follower that selects the most adverse problem instance for the algorithm. In many cases, randomized algorithms, which employ randomness as part of their logic, offer probabilistic runtime guarantees that are competitive with those of the best deterministic algorithms, while at the same time being significantly easier to implement. In the design of error-correcting codes in communication theory, Shannon's noisy channel coding theorem shows that a randomly constructed error-correcting code is essentially as good as the best possible code (Cover and Thomas 1991).

3 Randomization under Stochastic Uncertainty

Assume that uncertainty is modeled via a probability space $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$, where Ω_0 represents the set of outcomes, $\mathcal{F}_0$ the σ-algebra of events and $\mathbb{P}_0$ the probability measure. We denote by $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ the space of all real-valued random variables that are essentially bounded with respect to $\mathbb{P}_0$. We can think of each $X \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ as a financial position or revenue, and we denote by F_X the distribution function of X under $\mathbb{P}_0$ defined through $F_X(x) = \mathbb{P}_0(X \leq x)$ for all $x \in \mathbb{R}$. Throughout the paper we will assume that the probability space $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ is non-atomic.

Definition 1 (Non-Atomic Probability Space)**.** *The probability space $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ is non-atomic if there exists $U_0 \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ that follows a uniform distribution on $[0,1]$.*

Definition 1 is equivalent to the requirement that the probability space does not contain any atoms, that is, any smallest measurable sets with strictly positive probability, see Lemma 6 in Dhaene and Kukush (2011). Working with a non-atomic probability space implies that $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ is rich enough to contain random variables with arbitrary distributions. To formalize this statement, we define $\mathcal{D}$ as the set of all distributions with bounded support, *i.e.*, $\mathcal{D}$ comprises all nondecreasing right-continuous functions $F : \mathbb{R} \to [0,1]$ that attain both 0 and 1.

Lemma 1 (Richness of $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$)**.** *If $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ is non-atomic, then for all $F \in \mathcal{D}$ there exists $X \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ with $F_X = F$.*

Proof. Proof. As $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ is non-atomic, there exists a standard uniformly distributed random variable $U_0 \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$. Choose now an arbitrary $F \in \mathcal{D}$ and denote by F^{-1} the corresponding quantile function defined through $F^{-1}(y) = \inf\{x : F(x) \geq y\}$ for all $y \in \mathbb{R}$. By construction, $X = F^{-1}(U_0)$ is an element of $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ with distribution function $F_X = F$. $\qquad\square \qquad\square$

A risk measure is a functional $\rho_0 : \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0) \to \mathbb{R}$ that assigns each financial position a risk index. We will focus exclusively on law invariant risk measures, which quantify the risk of a position X solely on the basis of its distribution function F_X.

Definition 2 (Law Invariance)**.** *A risk measure $\rho_0 : \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0) \to \mathbb{R}$ is called law invariant if it satisfies the following condition for all random variables $X, Y \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$:*

Law invariance: *If $F_X = F_Y$, then $\rho_0(X) = \rho_0(Y)$.*

Any law invariant risk measure ρ_0 on a non-atomic probability space corresponds to a real-valued functional ϱ_0 on the space $\mathcal{D}$ of all distribution functions with bounded support.

Proposition 1 (Existence and Uniqueness of ϱ_0). *If $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ is non-atomic and ρ_0 is a law invariant risk measure on $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$, then there is a unique functional ϱ_0 on $\mathcal{D}$ with*

$$\rho_0(X) = \varrho_0(F_X) \quad \forall X \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0). \tag{2}$$

Proof. Proof. The existence of ϱ_0 follows immediately from the law invariance of ρ_0. As for the uniqueness, assume to the contrary that (2) is satisfied by two functionals ϱ_0, ϱ_0' such that there is $F \in \mathcal{D}$ with $\varrho_0(F) \neq \varrho_0'(F)$. Lemma 1 then implies that $F = F_X$ for some $X \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$. We then obtain $\rho_0(X) = \varrho_0(F_X) = \varrho_0(F) \neq \varrho_0'(F) = \varrho_0'(F_X) = \rho_0(X)$, which is a contradiction. $\square$ $\square$

Proposition 1 shows that the assumptions of non-atomicity and law invariance allow us to conclude that the distribution mapping ϱ_0 associated with the risk measure ρ_0 is defined over all distribution functions in $\mathcal{D}$. In other words, under these two assumptions the decision maker is able to quantify the risk of every random outcome associated with any possible probability distribution $F \in \mathcal{D}$, not just those random outcomes that correspond to feasible decisions. Below we will use this insight to assign a unique risk to randomized decisions, some of which may correspond to random outcomes whose probability distributions are not attained by any feasible deterministic decision. In principle, we could drop the assumption of non-atomicity as long as ϱ_0 is naturally defined over all distribution functions in $\mathcal{D}$ (as is the case for any commonly used law invariant risk measure ρ_0 from the literature). To keep the discussion general, we prefer to postulate non-atomicity explicitly.

Several commonly encountered properties of risk measures are reviewed below.

Definition 3 (Monetary Risk Measure). *A risk measure $\rho_0 : \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0) \to \mathbb{R}$ is called monetary if it satisfies the following conditions for all random variables $X, Y \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$:*

Monotonicity: *If $X \geq Y$ $\mathbb{P}_0$-almost surely, then $\rho_0(X) \leq \rho_0(Y)$.*

Translation invariance: *If $t \in \mathbb{R}$, then $\rho_0(X + t) = \rho_0(X) - t$.*

Definition 4 (Convex Risk Measure). *A risk measure $\rho_0 : \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0) \to \mathbb{R}$ is called convex if it satisfies the following condition for all random variables $X, Y \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$:*

Convexity: *If $\theta \in [0, 1]$, then $\rho_0(\theta X + (1 - \theta)Y) \leq \theta \rho_0(X) + (1 - \theta)\rho_0(Y)$.*

We alert the reader that we do not implicitly assume convex risk measures to be monetary as is frequently done in the literature. The explicit distinction between convex and convex monetary risk measures will facilitate a more concise description of the main results in this paper.

Definition 5 (Coherent Risk Measure). *A risk measure $\rho_0 : \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0) \to \mathbb{R}$ is called coherent if it is convex and monetary, and if it additionally satisfies the following condition for all random variables $X \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$:*

Scale invariance: *If $\lambda \geq 0$, then $\rho_0(\lambda X) = \lambda \rho_0(X)$.*

The class of coherent risk measures has been popularized by Artzner et al. (1999). However, many widely-used risk measures are not scale invariant and thus fail to be coherent.

Consider now a decision maker who needs to select a financial position from a feasible set $\mathcal{X}_0 \subseteq \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$. If the disutility of each position is quantified using a risk measure ρ_0, then the decision maker may solve the following optimization problem in pure strategies.

$$\text{(Pure Strategy Problem)} \qquad \underset{X \in \mathcal{X}_0}{\text{minimize}} \qquad \rho_0(X) \tag{3}$$

Alternatively, the decision maker might optimize over a class of randomized strategies by using a randomization device that is independent of all $X \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$. We assume henceforth that this randomization device generates samples from the uniform distribution $\mathbb{U}$ on the interval $[0, 1]$ equipped with the Borel σ-algebra $\mathcal{B}_{[0,1]}$. To formalize the optimization problem over randomized strategies, we thus introduce an augmented probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\Omega = \Omega_0 \times [0, 1]$, $\mathcal{F} = \mathcal{F}_0 \otimes \mathcal{B}_{[0,1]}$ and $\mathbb{P} = \mathbb{P}_0 \times \mathbb{U}$. Elements of Ω_0 will usually be denoted by ω_0, while elements of Ω will be denoted by $\omega = (\omega_0, u)$. For $X \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P})$, we denote by F_X the distribution function of X under the product measure $\mathbb{P}_0 \times \mathbb{U}$. We further define the feasible set of randomized strategies as

$$\mathcal{X} = \{X \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P}) \ : \ X(\cdot, u) \in \mathcal{X}_0 \ \forall u \in [0, 1]\}.$$

Note that any randomized strategy $X \in \mathcal{X}$ can be viewed as a decision rule that assigns each possible outcome $u \in [0, 1]$ of the randomization device a pure strategy $X(\cdot, u) \in \mathcal{X}_0$. Moreover, the space of pure strategies $\mathcal{X}_0$ can naturally be embedded into $\mathcal{X}$ by identifying the elements of $\mathcal{X}_0$ with those elements of $\mathcal{X}$ that are constant in u. We will henceforth refer to such random variables as being independent of the randomization device.

Remark 1 (Expressiveness of the Randomization Device). *Focusing on a randomization device that generates uniform samples from $[0,1]$ is non-restrictive as the pure strategies $X \in \mathcal{X}_0$ often admit a d-dimensional parameterization for some $d \in \mathbb{N}$. For instance, in portfolio selection the pure strategies are parameterized by a portfolio weight vector in $\mathbb{R}^d$, where d equals the number of assets. Then, all randomized portfolio strategies are implementable if we can sample from general distributions on $\mathbb{R}^d$. A uniform randomization device indeed enables us to do so:*

(i) *If the samples of the randomization device are represented in binary as $u = 0.u_1 u_2 \ldots$, then all digits u_i, $i \in \mathbb{N}$, follow independent Bernoulli random variables that are equal to 0 or 1 with probability $\frac{1}{2}$ each. The binary numbers $0.u_i\, u_{i+d}\, u_{i+2d} \ldots$ for $i = 1, \ldots, d$ constructed from u are thus independent and uniformly distributed on $[0,1]$. If we can sample uniformly from the interval $[0,1]$, we can therefore also sample uniformly from the hypercube $[0,1]^d$.*

(ii) *Uniform samples from $[0,1]^d$ can systematically be transformed to samples from an arbitrary distribution on $\mathbb{R}^d$ via the inverse Rosenblatt transformation (Rosenblatt 1952). Thus, if we can sample uniformly from $[0,1]^d$, then we can sample from any multivariate distribution.*

In order to be able to quantify the risk of a randomized strategy, we need to extend the risk measure ρ_0 on $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ to another risk measure ρ on $\mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P})$. We will show below in Proposition 2 that this extension is unique if we insist that ρ must be law invariant and coincide with ρ_0 on the subspace of all random variables that are independent of the randomization device. By slight abuse of notation, for $Z \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ we use $\rho(Z)$ to denote the risk of the natural extension of the random variable Z to the space $\mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P})$, where Z is constant in u and thus is independent of the randomization device.

Proposition 2 (Law Invariant Extension). *If $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ is non-atomic and ρ_0 is a law invariant risk measure on $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$, then the unique extension of ρ_0 to a law invariant risk measure ρ on $\mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P})$ is given by $\rho(X) = \varrho_0(F_X)$ for all $X \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P})$.*

Proof. Proof. Let ρ be any law invariant risk measure that extends ρ_0 to $\mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P})$, and note that $(\Omega, \mathcal{F}, \mathbb{P})$ is non-atomic as it is an extension of a non-atomic probability space. Proposition 1 thus ensures the existence of a mapping $\varrho : \mathcal{D} \to \mathbb{R}$ with $\rho(X) = \varrho(F_X)$ for all randomized strategies $X \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P})$. Hence, for all $X \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$, we have $\varrho(F_X) = \rho(X) = \rho_0(X) = \varrho_0(F_X)$, where the second equality follows from the requirement that ρ must coincide with ρ_0 on the subspace

of all random variables that are independent of the randomization device, while the third equality holds by the definition of ϱ_0. Lemma 1 then implies that $\varrho(F) = \varrho_0(F)$ for all $F \in \mathcal{D}$. Thus, we indeed have $\rho(X) = \varrho(F_X) = \varrho_0(F_X)$ for all randomized strategies $X \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P})$, and ρ is the only law invariant risk measure that extends ρ_0 to $\mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P})$. □ □

We note that the assumption of law invariance implies that the decision maker agrees with the reduction of compound lotteries axiom (Samuelson 1952) over the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and therefore measures risk solely based on the distribution function of the random outcome. In other words, she is indifferent between the uncertainty due to a selected decision and the uncertainty caused by the randomization process. This could possibly be problematic when the outcome u of the randomization device is observed earlier than the outcome ω_0 of the implemented decision. We will discuss this issue in Section 7.

We can now formulate the counterpart of (3) that optimizes over all randomized strategies as

$$\text{(Randomized Strategy Problem)} \qquad \underset{X \in \mathcal{X}}{\text{minimize}} \quad \rho(X). \qquad (4)$$

It is clear that the optimal value of (4) is never worse (that is, never strictly larger) than the optimal value of (3). Indeed, problem (3) can be recovered from (4) by restricting $\mathcal{X}$ to those random variables that are independent of the randomization device. Instead, there are interesting situations where the best randomized strategy is strictly preferred to the best pure one, that is, where the optimal value of (4) is strictly smaller than the optimal value of (3), see, *e.g.*, Example 1. Popular law invariant risk measures that can benefit from randomization are presented in Section 4.

In the following we demonstrate that there is no benefit in adopting a randomized strategy when either (i) ρ_0 is convex and $\mathcal{X}_0$ is a convex feasible set, or (ii) ρ_0 is mixture-quasiconcave.

Definition 6 (Mixture-Quasiconcavity). *A law invariant risk measure ρ_0 is called mixture-quasi-concave if the corresponding mapping ϱ_0, defined on the set of distribution functions, is quasicon-cave, that is, if ϱ_0 satisfies the following property for all $X, Y \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$:*

Mixture-quasiconcavity: *If $\theta \in [0, 1]$, then $\varrho_0(\theta F_X + (1 - \theta) F_Y) \geq \min\{\varrho_0(F_X), \varrho_0(F_Y)\}$.*

To simplify the subsequent arguments, we introduce the following notational shorthand.

Definition 7. *For any $X, Y \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ and $\theta \in [0, 1]$, we define the random variable $X \oplus_\theta Y \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P})$ through $X \oplus_\theta Y(\omega_0, u) = X(\omega_0)$ if $u \leq \theta$ and $X \oplus_\theta Y(\omega_0, u) = Y(\omega_0)$ if $u > \theta$.*

Remark 2. *The distribution function F_X of a random variable $X \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P})$ is given by*

$$F_X(x) = \int_\Omega \mathbf{1}_{\{X(\omega) \leq x\}} \, \mathbb{P}(\mathrm{d}\omega) = \int_0^1 \int_{\Omega_0} \mathbf{1}_{\{X(\omega_0, u) \leq x\}} \, \mathbb{P}_0(\mathrm{d}\omega_0) \, \mathrm{d}u = \int_0^1 F_{X(\cdot, u)}(x) \, \mathrm{d}u.$$

Thus, F_X corresponds to an infinite convex combination of the distribution functions associated with the random variables $X(\cdot, u) \in \mathcal{X}_0$ for $u \in [0, 1]$. In particular, the distribution function of the mixture $X \oplus_\theta Y$ is given by the convex combination $\theta F_X + (1 - \theta) F_Y$.

We first demonstrate that under a convex risk measure the convex combination of two pure positions is always weakly preferred to the mixture of these positions when the mixture probabilities are used as convex weights.

Theorem 1. *If ρ is the unique law invariant extension of a convex law invariant risk measure ρ_0, then we have for any $X, Y \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ and $\theta \in [0, 1]$ that $\rho(X \oplus_\theta Y) \geq \rho(\theta X + (1 - \theta) Y)$.*

Proof. Proof. For any $s \in [0, 1]$, we define the auxiliary random variable $X \oplus_\theta^s Y \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P})$ through $X \oplus_\theta^s Y(\omega_0, u) = X(\omega_0)$ if $(u + s \bmod 1) \in [0, \theta]$ and $X \oplus_\theta^s Y(\omega_0, u) = Y(\omega_0)$ if $(u + s \bmod 1) \in (\theta, 1]$. Note that, for any $s \in [0, 1]$, the random variable $X \oplus_\theta^s Y$ mixes X and Y with probabilities θ and $1 - \theta$, respectively. Thus, $X \oplus_\theta^s Y$ has the same law as $X \oplus_\theta Y$ irrespective of s.

By construction, integrating $X \oplus_\theta^s Y$ scenariowise over $s \in [0, 1]$ yields the convex combination $\theta X + (1 - \theta) Y$, which is independent of the randomization device. To see this, observe that

$$\int_0^1 X \oplus_\theta^s Y(\omega_0, u) \, \mathrm{d}s = \int_0^1 X(\omega_0) \mathbf{1}_{\{(u + s \bmod 1) \in [0, \theta]\}} \, \mathrm{d}s + \int_0^1 Y(\omega_0) \mathbf{1}_{\{(u + s \bmod 1) \in (\theta, 1]\}} \, \mathrm{d}s$$

$$= \theta X(\omega_0) + (1 - \theta) Y(\omega_0).$$

Since ρ inherits convexity from ρ_0, which we will prove below, Jensen's inequality implies

$$\rho(\theta X + (1 - \theta) Y) = \rho \left(\int_0^1 X \oplus_\theta^s Y \, \mathrm{d}s \right)$$

$$\leq \int_0^1 \rho \left(X \oplus_\theta^s Y \right) \, \mathrm{d}s = \int_0^1 \rho \left(X \oplus_\theta Y \right) \, \mathrm{d}s = \rho(X \oplus_\theta Y),$$

where the second equality follows from the law invariance of ρ.

We are left with demonstrating that ρ inherits convexity from ρ_0. To this end, let (X, Y) be any pair of random variables in $\mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P})$ with the joint distribution function $F_{X,Y} : \mathbb{R}^2 \to [0, 1]$. Since $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ is non-atomic, it contains a uniformly distributed random variable U_0 which we

can use to create a pair of random variables (X_0, Y_0) in $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ whose distribution function equals $F_{X,Y}$, see Remark 1. For any $\theta \in [0, 1]$, we then obtain that

$$\rho(\theta X + (1 - \theta)Y) \;=\; \rho_0(\theta X_0 + (1 - \theta)Y_0) \;\leq\; \theta \rho_0(X_0) + (1 - \theta)\rho_0(Y_0) \;=\; \theta \rho(X) + (1 - \theta)\rho(Y),$$

where the equalities hold since ρ is the law invariant extension of ρ_0, and the inequality follows from the assumed convexity of ρ_0. This completes the proof. $\qquad \square \qquad\qquad \square$

Unless Jensen's inequality in the proof of Theorem 1 is binding, the risk of the mixture $X \oplus_\theta Y$ is strictly larger than the risk of the convex combination $\theta X + (1 - \theta)Y$. This is the case, for instance, if $X \neq Y$, $\theta \in (0, 1)$ and ρ is strictly convex.

Using a similar strategy as in the proof of Theorem 1, we can further show that randomization is never beneficial if ρ_0 is convex and $\mathcal{X}_0$ constitutes a convex feasible set.

Theorem 2. *Assume that ρ is the unique law invariant extension of the convex law invariant risk measure ρ_0 and that $\mathcal{X}_0$ is convex. Then, problems (3) and (4) attain the same optimal value.*

Proof. Proof. As (3) is a restriction of (4), it suffices to show that the optimal value of (3) is never larger than that of (4). To this end, we define $T_s : \Omega \to \Omega$ through $T_s(\omega_0, u) = (\omega_0, (u + s \bmod 1))$ for all $(\omega_0, u) \in \Omega$. By construction, T_s is a measurable bijection. A direct calculation shows that $\mathbb{P}(T_s^{-1}(A)) = \mathbb{P}(A)$ for every $A \in \mathcal{F}$, and thus T_s is in fact a measure preserving transformation.

Choose now any $X \in \mathcal{X}$, and note that $X \circ T_s \in \mathcal{X}$ has the same law as X. Next, define $Y \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P})$ via

$$Y(\omega_0, u) = \int_0^1 X \circ T_s(\omega_0, u)\, \mathrm{d}s = \int_0^1 X(\omega_0, (u + s \bmod 1))\, \mathrm{d}s \quad \forall (\omega_0, u) \in \Omega.$$

By construction, $Y(\omega_0, u)$ is constant in u implying that Y is independent of the randomization device. Moreover, for all $u \in [0, 1]$, we have that $Y(\cdot, u) \in \mathcal{X}_0$ because $\mathcal{X}_0$ is convex and $Y(\cdot, u)$ can be viewed as an infinite convex combination of the random variables $X(\cdot, (u + s \bmod 1)) \in \mathcal{X}_0$ across all $s \in [0, 1]$. As ρ inherits convexity from ρ_0 (see proof of Theorem 1), Jensen's inequality implies

$$\rho(Y) = \rho\left(\int_0^1 X \circ T_s\, \mathrm{d}s\right) \leq \int_0^1 \rho(X \circ T_s)\, \mathrm{d}s = \int_0^1 \rho(X)\, \mathrm{d}s = \rho(X),$$

where the second equality follows from the law invariance of ρ. As any $X \in \mathcal{X}$ gives rise to some $Y \in \mathcal{X}_0$ with weakly smaller risk, the optimal value of (3) is never larger than that of (4). $\square \ \square$

In the following we show that there is no benefit in randomization if the risk measure ρ_0 is mixture-quasiconcave, even if $\mathcal{X}_0$ is nonconvex.

Theorem 3. *Assume that ρ is the unique law invariant extension of a law invariant risk measure ρ_0 that is mixture-quasiconcave. Then, problems (3) and (4) attain the same optimal values.*

Proof. Proof. As (3) is a restriction of (4), it suffices to show that the optimal value of (3) is never larger than that of (4). To this end, we recall that the law invariance of ρ implies that there exists a functional ϱ from the space of distribution functions of all bounded random variables to $\mathbb{R}$ with the property that $\rho(X) = \varrho(F_X)$ for all $X \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P})$. Using similar arguments as in the proof of Theorem 1, one can show that ρ inherits mixture-quasiconcavity from ρ_0, which means that ϱ is quasiconcave. Select now any $X \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P})$ and recall from Remark 2 that the distribution of X can be expressed as the infinite convex combination $F_X(x) = \int_0^1 F_{X(\cdot,u)} \, \mathrm{d}u$. Hence, we have

$$\rho(X) = \varrho(F_X) = \varrho \left(\int_0^1 F_{X(\cdot,u)} \, \mathrm{d}u \right) \geq \inf_{u \in [0,1]} \varrho \left(F_{X(\cdot,u)} \right) \geq \inf_{X_0 \in \mathcal{X}_0} \rho_0(X_0),$$

where the first inequality follows from the quasiconcavity of ϱ, while the second inequality holds because $X(\cdot, u) \in \mathcal{X}_0$ for all $u \in [0,1]$. As the risk of any $X \in \mathcal{X}$ is bounded below by the optimal value of (3), the claim follows. $\qquad\square$ $\qquad\qquad\square$

Theorem 3 provides the justification for the following definition.

Definition 8 (Randomization-Proofness and -Receptiveness)**.** *A law invariant risk measure ρ_0 is called randomization-proof if its law invariant extension ρ satisfies $\rho(X \oplus_\theta Y) \geq \min\{\rho(X), \rho(Y)\}$ for all $\theta \in [0,1]$ and all $X, Y \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$. Otherwise, ρ_0 is called randomization-receptive.*

Theorem 3 shows that mixture-quasiconcavity implies randomization-proofness in the sense of Definition 8. If, on the other hand, ρ_0 is randomization-proof, then its extension ρ must satisfy $\rho(X \oplus_\theta Y) \geq \min\{\rho_0(X), \rho_0(Y)\}$ for all $\theta \in [0,1]$ and all $X, Y \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$. Proposition 2 and Remark 2 then imply that $\varrho_0(\theta F_X + (1-\theta)F_Y) = \rho(X \oplus_\theta Y) \geq \min\{\rho_0(X), \rho_0(Y)\} = \min\{\varrho_0(F_X), \varrho_0(F_Y))$, that is, ρ_0 is mixture-quasiconcave. We thus conclude that a risk measure is randomization-proof if and only if it is mixture-quasiconcave.

Remark 3. *Randomization-receptiveness (*i.e., ρ_0 not being mixture-quasiconcave) is closely related to but distinct from the axiom of uncertainty aversion in maximin expected utility theory*

(Gilboa and Schmeidler 1989). Translated into our context, uncertainty aversion (as initially introduced in Schmeidler 1989) requires the risk measure ρ_0 to be mixture-quasiconvex, that is, the law invariant extension ρ has to satisfy $\rho(X \oplus_\theta Y) \leq \max\{\rho(X), \rho(Y)\}$ for all $\theta \in [0,1]$ and all $X, Y \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$. Uncertainty aversion states that mixing risky positions cannot increase the risk, whereas randomization-receptiveness implies that there are mixtures of risky positions that strictly decrease the risk.

4 Randomization-Receptive Risk Measures

We now identify 4 classes of randomization-receptive risk measures. All of these risk measures are law invariant, translation invariant and convex, but some are not monotonic and/or scale invariant.

Mean Moment of Order p. The order-p mean moment risk measure $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \alpha\mathbb{E}_{\mathbb{P}_0}|\mathbb{E}_{\mathbb{P}_0}[X] - X|^p$ with $p \geq 1$ and $\alpha > 0$ is law invariant, translation invariant and convex, but it fails to be monotonic or scale invariant in general.

We will show in Section 5.4 that order-2 mean moment (*i.e.*, mean variance) risk measures are randomization-proof. For every other $p \geq 1$ and any $\alpha > 0$, however, these risk measures are randomization-receptive. Indeed, consider the following two random variables:

$$X \sim \begin{cases} 1 & \text{w.p. } 1/2, \\ -1 & \text{w.p. } 1/2, \end{cases} \qquad Y \sim \begin{cases} 0.91 & \text{w.p. } 2/3, \\ -1.11 & \text{w.p. } 1/3. \end{cases} \tag{5}$$

For $p = 4$ and $\alpha = 1$, we obtain $\rho(X) = 1$, $\rho(Y) = 0.997$ and $\rho(X \oplus_{1/2} Y) = 0.934$, that is, $\rho(X \oplus_{1/2} Y) < \min\{\rho(X), \rho(Y)\}$. Indeed, we have the following general result.

Theorem 4. *The order-p mean moment risk measure $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \alpha\mathbb{E}_{\mathbb{P}_0}|\mathbb{E}_{\mathbb{P}_0}[X] - X|^p$ is randomization-receptive for every order $p \in [1, 2) \cup (2, \infty)$ and for any weighting $\alpha > 0$.*

We relegate the proof of Theorem 4 to the appendix. The above example (5) with $p = 4$ and $\alpha = 1$ follows the construction proposed in that proof if we fix $\epsilon = 0.238$.

Mean Deviation of Order p. The order-p mean deviation risk measure $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \alpha(\mathbb{E}_{\mathbb{P}_0}|\mathbb{E}_{\mathbb{P}_0}[X] - X|^p)^{1/p}$ with $p \geq 1$ and $\alpha > 0$ is law invariant, translation invariant, convex and scale invariant, but it typically fails to be monotonic (Shapiro et al. 2009, Example 6.19).

We will show in Section 5.4 that order-2 mean deviation risk measures are randomization-proof. For every other p and any $\alpha > 0$, however, these risk measures are randomization-receptive. To see this, consider the following two random variables:

$$X \sim \begin{cases} 1 & \text{w.p. } 1/2, \\ -1 & \text{w.p. } 1/2, \end{cases} \qquad Y \sim \begin{cases} 0.88 & \text{w.p. } 2/3, \\ -1.32 & \text{w.p. } 1/3. \end{cases} \tag{6}$$

For $p = 4$ and $\alpha = 1$, we obtain $\rho(X) = 1$, $\rho(Y) = 1.001$ and $\rho(X \oplus_{1/2} Y) = 0.992$, which implies that $\rho(X \oplus_{1/2} Y) < \min\{\rho(X), \rho(Y)\}$. Indeed, we have the following general result.

Theorem 5. *The order-p mean deviation risk measure $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \alpha(\mathbb{E}_{\mathbb{P}_0}|\mathbb{E}_{\mathbb{P}_0}[X] - X|^p)^{1/p}$ is randomization-receptive for every order $p \in [1, 2) \cup (2, \infty)$ and for any weighting $\alpha > 0$.*

The proof of Theorem 5 can be found in the appendix. The above example (6) with $p = 4$ and $\alpha = 1$ follows the construction proposed in that proof if we fix $\epsilon = 0.146$.

Mean Semi-Moment of Order p. The order-p mean semi-moment risk measure $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \alpha\mathbb{E}_{\mathbb{P}_0}[\mathbb{E}_{\mathbb{P}_0}[X] - X]_+^p$ with $p \geq 1$ and $\alpha > 0$ is law invariant, translation invariant and convex, but it is neither monotonic nor scale invariant in general. For $p = 2$, it reduces to the mean semi-variance risk measure, which we have shown to be randomization-receptive in Example 1.

Order-p mean semi-moment risk measures are randomization-receptive for every $p > 1$ and any $\alpha > 0$. To see this, consider the following two random variables:

$$X \sim \begin{cases} \sqrt{2} & \text{w.p. } 1/2, \\ -\sqrt{2} & \text{w.p. } 1/2, \end{cases} \qquad Y \sim \begin{cases} -0.97 & \text{w.p. } 2/3, \\ -1.05 & \text{w.p. } 1/3. \end{cases} \tag{7}$$

For $p = 2$ and $\alpha = 1$, we obtain $\rho(X) = 1$, $\rho(Y) = 0.998$ and $\rho(X \oplus_{1/2} Y) = 0.833$, which implies that $\rho(X \oplus_{1/2} Y) < \min\{\rho(X), \rho(Y)\}$. Indeed, we have the following general result.

Theorem 6. *The order-p mean semi-moment risk measure $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \alpha\mathbb{E}_{\mathbb{P}_0}[\mathbb{E}_{\mathbb{P}_0}[X] - X]_+^p$ is randomization-receptive for every order $p \geq 1$ and for any weighting $\alpha > 0$.*

The proof of Theorem 6 is relegated to the appendix. The above example (7) with $p = 2$ and $\alpha = 1$ follows the construction proposed in that proof if we fix $\epsilon = 0.999$.

Mean Semi-Deviation of Order p. The order-p mean semi-deviation risk measure $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \alpha(\mathbb{E}_{\mathbb{P}_0}[\mathbb{E}_{\mathbb{P}_0}[X] - X]_+^p)^{1/p}$ with $p \geq 1$ and $\alpha > 0$ is translation invariant, convex and scale invariant. Additionally, it is monotonic (and thus coherent) when $\alpha \in [0, 1]$, see Example 6.20 in Shapiro et al. (2009).

Order-p mean semi-deviation risk measures are randomization-receptive for every $p > 1$ and any $\alpha > 0$. To see this, consider the following two random variables:

$$X \sim \begin{cases} \sqrt{2} & \text{w.p. } 1/2, \\ -\sqrt{2} & \text{w.p. } 1/2, \end{cases} \qquad Y \sim \begin{cases} -0.22 & \text{w.p. } 2/3, \\ -1.31 & \text{w.p. } 1/3. \end{cases} \tag{8}$$

For $p = 2$ and $\alpha = 1$, we obtain $\rho(X) = 1$, $\rho(Y) = 1.029$ and $\rho(X \oplus_{1/2} Y) = 0.990$, which implies that $\rho(X \oplus_{1/2} Y) < \min\{\rho(X), \rho(Y)\}$. Indeed, we have the following general result.

Theorem 7. *The order-p mean semi-deviation risk measure* $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \alpha(\mathbb{E}_{\mathbb{P}_0}[\mathbb{E}_{\mathbb{P}_0}[X] - X]_+^p)^{1/p}$ *is randomization-receptive for every order* $p \geq 1$ *and for any weighting* $\alpha > 0$.

The proof of Theorem 7 can be found in the appendix. The above example (8) with $p = 2$ and $\alpha = 1$ follows the construction proposed in that proof if we fix $\epsilon = 0.581$.

5 Randomization-Proof Risk Measures

We now present commonly used law invariant risk measures that are randomization-proof. Specifically, we discuss spectral risk measures in Section 5.1, divergence risk measures in Section 5.2 and shortfall risk measures in Section 5.3. Other randomization-proof risk measures that fit into none of these categories are discussed in Section 5.4.

5.1 Spectral Risk Measures

Spectral risk measures[3] can be defined axiomatically as convex law invariant monetary risk measures that are also comonotonic. The latter property captures the intuitive idea that if two positions X and Y are comonotone, then X cannot serve as a hedge for Y, and thus the risks of X and Y should add up.

---

[3]The class of spectral risk measures coincides with the class of concave distortion risk measures, see, *e.g.*, Föllmer and Schied (2011).

Definition 9 (Spectral Risk Measure). *A monetary risk measure $\rho_0 : \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0) \to \mathbb{R}$ is called spectral if it is convex and law invariant and satisfies the following condition for all $X, Y \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$:*

Comonotonicity: *If X and Y are comonotone, that is, if $(X(\omega_0) - X(\omega_0'))(Y(\omega_0) - Y(\omega_0')) \geq 0$ for all $\omega_0, \omega_0' \in \Omega_0$, then $\rho_0(X + Y) = \rho_0(X) + \rho_0(Y)$.*

Lemma 2. *For convex risk measures, comonotonicity implies scale invariance.*

Proof. Proof. Let $\rho_0 : \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0) \to \mathbb{R}$ be a comonotonic convex risk measure. Note first that $\rho_0(0) = \rho_0(0+0) = 2\rho_0(0)$ by comonotonicity. Thus, we find $\rho_0(0) = 0$. Moreover, for any $i \in \mathbb{N}$ we find $\rho_0(iX) = \rho_0(X+(i-1)X) = \rho_0(X)+\rho_0((i-1)X) = 2\rho_0(X)+\rho_0((i-2)X) = \cdots = i\rho_0(X)$, where the second and third equality etc. follow again from comonotonicity. As the function $f(\lambda) = \rho_0(\lambda X)$ inherits convexity from ρ_0 for every fixed X and as $f(i) = if(1)$ for every $i \in \mathbb{N} \cup \{0\}$, we may conclude that $f(\lambda) = \lambda f(1)$ for every $\lambda \geq 0$. Thus, ρ_0 is scale invariant. $\square$ $\square$

Lemma 2 implies the well-known fact that every spectral risk measure is coherent.

Example 4 (Spectral Risk Measures). *We list examples of spectral risk measures.*

(i) *Expected loss:* $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] = -\int_{\Omega_0} X(\omega_0)\,\mathbb{P}_0(\mathrm{d}\omega_0)$. *Note that the expected value is additive for all pairs and, in particular, for all pairs of comonotone random variables.*

(ii) *Essential supremum:* $\rho_0(X) = \mathrm{ess\,sup}(-X) = \inf_{s \in \mathbb{R}}\{s : \mathbb{P}_0(-X \leq s) = 1\}$.

(iii) *Conditional value-at-risk:* $\rho_0(X) = \mathrm{CV@R}_\alpha(X) = \inf_{s \in \mathbb{R}} s + \frac{1}{1-\alpha}\mathbb{E}_{\mathbb{P}_0}[-X-s]_+$ *with $\alpha \in [0,1)$.* *Note that the conditional value-at-risk reduces to the expected value for $\alpha = 0$ and to the essential supremum for $\alpha \uparrow 1$. Thus, it is common to define $\mathrm{CV@R}_1(X) = \mathrm{ess\,sup}(-X)$.*[4]

(iv) *Mixture of conditional value-at-risk:* $\rho_0(X) = \int_0^1 \mathrm{CV@R}_\alpha(X)\,\mu(\mathrm{d}\alpha)$ *with μ a Borel probability measure on $[0,1]$. In fact, every spectral risk measure can be represented as a mixture of conditional value-at-risk (Kusuoka 2001).*

Theorem 8. *Every spectral risk measure $\rho_0 : \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0) \to \mathbb{R}$ is randomization-proof.*

---

[4]The conditional value-at-risk can even strictly discourage randomization. To see this, set $X = -1$ and assume that the random variable Y takes the values 0 and -2 with probabilities 0.75 and 0.25, respectively. In this case, we have $\mathrm{CV@R}_{50\%}(X \oplus_{1/2} Y) = 1.25$ even though $\mathrm{CV@R}_{50\%}(X) = \mathrm{CV@R}_{50\%}(Y) = 1$.

Proof. Proof. Select any $X, Y \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$, and denote by F_X^{-1} and F_Y^{-1} their respective quantile functions. As $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ is non-atomic, there exists a standard uniformly distributed random variable $U_0 \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$, which we use to define the random variables $X' = F_X^{-1}(U_0)$ and $Y' = F_Y^{-1}(U_0)$. By construction, X' and Y' have the same distributions as X and Y, respectively. As distribution and quantile functions are nondecreasing, we find that $X'(\omega_0) \geq X'(\omega_0')$ is equivalent to $U_0(\omega_0) \geq U_0(\omega_0')$, which in turn is equivalent to $Y'(\omega_0) \geq Y'(\omega_0')$ for any $\omega_0, \omega_0' \in \Omega_0$. Thus, X' and Y' are comonotone. Next, choose any $\theta \in [0, 1]$. Remark 2 implies that $X \oplus_\theta Y$ and $X' \oplus_\theta Y'$ share the same distribution. If ρ is the unique law invariant extension of ρ_0, then

$$
\begin{aligned}
\rho(X \oplus_\theta Y) &= \rho(X' \oplus_\theta Y') && \text{(by law invariance)} \\
&\geq \rho(\theta X' + (1 - \theta) Y') && \text{(by Theorem 1)} \\
&= \rho(\theta X') + \rho((1 - \theta) Y') && \text{(by comonotonicity)} \\
&= \theta \rho(X') + (1 - \theta) \rho(Y') && \text{(by scale invariance)} \\
&\geq \min\{\rho(X'), \rho(Y')\} && \text{(since } \theta \in [0, 1]) \\
&= \min\{\rho(X), \rho(Y)\} && \text{(by law invariance).}
\end{aligned}
$$

The second equality in the above derivation holds because ρ inherits comonotonicity from ρ_0. Thus, ρ_0 is randomization-proof. $\qquad\square$ $\qquad\qquad\qquad\qquad\qquad\square$

As we have seen in Section 4, the mean semi-deviation is randomization-receptive despite being coherent and law invariant. This is only possible because it is not comonotonic.

5.2 Divergence Risk Measures

The divergence risk of a position X is defined as the worst-case expectation of the corresponding loss $-X$ across all probability measures $\mathbb{Q}$ that are absolutely continuous with respect to $\mathbb{P}_0$, where any $\mathbb{Q} \neq \mathbb{P}_0$ is assigned a divergence penalty.

Definition 10 (Divergence Risk Measure)**.** *A mapping* $\rho_0 : \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0) \to \mathbb{R}$ *is a divergence risk measure if there exists a convex lower semicontinuous function* $\varphi : \mathbb{R} \to [0, \infty]$ *with* $\varphi(1) = 0$ *and if*

$$
\rho_0(X) = \sup_{\mathbb{Q} \ll \mathbb{P}_0} \mathbb{E}_\mathbb{Q}[-X] - I_\varphi(\mathbb{Q}, \mathbb{P}_0),
$$

21

where $I_\varphi(\mathbb{Q}, \mathbb{P}_0) = \mathbb{E}_{\mathbb{P}_0}[\varphi(\frac{d\mathbb{Q}}{d\mathbb{P}_0})]$ *is the φ-divergence of the probability measure $\mathbb{Q}$ with respect to $\mathbb{P}_0$,*
while $\mathbb{Q} \ll \mathbb{P}_0$ indicates that $\mathbb{Q}$ admits a Radon-Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}_0}$ with respect to $\mathbb{P}_0$.

It follows directly from their definition that divergence risk measures are monetary and convex. The following theorem offers an intuitive physical interpretation of divergence risk measures.

Theorem 9 (Optimized Certainty Equivalent). *If $\rho_0 : \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0) \to \mathbb{R}$ is the divergence risk measure generated by a convex lower semicontinuous function $\varphi : \mathbb{R} \to [0, \infty]$ with $\varphi(1) = 0$, then*

$$\rho_0(X) = -\sup_{s \in \mathbb{R}} s - \mathbb{E}_{\mathbb{P}_0}[\varphi^*(s - X)],$$

where $\varphi^*(s) = \sup_{t \in \mathbb{R}} st - \varphi(t)$ is the convex conjugate of φ.

Proof. Proof. See Theorem 4.2 in Ben-Tal and Teboulle (2007). □   □

On the one hand, Theorem 9 implies that divergence risk measures are law invariant. Moreover, note that $u(s) = -\varphi^*(-s)$ can be interpreted as a normalized utility function. Specifically, we have $u(0) = 0$ because 0 is the minimum of φ, and $0 \in \partial u(0)$ because 1 is a minimizer of φ. The utility function u can be used to define the optimized certainty equivalent $\sup_{s \in \mathbb{R}} s + \mathbb{E}_{\mathbb{P}_0}[u(X - s)]$, which quantifies the expected utility of an optimized payment schedule that splits an uncertain future revenue X into an amount s that is paid today and a remainder $X - s$ that is paid after the uncertainty has been revealed. By Theorem 9, the divergence risk measure induced by φ thus coincides with the negative optimized certainty equivalent corresponding to the utility function u.

Example 5 (Divergence Risk Measures). *We list examples of divergence risk measures.*

  (i) *Expected value: $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X]$ is obtained by setting $\varphi(t) = 0$, that is, $u(s) = s$.*

  (ii) *Conditional value-at-risk: $\rho_0(X) = \mathrm{CV@R}_\alpha(X)$ for $\alpha \in [0, 1)$ is obtained by setting $\varphi(t) = 0$ if $0 \le t \le \frac{1}{1-\alpha}$ and $\varphi(t) = \infty$ else, which corresponds to $u(s) = \frac{1}{1-\alpha} \min\{s, 0\}$.*

  (iii) *Entropic risk measure: $\rho_0(X) = \frac{1}{\theta} \ln \mathbb{E}_{\mathbb{P}_0}[e^{-\theta X}]$ with $\theta > 0$ is obtained by setting $\varphi(t) = \frac{1}{\theta}(1 - t + t \ln t)$, which corresponds to $u(s) = \frac{1}{\theta}(1 - e^{-\theta s})$.*

The example of the entropic risk measure indicates that divergence risk measures are not necessarily scale invariant and may thus fail to be coherent, see also Theorem 3.1 in Ben-Tal and Teboulle (2007).

Theorem 10. *Every divergence risk measure $\rho_0 : \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0) \to \mathbb{R}$ is randomization-proof.*

Proof. Proof. Select any $X, Y \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ and $\theta \in [0, 1]$, and denote by ρ the unique law invariant extension of ρ_0. Then, we have

$$
\begin{aligned}
\rho(X \oplus_\theta Y) &= \inf_{s \in \mathbb{R}} -s + \mathbb{E}_\mathbb{P}\left[\varphi^*(s - X \oplus_\theta Y)\right] \\
&= \inf_{s \in \mathbb{R}} -s + \theta \mathbb{E}_\mathbb{P}\left[\varphi^*(s - X)\right] + (1 - \theta)\mathbb{E}_\mathbb{P}\left[\varphi^*(s - Y)\right] \\
&\geq \theta\left(\inf_{s \in \mathbb{R}} -s + \mathbb{E}_\mathbb{P}\left[\varphi^*(s - X)\right]\right) + (1 - \theta)\left(\inf_{s \in \mathbb{R}} -s + \mathbb{E}_\mathbb{P}\left[\varphi^*(s - Y)\right]\right) \\
&= \theta\rho(X) + (1 - \theta)\rho(Y) \; \geq \; \min\{\rho(X), \rho(Y)\},
\end{aligned}
$$

where the first equality follows from Theorem 9, and the second equality holds due to Remark 2. The two inequalities follow from the superadditivity of the infimum operator and the fact that $\theta \in [0, 1]$, respectively. Thus, ρ_0 is randomization-proof. $\qquad\square \qquad\qquad \square$

5.3 Shortfall Risk Measures

The shortfall risk of a position X is the smallest cash amount that must be added to X in order to lift the expected utility of X above a prescribed threshold or acceptability level. The shortfall risk measures belong to the class of elicitable risk measures (Bellini and Bignozzi 2015), which encompasses also nonconvex risk measures such as the value-at-risk. All elicitable and—*a fortiori*—all shortfall risk measures exhibit favorable statistical properties and thus lend themselves to historical backtesting (Bellini and Bignozzi 2015).

Definition 11 (Shortfall Risk Measure). *A mapping $\rho_0 : \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0) \to \mathbb{R}$ is called a utility-based shortfall risk measure if there is a concave nondecreasing utility function $u : \mathbb{R} \to \mathbb{R}$ with*

$$
\rho_0(X) = \inf_{s \in \mathbb{R}}\{s : \mathbb{E}_{\mathbb{P}_0}[u(X + s)] \geq u(0)\}.
$$

One can readily verify that shortfall risk measures are monetary, convex and law invariant.

Example 6 (Shortfall Risk Measures). *We list examples of shortfall risk measures.*

(i) *Expected value: $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X]$ is obtained by setting $u(s) = s$.*

(ii) *Entropic risk measure: $\rho_0(X) = \frac{1}{\theta}\ln \mathbb{E}_{\mathbb{P}_0}[e^{-\theta X}]$ with $\theta > 0$ is obtained by setting $u(s) = -e^{-\theta s}$.*

(iii) Expectiles are shortfall risk measures induced by positive homogeneous utility functions of the form $u(s) = \min\{\tau s, (1-\tau)s\}$ for $\tau \in [\frac{1}{2}, 1]$. The family of the expectiles coincides with the class of all coherent shortfall risk measures (Bellini and Bignozzi 2015, Theorem 4.9).

We now establish that shortfall risk measures are randomization-proof.

Theorem 11. *Every shortfall risk measure $\rho_0 : \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0) \to \mathbb{R}$ is randomization-proof.*

Proof. Proof. Select any $X, Y \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ and $\theta \in [0,1]$, and denote by ρ the unique law invariant extension of ρ_0 to $(\Omega, \mathcal{F}, \mathbb{P})$. Then, we have

$$
\begin{aligned}
\rho(X \oplus_\theta Y) &= \inf\{s : \mathbb{E}_\mathbb{P}[u(X \oplus_\theta Y + s)] \geq u(0)\} \\
&= \inf\{s : \theta\mathbb{E}_{\mathbb{P}_0}[u(X+s)] + (1-\theta)\mathbb{E}_{\mathbb{P}_0}[u(Y+s)] \geq u(0)\} \\
&\geq \min\{\inf\{s_X : \mathbb{E}_{\mathbb{P}_0}[u(X+s_X)] \geq u(0)\}, \ \inf\{s_Y : \mathbb{E}_{\mathbb{P}_0}[u(Y+s_Y)] \geq u(0)\}\} \\
&= \min\{\rho(X), \rho(Y)\},
\end{aligned}
$$

where the second equality is due to Remark 2, and the inequality holds because for every s feasible in the second line either $s_X = s$ is feasible in the first minimization problem or $s_Y = s$ is feasible in the second minimization problem in the third line. Thus, the claim follows. $\square$ $\square$

5.4 Other Randomization-Proof Risk Measures

Below we survey several popular law invariant risk measures that are randomization-proof but do not fit into any of the categories considered so far.

Value-at-Risk. The value-at-risk $\rho_0(X) = \inf\{x \in \mathbb{R} : \mathbb{P}_0(X \leq -x) \geq \alpha\}$ at level $\alpha \in [0,1]$ is a monetary risk measure but fails to be convex (Föllmer and Schied 2011, Example 4.11). Note that it is reminiscent of a shortfall risk measure with a non-concave utility function. To show that the value-at-risk is randomization-proof, one can thus use similar arguments as in the proof of Theorem 11. Details are omitted for brevity of exposition.

Mean Variance. The mean variance or Markowitz risk measure $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \alpha\mathbb{E}_{\mathbb{P}_0}[\mathbb{E}_{\mathbb{P}_0}[X] - X]^2$ with $\alpha > 0$ is convex and translation invariant but fails to be monotonic and scale invariant (Shapiro et al. 2009, Example 6.18). It is well-known that the mean variance risk measure can

24

be reformulated as $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \alpha \inf_{t \in \mathbb{R}} \mathbb{E}_{\mathbb{P}_0}[X-t]^2$, see, *e.g.*, Example 2.2.6 in Casella and Berger (2002), which implies that it is mixture-concave and thus randomization-proof.

Mean Standard-Deviation. Similar to the mean variance, the convex, translation invariant and scale invariant (but not monotonic) risk measure $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \alpha \sqrt{\mathbb{E}_{\mathbb{P}_0}[(X - \mathbb{E}_{\mathbb{P}_0}[X])^2]}$, where $\alpha > 0$, can be reformulated as $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \inf_{t \in \mathbb{R}} \alpha \sqrt{\mathbb{E}_{\mathbb{P}_0}[(X-t)^2]}$. We thus conclude that the mean standard-deviation is mixture-concave and therefore randomization-proof.

Mean (Semi-)Deviation from Target. The mean deviation from target $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \alpha \left(\mathbb{E}_{\mathbb{P}_0} |\tau - X|^p \right)^{1/p}$ and the mean semi-deviation from target $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \alpha \left(\mathbb{E}_{\mathbb{P}_0} [\tau - X]_+^p \right)^{1/p}$ with $\alpha > 0$, $p \geq 1$ and $\tau \in \mathbb{R}$ are monotonic and convex but fail to be translation invariant and scale invariant. By construction, both risk measures are mixture-concave and thus randomization-proof.

Mean (Semi-)Moment from Target. Consider a risk measure of the form $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \alpha \mathbb{E}_{\mathbb{P}_0} |\tau - X|^p$ or $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \alpha \mathbb{E}_{\mathbb{P}_0} [\tau - X]_+^p$, where $\alpha > 0$, $p \geq 1$ and $\tau \in \mathbb{R}$. This risk measure is monotonic and convex, but it is neither translation invariant nor scale invariant. By definition, ρ_0 is mixture-affine and thus randomization-proof.

Mean Weighted Mean-Deviation from Quantile. The mean weighted mean-deviation from quantile is defined as $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \alpha \mathbb{E}_{\mathbb{P}_0} \left[\max\{\epsilon(X - F_X^{-1}(1-\epsilon)), (1-\epsilon)(F_X^{-1}(1-\epsilon) - X)\} \right]$, where $\alpha > 0$, $\epsilon \in [0,1]$ and F_X^{-1} represents the quantile function of X. It is translation invariant, convex and scale invariant but fails to be monotonic in general. It is shown in §6.2.3 of Shapiro et al. (2009) that $\rho_0(X)$ can be reformulated as $\rho_0(X) = \mathbb{E}_{\mathbb{P}_0}[-X] + \inf_{t \in \mathbb{R}} \alpha \mathbb{E}_{\mathbb{P}_0} \left[\max\{\epsilon(X - t), (1-\epsilon)(t - X)\} \right]$. From this representation we conclude that ρ_0 is mixture-concave and thus randomization-proof.

Negative Sharpe Ratio. The negative Sharpe ratio or signal-to-noise ratio $\rho_0(X) = -(\mathbb{E}_{\mathbb{P}_0}[X] - \tau)/\sqrt{\mathbb{E}_{\mathbb{P}_0}[\mathbb{E}_{\mathbb{P}_0}[X] - X]^2}$ with $\tau \in \mathbb{R}$ is defined for all random variables satisfying $\mathbb{E}_{\mathbb{P}_0}[X] \geq \tau$.[5] It fails to be monotonic, translation invariant, convex and scale invariant but enjoys quasiconvexity in X. To prove that ρ_0 is mixture-quasiconcave, it suffices to show that the set of all distribution functions F_X corresponding to random variables satisfying $\mathbb{E}_{\mathbb{P}_0}[X] \geq \tau$ and $\rho_0(X) \geq \beta$ is convex

---

[5]Here we adopt the convention that $c/0 = \infty$ for all $c \geq 0$.

for every $\beta \in \mathbb{R}$. This holds, however, because the feasible set of the constraint

$$
\begin{aligned}
\rho_0(X) \geq \beta \quad &\iff \quad (\mathbb{E}_{\mathbb{P}_0}[X] - \tau) + \beta \sqrt{\mathbb{E}_{\mathbb{P}_0}[\mathbb{E}_{\mathbb{P}_0}[X] - X]^2} \leq 0 \\
&\iff \quad (\mathbb{E}_{\mathbb{P}_0}[X] - \tau) + \beta \inf_{t \in \mathbb{R}} \sqrt{\mathbb{E}_{\mathbb{P}_0}[X - t]^2} \leq 0 \\
&\iff \quad (\mathbb{E}_{\mathbb{P}_0}[X] - \tau) + \beta \sqrt{\mathbb{E}_{\mathbb{P}_0}[X - t]^2} \leq 0 \quad \forall t \in \mathbb{R} \\
&\iff \quad \left(\int_{\mathbb{R}} x \, \mathrm{d}F_X(x) - \tau \right) + \beta \left(\int_{\mathbb{R}} (x - t)^2 \, \mathrm{d}F_X(x) \right)^{\frac{1}{2}} \leq 0 \quad \forall t \in \mathbb{R}
\end{aligned}
$$

is convex in F_X whenever $\beta \leq 0$ and collapses to the (convex) singleton set $\{F_\tau\}$ for $\beta > 0$, where F_τ is the Dirac distribution that places unit mass on τ. Our argument implies that ρ_0 is mixture-quasiconcave and thus randomization-proof.

6 Randomization under Distributional Ambiguity

In many decision-making situations the probability measure $\mathbb{P}_0$ governing all relevant random variables is itself uncertain. Indeed, $\mathbb{P}_0$ must usually be estimated from limited statistical data or guessed on the basis of expert judgment, and therefore one can typically only guarantee that $\mathbb{P}_0$ belongs to a (nonempty) family $\mathcal{P}_0$ of distributions that share certain structural or statistical properties. We will henceforth refer to $\mathcal{P}_0$ as the ambiguity set.

In analogy to Section 3, we first formalize the pure and randomized strategy problems under distributional ambiguity. We then show that under mild assumptions, these decision problems correspond to distributionally robust optimization problems (Delage and Ye 2010, Goh and Sim 2010, Wiesemann et al. 2014b). At the end of the section, we prove that for any risk measure satisfying a mild continuity property, we can construct a distributionally robust optimization problem with a nonconvex feasible region in which a randomized strategy strictly dominates all pure strategies.

We introduce the notion of an ambiguous probability space $(\Omega, \mathcal{F}_0, \mathcal{P}_0)$, $i.e.$, a measurable space $(\Omega, \mathcal{F}_0)$ endowed with an ambiguity set $\mathcal{P}_0$. Let $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0) = \cap_{\mathbb{P} \in \mathcal{P}_0} \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathbb{P})$ be the space of all random variables that are essentially bounded with respect to every probability measure in the ambiguity set, and note that any $X \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ can have multiple distribution functions. Below, we denote by $F_X^{\mathbb{P}}$ the distribution function of X under the probability measure $\mathbb{P} \in \mathcal{P}_0$. We can then construct the ambiguity set of all possible distributions of X as $\{F_X^{\mathbb{P}} : \mathbb{P} \in \mathcal{P}_0\}$.

We will henceforth assume that the ambiguous probability space $(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ is non-atomic.

Definition 12 (Non-Atomic Ambiguous Probability Space). *The ambiguous probability space* $(\Omega, \mathcal{F}_0, \mathcal{P}_0)$ *is non-atomic if there exists a random variable* $U_0 \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ *that follows a standard uniform distribution under every probability measure* $\mathbb{P} \in \mathcal{P}_0$.

Definition 12 can always be enforced by introducing a dummy random variable that is known to follow the standard uniform distribution, and it implies that $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ is rich enough to contain—modulo law invariance—all bounded random variables with an *un*ambiguous distribution.

Lemma 3 (Richness of $\mathcal{L}_\infty(\Omega, \mathcal{F}, \mathcal{P}_0)$). *If* $(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ *is non-atomic, then for every* $F \in \mathcal{D}$ *there exists* $X \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ *with* $F_X^{\mathbb{P}} = F$ *for all* $\mathbb{P} \in \mathcal{P}_0$.

Proof. Proof. The proof closely parallels that of Lemma 1 and is thus omitted. $\square$ $\square$

The definition of monetary risk measures naturally extends to ambiguous probability spaces if we require $X \geq Y$ $\mathbb{P}$-almost surely for every $\mathbb{P} \in \mathcal{P}$ in the antecedent of the monotonicity condition. The definitions of convex and coherent risk measures also directly extend to ambiguous probability spaces. However, the notion of law invariance has to be reconciled in the presence of ambiguity.

Definition 13 (Law Invariance Revisited). *A risk measure* $\rho_0 : \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0) \to \mathbb{R}$ *is called law invariant if it satisfies the following condition for all* $X, Y \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$:

Law invariance: *If* $\{F_X^{\mathbb{P}} : \mathbb{P} \in \mathcal{P}_0\} = \{F_Y^{\mathbb{P}} : \mathbb{P} \in \mathcal{P}_0\}$, *then* $\rho_0(X) = \rho_0(Y)$.

Note that Definition 13 reduces to Definition 2 for the singleton ambiguity set $\mathcal{P} = \{\mathbb{P}_0\}$. We can now show that for any law invariant risk measure ρ_0 on a non-atomic ambiguous probability space there is a real-valued functional ϱ_0 on the set $\mathcal{D}$ of all distribution functions with bounded support such that $\rho_0(X) = \varrho_0(F_X)$ for each random variable $X \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ with an *un*ambiguous distribution function $F_X = F_X^{\mathbb{P}}$ for all $\mathbb{P} \in \mathcal{P}_0$.

Proposition 3 (Existence and Uniqueness of ϱ_0). *If* $(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ *is non-atomic and* ρ_0 *is a law invariant risk measure on* $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$, *then there is a unique functional* ϱ_0 *on* $\mathcal{D}$ *such that for all* $F \in \mathcal{D}$, *we have*

$$\varrho_0(F) = \rho_0(X) \qquad \forall X \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0) : F_X^{\mathbb{P}} = F \;\; \forall \mathbb{P} \in \mathcal{P}_0. \tag{9}$$

Proof. Proof. The proof parallels that of Proposition 1, *i.e.*, the mapping ϱ_0 exists because ρ_0 is law invariant, and it is unique because $(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ is non-atomic. Details are omitted for brevity. $\square$ $\square$

By definition, ϱ_0 is constructed from ρ_0. But it is not possible to reconstruct ρ_0 from ϱ_0 because ϱ_0 only contains information on the restriction of ρ_0 to the space of *un*ambiguous random variables, *i.e.*, the random variables with an *un*ambiguous distribution function. A unique backtransformation exists, however, if ρ_0 constitutes an ambiguity averse risk measure.

Definition 14 (Ambiguity Averse Risk Measure)**.** *A risk measure* $\rho_0 : \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0) \to \mathbb{R}$ *is called ambiguity averse if it satisfies the following conditions for all* $X, Y \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$:

Ambiguity aversion: *If* $\{F_X^{\mathbb{P}} : \mathbb{P} \in \mathcal{P}_0\} \subseteq \{F_Y^{\mathbb{P}} : \mathbb{P} \in \mathcal{P}_0\}$, *then* $\rho_0(X) \leq \rho_0(Y)$.

Ambiguity monotonicity: *If* $\varrho_0(F_X^{\mathbb{P}}) \leq \varrho_0(F_Y^{\mathbb{P}})$ *for all* $\mathbb{P} \in \mathcal{P}_0$, *then* $\rho_0(X) \leq \rho_0(Y)$.

Ambiguity aversion implies law invariance in the sense of Definition 13. Thus, Proposition 3 ensures that the functional ϱ_0 in the ambiguity monotonicity condition is well-defined. Ambiguity aversion states that X is weakly preferred to Y if the distribution of X is less ambiguous than that of Y with respect to the partial order induced by set inclusion. It captures the empirical observation that many decision makers display a very low tolerance for ambiguity and is in line with the current thinking in distributionally robust optimization, where one seeks decisions that are worst-case optimal with respect to all distributions in a prescribed ambiguity set (Delage and Ye 2010, Goh and Sim 2010, Wiesemann et al. 2014b). Ambiguity monotonicity states that X is weakly preferred to Y under an ambiguous probability measure if X is weakly preferred to Y under each fixed probability measure in the given ambiguity set. Ambiguity monotonicity is reminiscent of the classical monotonicity axiom from Definition 4. We note that there are ambiguity monotone risk measures that violate the monotonicity axiom from Definition 4. Thus, ambiguity monotonicity does not imply classical monotonicity in general.

Our next theorem provides an important representation for ambiguity averse risk measures.

Theorem 12 (Reconstructing ρ_0 from ϱ_0)**.** *If* $(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ *is non-atomic and* ρ_0 *represents an ambiguity averse monetary risk measure on* $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$, *then* $\rho_0(X) = \sup_{\mathbb{P} \in \mathcal{P}_0} \varrho_0(F_X^{\mathbb{P}})$.

We relegate the proof of Theorem 12 to the appendix. The theorem provides an axiomatic justification for the distributionally robust optimization approach advocated, among others, by Delage and Ye (2010), Goh and Sim (2010) or Wiesemann et al. (2014b). In fact, it shows that any decision maker who employs an ambiguity averse risk measure in the sense of Definition 14 should select actions in view of the worst probability distribution in the ambiguity set. Ambiguity aversion is a natural property for risk measures under distributional ambiguity, and it is satisfied by all commonly used risk measures in the distributionally robust optimization literature. Examples include the worst-case expectation as well as the worst-case variants of the value-at-risk and the conditional vaue-at-risk, see, *e.g.*, Ghaoui et al. (2003), Postek et al. (2016), Zhu and Fukushima (2009) and Zymler et al. (2013), among many others.

Let us now contemplate decision problems of the form (3), where ρ_0 represents an ambiguity averse monetary risk measure, while $\mathcal{X}_0$ constitutes a subset of $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$. As in Section 3, our main goal is to identify situations in which the optimal value of (3) can be strictly improved by allowing for randomized strategies. Thus, we introduce a randomization device generating standard uniformly distributed samples that are independent of all random variables in $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ under every probability measure in $\mathcal{P}_0$. Formally, we augment the ambiguous probability space $(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ to $(\Omega, \mathcal{F}, \mathcal{P})$, where $\Omega = \Omega_0 \times [0, 1]$, $\mathcal{F} = \mathcal{F}_0 \otimes \mathcal{B}_{[0,1]}$ and $\mathcal{P} = \{\mathbb{P} \times \mathbb{U} : \mathbb{P} \in \mathcal{P}_0\}$ with $\mathbb{U}$ denoting the uniform distribution on $[0, 1]$. The set of all admissible randomized strategies is then defined as

$$\mathcal{X} = \{X \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathcal{P}) \ : \ X(\cdot, u) \in \mathcal{X}_0 \ \forall u \in [0, 1]\}.$$

It remains to be shown that ρ_0 can be uniquely extended to an ambiguity averse monetary risk measure ρ on $(\Omega, \mathcal{F}, \mathcal{P})$ that coincides with ρ_0 on $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$.

Proposition 4 (Ambiguity Averse Extension). *If $(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ is non-atomic and ρ_0 is an ambiguity averse monetary risk measure on $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$, then the unique extension of ρ_0 to an ambiguity averse monetary risk measure ρ on $\mathcal{L}_\infty(\Omega, \mathcal{F}, \mathcal{P})$ is given by $\rho(X) = \sup_{\mathbb{P} \in \mathcal{P}} \varrho_0(F_X^{\mathbb{P}})$.*

Proof. Proof. Let ρ be any ambiguity averse risk measure that extends ρ_0 to $\mathcal{L}_\infty(\Omega, \mathcal{F}, \mathcal{P})$, and note that $(\Omega, \mathcal{F}, \mathcal{P})$ is non-atomic as an extension of a non-atomic ambiguous probability space. Proposition 3 and Theorem 12 thus ensure the existence of a mapping $\varrho : \mathcal{D} \to \mathbb{R}$ with $\rho(X) = \sup_{\mathbb{P} \in \mathcal{P}} \varrho(F_X^{\mathbb{P}})$ for all $X \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathcal{P})$. Hence, if $X \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ with $F_X^{\mathbb{P}} = F_X^{\mathbb{Q}} = F_X$ for all

$\mathbb{P}, \mathbb{Q} \in \mathcal{P}$, then

$$\varrho(F_X) = \sup_{\mathbb{P} \in \mathcal{P}} \varrho(F_X^{\mathbb{P}}) = \rho(X) = \rho_0(X) = \sup_{\mathbb{P} \in \mathcal{P}} \varrho_0(F_X^{\mathbb{P}}) = \varrho_0(F_X),$$

where the first and the fifth equalities hold because X is unambiguous, the second and the fourth equalities follow from the definitions of ϱ and ϱ_0, respectively, and the third equality holds because ρ must equal ρ_0 on $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$. Lemma 3 then implies that $\varrho(F) = \varrho_0(F)$ for all $F \in \mathcal{D}$. Thus, $\rho(X) = \sup_{\mathbb{P} \in \mathcal{P}} \varrho_0(F_X^{\mathbb{P}})$ is the only ambiguity averse risk measure extending ρ_0 to $\mathcal{L}_\infty(\Omega, \mathcal{F}, \mathcal{P})$. □

□

Remark 4 (Law Invariance in $(\Omega, \mathcal{F}, \mathcal{P})$). *Similar to the law invariance in Section 3, the assumption that ρ is ambiguity averse on $(\Omega, \mathcal{F}, \mathcal{P})$ implies that the decision maker agrees with a reduction of compound lotteries axiom (Samuelson 1952) over the ambiguous probability space. It also means that a random variable X depending on both the randomization device and $\omega_0 \in \Omega_0$ is considered equivalent to a random variable Y that is independent of the randomization device whenever Y shares the same set of cumulative distribution functions as X. In the context of Example 2 this implies that "picking a color based on the throw of a fair coin is considered exactly equivalent to playing a game in which the urn has known equal proportions of blue and red balls."*

By an immediate extension of Theorem 2, one can prove that there is no benefit in adopting a randomized strategy when minimizing a convex ambiguity averse monetary risk measure over a convex set $\mathcal{X}_0$. In stark contrast to Section 3, however, most ambiguity averse monetary risk measures induce a strict preference for randomization for some $\mathcal{X}_0$. Indeed, under a mild regularity condition, for any ambiguity averse monetary risk measure there exists a variant of the Ellsberg urn problem from Section 1 in which a randomized strategy strictly dominates all pure strategies. We will first motivate this statement through an example and then provide a formal proof.

Example 7 (The Rainbow Urn Game). *An urn contains balls of K different colors (e.g., red, orange, yellow, ..., indigo, violet), but the number of balls and the proportions of the colors are unknown. A player is asked to name a color and to draw a random ball from the urn. If the chosen ball has the stated color, the player pays a penalty of \$1. Otherwise, there is a reward of \$1. As the distribution of colors is completely ambiguous, the worst scenario for a player selecting color k is that all balls in the urn are painted in color k, in which case the penalty of \$1 arises*

with certainty. An ambiguity averse player restricting attention to pure strategies is thus indifferent between participating in the game and paying a fixed amount of $1. If the player were to select one of the K pure strategies uniformly at random (e.g., by rolling a fair K-sided die), however, the ambiguity could be completely suppressed. Specifically, the player would face a reward of $1 with probability $\frac{K-1}{K}$ and a penalty of $1 with probability $\frac{1}{K}$. Under the vast majority of commonly used risk measures,[6] this option is asymptotically (i.e., for large K) as attractive as receiving a certain reward of $1.

In order to facilitate a rigorous treatment of the rainbow urn game, we have to introduce the notion of a maximally ambiguous random variable.

Definition 15 (Maximally Ambiguous Random Variable). *A random variable $V \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ is maximally ambiguous if it can have any bounded distribution, that is, if $\{F_V^{\mathbb{P}} : \mathbb{P} \in \mathcal{P}_0\} = \mathcal{D}$.*

Similar to the assumption of non-atomicity, the existence of a maximally ambiguous random variable is a mild assumption that can be met by augmenting the ambiguous probability space.

We will also focus on ambiguity averse monetary risk measures satisfying the Lebesgue property. In order to define the Lebesgue property, recall that a sequence of distribution functions $F_k \in \mathcal{D}$, $k \in \mathbb{N}$, converges weakly to $F \in \mathcal{D}$ if $\lim_{k\uparrow\infty} F_k(x) = F(x)$ for all $x \in \mathbb{R}$ where F is continuous.

Definition 16 (Lebesgue Property). *An ambiguity averse monetary risk measure representable as $\rho_0(X) = \sup_{\mathbb{P} \in \mathcal{P}_0} \varrho_0(F_X^{\mathbb{P}})$ is said to satisfy the Lebesgue property if $\lim_{k\uparrow\infty} \varrho_0(F_k) = \varrho_0(F)$ whenever F_k converges weakly to F.*

Definition 16 generalizes the original definition of the Lebesgue property in Jouini et al. (2006) to ambiguous probability spaces. It is a mild technical condition which ensures that the mapping ϱ_0 is continuous with respect to weak convergence of distribution functions. Apart from the essential supremum in Example 4, all risk measures studied in this paper satisfy the Lebesgue property. In the context of Example 7, the Lebesgue property ensures that for a rainbow urn with sufficiently many colors K, randomizing uniformly over all K colors outperforms every pure decision.

We are now armed to state the main result of this section.

Theorem 13. *Assume that $(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ is non-atomic and that there is a maximally ambiguous random variable $V \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$. Then, all ambiguity averse monetary risk measures*

---

[6]The risk measure must satisfy the Lebesgue property introduced in Definition 16 below.

ρ_0 *that satisfy the Lebesgue property are randomization-receptive,* i.e.*, there exists a pair $X, Y \in$* $\mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ *and a $\theta \in [0, 1]$ such that $\rho(X \oplus_\theta Y) < \min\{\rho(X), \rho(Y)\}$.*

We relegate the proof of Theorem 13 to the appendix. At its core, the proof constructs for any ambiguity averse monetary risk measure ρ_0 satisfying the Lebesgue property an instance of the rainbow urn game such that the randomized strategy from Example 7 strictly dominates all pure strategies. To this end, the number of potential colors K in the rainbow urn has to be chosen large enough so that the probability $\frac{1}{K}$ of losing under the randomized strategy has a strictly lower risk than the certain loss under any deterministic decision. The Lebesgue property guarantees that this is the case for some finite K. It is worth noting that the proof does not exploit the convexity of ρ_0, and thus the theorem generalizes to risk measures that do not capture risk aversion.

Theorem 13 suggests that ambiguity averse decision makers solving problem (3) instead of (4) may unwittingly expose themselves to avoidable risks and thus act irrationally. Furthermore, it indicates that various recent distributionally robust optimization problems involving discrete choices such as appointment scheduling (Mittal et al. 2014), knapsack (Cheng et al. 2014), hospital admission (Meng et al. 2015), shortest path (Cheng et al. 2016), unit commitment (Jiang et al. 2016), facility location (Chan et al. 2015), and vehicle routing problems (Gounaris et al. 2013) etc. might benefit from randomization.

7   A Remark about Time Consistency

Consider again Example 1 from Section 1, where the attractiveness of a project was assessed via the mean semi-variance of its uncertain NPV. We have seen that if exactly one out of the five projects specified in Table 1 can be chosen, then it is optimal to randomly select projects 1 and 4 with probabilities 81% and 19%, respectively, which results in a mean semi-variance of -0.69.

Upon further reflection one realizes that, by allowing for randomized strategies, the single-stage project selection problem has essentially been transformed to a two-stage problem. Indeed, the succession of events is as follows. First, the manager selects a decision rule or contingency plan that assigns each possible outcome $u \in [0, 1]$ of the randomization device a project to be chosen upon observation of u. Then, "the dice are rolled" and a particular outcome u is observed. Finally, the project associated with outcome u is implemented. Only then the outcome ω_0 associated with

the exogenous uncertainty is observed, and the corresponding NPV is realized.

An optimal strategy for this two-stage problem is to select project 1 if $u \leq 81\%$ and project 4 if $u > 81\%$. Unfortunately, this strategy suffers from a time inconsistency and may therefore not be implemented. Indeed, after observing u, the manager could follow the recommendation of the randomization device and implement project 1 or project 4, which would result in a mean semi-variance of -0.52 or 1.13, respectively. Instead, she could implement project 2, which would result in a lower mean semi-variance of -0.53; see Table 1. Thus, as soon as the outcome of the randomization device is known (and there is no possibility to roll the dice for a second time), there is an incentive to deviate from the action prescribed by the optimal randomized strategy and to select the optimal pure strategy instead.

This issue of *time inconsistency* commonly arises in dynamic decision-making situations where multiple actions must be implemented over time. A seminal paper in this regard is due to Strotz (1955–56), who shows that when decision makers maximize their intertemporal utility subject to a budget constraint, their optimal course of actions will be time inconsistent, even in the absence of uncertainty, unless future utility is discounted at a constant rate of interest. The paper suggests that decision makers can overcome this dilemma by locking themselves into implementing the envisaged course of actions today (in our case, by locking oneself into implementing the randomized decision) through a *commitment device*. In our context, the decision maker might settle for a randomized strategy but outsource the implementation of this policy to a third party. Alternatively, she could enforce the implementation of the optimal strategy contractually by agreeing to pay a penalty for any deviations from this strategy. Yet another possibility for the manager would be to publicize the optimal strategy upfront, thus putting her reputation at stake. For a detailed discussion of commitment devices and their applications in economics and finance, we refer to Bryan et al. (2010).

Some decision makers might feel that the time inconsistency phenomenon outlined above is unacceptable and that any decision criteria susceptible to randomization should therefore be rejected as inappropriate. While we would not argue against such a position, we highlight again that many randomization-receptive risk measures (both classical and ambiguity averse) are used routinely in academia and practice. When such risk measures are employed, it is a mathematical fact that decision makers who neglect randomized strategies might sacrifice performance.

8 Conclusion

In this paper we investigated whether randomized strategies can strictly outperform the best deterministic strategy in stochastic and distributionally robust optimization problems that involve (worst-case) risk measures. We confirmed that randomization does not help in many common settings, such as in optimization problems that are convex or that minimize spectral, divergence or shortfall risk measures. Interestingly, however, the optimality of deterministic strategies is *not* a direct consequence of the axioms of law invariant and coherent risk measures (see, *e.g.*, the mean semi-deviation measure in Section 4), and it is not tenable in a distributionally robust setting. This is in sharp contrast to the expected utility framework commonly adopted in economics, where the optimality of deterministic strategies directly follows from the independence axiom.

Our results provide a justification for the common practice of focusing on deterministic strategies in stochastic programming whenever the employed risk measure is randomization-proof. On the other hand, our results indicate that randomized decisions may be beneficial in several classes of recently studied distributionally robust optimization problems, such as project management, appointment scheduling, vehicle routing and supply chain design problems. It would be instructive to explore whether the theoretical analysis in this paper, which shows that randomized decisions may outperform pure strategies in these settings, translates into practical benefits in these applications. We thus identify the design of tractable algorithms for optimizing over randomized strategies, as well as their application in the aforementioned domains, as a promising area for future research.

References

Agranov M, Ortoleva P (2017) Stochastic choice and preferences for randomization. *Journal of Political Economy* 125(1):40–68.

An B, Tambe M, Sinha A (2016) Stackelberg security games (SSG): Basics and application overview. Abbas A, Tambe M, von Winterfeldt D, eds., *Improving Homeland Security Decisions* (Cambridge University Press).

Artzner P, Delbaen F, Eber JM, Heath D (1999) Coherent measures of risk. *Mathematical Finance* 9(3):203–228.

Bäuerle N, Müller A (2006) Stochastic orders and risk measures: Consistency and bounds. *Insurance: Mathematics and Economics* 38(1):132–148.

Bellini F, Bignozzi V (2015) On elicitable risk measures. *Quantitative Finance* 15(5):725–733.

Ben-Tal A, Ghaoui LE, Nemirovski A (2009) *Robust Optimization* (Princeton University Press).

Ben-Tal A, Teboulle M (2007) An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance* 17(3):449–476.

Bertsimas D, Nasrabadi E, Orlin JB (2016) On the power of randomization in network interdiction. *Operations Research Letters* 44(1):114–120.

Birge JR, Louveaux F (2000) *Introduction to Stochastic Programming* (Springer).

Bryan G, Karlan D, Nelson S (2010) Commitment devices. *Annual Review of Economics* 2:671–698.

Campi M (2010) Why is resorting to fate wise? a critical look at randomized algorithms in systems and control. *European Journal of Control* 16(5):419–430.

Casella G, Berger RL (2002) *Statistical Inference* (Duxbury Thomson Learning), 2nd edition.

Cettolin E, Riedl A (2016) Revealed incomplete preferences under uncertainty: Evidence for bewley preferences, available on SSRN.

Chan TCY, Shen ZJM, Siddiq A (2015) Robust facility location under demand location uncertainty, available on ArXiv.

Cheng J, Delage E, Lisser A (2014) Distributionally robust stochastic knapsack problem. *SIAM Journal on Optimization* 24(3):1485–1506.

Cheng J, Leung J, Lisser A (2016) New reformulations of distributionally robust shortest path problem. *Computers & Operations Research* 74:196–204.

Cover TM, Thomas JA (1991) *Elements of Information Theory* (Wiley).

DeJarnette P, Dillenberger D, Gottlieb D, Ortoleva P (2015) Time lotteries, available on SSRN.

Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* 58(3):595–612.

Dhaene J, Kukush A (2011) Comonotonic modification of random vector in its own probability space, available on SSRN.

Dominiak A, Schnedler W (2011) Attitudes towards uncertainty and randomization: An experimental study. *Economic Theory* 48(2):289–312.

Dwenger N, Kübler D, Weizsäcker G (2016) Flipping a coin: Evidence from laboratory and field, technical Report, University of Hohenheim.

Eichberger J, Grant S, Kelsey D (2016) Randomization and dynamic consistency. *Economic Theory* 63(3):547–566.

Eichberger J, Kelsey D (1996) Uncertainty aversion and preference for randomisation. *Journal of Economic Theory* 71(1):31–43.

Epstein LG (1999) A definition of uncertainty aversion. *Review of Economic Studies* 66(3):579–608.

Föllmer H, Schied A (2011) *Stochastic Finance – An Introduction in Discrete Time* (Walter de Gruyter & Co.), 3rd edition.

Ghaoui LE, Oks M, Oustry F (2003) Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research* 51(4):543–556.

Gilboa I, Schmeidler D (1989) Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18(2):141–153.

Goh J, Sim M (2010) Distributionally robust optimization and its tractable approximations. *Operations Research* 58(4):902–917.

Gounaris CE, Wiesemann W, Floudas CA (2013) The robust capacitated vehicle routing problem under demand uncertainty. *Operations Research* 61(3):677–693.

Gul F, Pesendorfer W (2006) Random expected utility. *Econometrica* 74(1):121–146.

Hey JD, Carbone E (1995) Stochastic choice with deterministic preferences: An experimental investigation. *Economics Letters* 47(2):161–167.

Jiang R, Guan Y, Watson JP (2016) Risk-averse stochastic unit commitment with incomplete information. *IIE Transactions*, Available online first.

Jouini E, Schachermayer W, Touzi N (2006) Law invariant risk measures have the Fatou property. Kusuoka S, Yamazaki A, eds., *Advances in Mathematical Economics* (Springer).

Kellerer H, Mansini R, Speranza MG (2000) Selecting portfolios with fixed costs and minimum transaction lots. *Annals of Operations Research* 99:287–304.

Keynes JM (1921) *A Treatise on Probability* (Macmillan & Co.).

Kircher P, Ludwig S, Sandroni A (2013) On the difference between social and private goods. *B.E. Journals of Theoretical Economics* 13(1):151–177.

Klibanoff P (1996) Uncertainty, decision and normal form games, Technical Report, Massachusetts Institute of Technology.

Knight FH (1921) *Risk, Uncertainty and Profit* (Houghton Mifflin).

Korzhyk D, Yin Z, Kiekintveld C, Conitzer V, Tambe M (2011) Stackelberg vs. Nash in security games: An extended investigation of interchangeability, equivalence, and uniqueness. *Journal of Artificial Intelligence Research* 41(2):297–327.

Kusuoka S (2001) On law invariant coherent risk measures. Kusuoka S, Maruyama T, eds., *Advances in Mathematical Economics*, volume 3, 83–95 (Springer).

Lo KC (1996) Equilibrium in beliefs under uncertainty. *Journal of Economic Theory* 71(2):443–484.

Manzini P, Mariotti M (2014) Stochastic choice and consideration sets. *Econometrica* 82(3):1153–1176.

Mastin A, Jaillet P, Chin S (2014) Randomized minmax regret for combinatorial optimization under uncertainty, available on arXiv.

Meng F, Qi J, Zhang M, Ang J, Chu S, Sim M (2015) A robust optimization model for managing elective admission in a public hospital. *Operations Research* 63(6):1452–1467.

Mittal S, Schulz A, Stiller S (2014) Robust appointment scheduling. *Proceedings of the 17th International Workshop on Approximation and Randomized Algorithms*, 356–370.

Motwani R, Raghavan P (1995) *Randomized Algorithms* (Cambridge University Press).

Nash J (1951) Non-cooperative games. *Annals of Mathematics* 54(2):286–295.

Postek KS, den Hertog D, Melenberg B (2016) Computationally tractable counterparts of distributionally robust constraints on risk measures. *SIAM Review* 58(4):603–650.

Rosenblatt M (1952) Remarks on a multivariate transformation. *The Annals of Mathematical Statistics* 23(3):470–472.

Rubinstein A (2002) Irrational diversification in multiple decision problems. *European Economic Review* 46(8):1369–1378.

Saito K (2012) Subjective timing of randomization and ambiguity, Working Paper, California Institute of Technology.

Saito K (2015) Preferences for flexibility and randomization under uncertainty. *American Economic Review* 105(3):1246–1271.

Samuelson P (1952) Probability, utility, and the independence axiom. *Econometrica* 20(4):1670–678.

Schmeidler D (1989) Subjective probability and expected utility without additivity. *Econometrica* 57(3):571–587.

Shapiro A, Dentcheva D, Ruszczyński A (2009) *Lectures on Stochastic Programming* (SIAM).

Sopher B, Narramore JM (2000) Stochastic choice and consistency in decision making under risk: An experimental study. *Theory and Decision* 48(4):323–349.

Stone P (2011) *A Treatise on Probability* (Oxford University Press).

Strotz RH (1955–56) Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies* 23(3):165–180.

Tallec YL (2007) *Robust, Risk-Sensitive, and Data-Driven Control of Markov Decision Processes.* Ph.D. thesis, Massachusetts Institute of Technology.

Wiesemann W, Kuhn D, Rustem B (2014a) Robust Markov decision processes. *Mathematics of Operations Research* 38(1):153–183.

Wiesemann W, Kuhn D, Sim M (2014b) Distributionally robust convex optimization. *Operations Research* 62(6):1358–1376.

Yu Z (2007) Spatial energy market risk analysis using the semivariance risk measure. *Energy Systems* 29(8):600–608.

Zhu S, Fukushima M (2009) Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research* 57(5):1155–1168.

Zymler S, Kuhn D, Rustem B (2013) Worst-case value at risk of nonlinear portfolios. *Management Science* 59(1):172–188.

A   Proofs: Randomization-Receptive Risk Measures

Proof of Theorem 4.   Set $c = (3/(2 + 2^p))^{1/p}$ and define the two independent random variables

$$X \sim \begin{cases} 1 & \text{w.p. } 1/2, \\ -1 & \text{w.p. } 1/2, \end{cases} \quad \text{and} \quad Y_0 \sim \begin{cases} c & \text{w.p. } 2/3, \\ -2c & \text{w.p. } 1/3. \end{cases}$$

By construction, we have $\mathbb{E}_{\mathbb{P}_0}[X] = \mathbb{E}_{\mathbb{P}_0}[Y_0] = 0$ and $\rho_0(X) = \rho_0(Y_0) = \alpha$.

Next, set $Y_\epsilon = a(\epsilon)Y_0 + \epsilon$ with $a(\epsilon) = (1 + \epsilon/\alpha)^{1/p}$ and $\epsilon \in (-\alpha, \alpha)$, implying that $\mathbb{E}_{\mathbb{P}_0}[Y_\epsilon] = \epsilon$ while still $\rho_0(Y_\epsilon) = \alpha$. To prove that ρ_0 is randomization-receptive for $p \neq 2$, we will argue that

$$f(\epsilon) \;=\; \rho\left(X \oplus_{\frac{1}{2}} Y_\epsilon\right) - \min\left\{\rho(X), \rho(Y)\right\} \;=\; \rho\left(X \oplus_{\frac{1}{2}} Y_\epsilon\right) - \frac{1}{2}\rho(X) - \frac{1}{2}\rho(Y_\epsilon)$$

is strictly negative if either $p < 2$ and $\epsilon < 0$ or $p > 2$ and $\epsilon > 0$ for sufficiently small values of ϵ. To this end, observe that $\rho(Y_\epsilon)$ can be written as $\rho(Y_\epsilon) = -\epsilon + \alpha(\frac{2}{3}[a(\epsilon)c]^p + \frac{1}{3}[2a(\epsilon)c]^p)$ and

$$\rho\left(X \oplus_{\frac{1}{2}} Y_\epsilon\right) = -\frac{\epsilon}{2} + \frac{\alpha}{2}\left(\frac{1}{2}\left[1 - \frac{\epsilon}{2}\right]^p + \frac{1}{2}\left[1 + \frac{\epsilon}{2}\right]^p + \frac{2}{3}\left[a(\epsilon)c + \frac{\epsilon}{2}\right]^p + \frac{1}{3}\left[2a(\epsilon)c - \frac{\epsilon}{2}\right]^p\right)$$

for all ϵ in a neighborhood of 0, which implies that

$$\begin{aligned} f(\epsilon) = {} & \frac{\alpha}{4}\left[1 + \frac{\epsilon}{2}\right]^p + \frac{\alpha}{4}\left[1 - \frac{\epsilon}{2}\right]^p - \frac{\alpha}{2} \\ & + \frac{\alpha}{3}\left[a(\epsilon)c + \frac{\epsilon}{2}\right]^p - \frac{\alpha}{3}[a(\epsilon)c]^p + \frac{\alpha}{6}\left[2a(\epsilon)c - \frac{\epsilon}{2}\right]^p - \frac{\alpha}{6}[2a(\epsilon)c]^p. \end{aligned}$$

Note that $a(0) = 1$ implies $f(0) = 0$. Moreover, differentiating with respect to ϵ yields

$$f'(\epsilon) = \frac{\alpha p}{8}\left[1 + \frac{\epsilon}{2}\right]^{p-1} - \frac{\alpha p}{8}\left[1 - \frac{\epsilon}{2}\right]^{p-1}$$
$$+ \frac{\alpha p}{3}\left[a(\epsilon)c + \frac{\epsilon}{2}\right]^{p-1}\left[a'(\epsilon)c + \frac{1}{2}\right] - \frac{\alpha p}{3}\left[a(\epsilon)c\right]^{p-1}a'(\epsilon)c$$
$$+ \frac{\alpha p}{6}\left[2a(\epsilon)c - \frac{\epsilon}{2}\right]^{p-1}\left[2a'(\epsilon)c - \frac{1}{2}\right] - \frac{\alpha p}{6}\left[2a(\epsilon)c\right]^{p-1}2a'(\epsilon)c.$$

For $\epsilon = 0$ we then obtain

$$f'(0) = \frac{\alpha p c^{p-1}}{6} - \frac{\alpha p (2c)^{p-1}}{12} \neq 0$$

unless $p = 2$. Thus, for any $p \in [1,2) \cup (2,\infty)$ and $\alpha > 0$ there is $\epsilon(p,\alpha)$ such that $\rho(X \oplus_{\frac{1}{2}} Y) < \min\{\rho(X), \rho(Y)\}$ for the random variable $Y = Y_{\epsilon(p,\alpha)}$, that is, ρ_0 is randomization-receptive. $\square$

Proof of Theorem 5. The proof follows the same lines as the proof of Theorem 4 if we replace the random variable Y_ϵ with $Y_\epsilon = a(\epsilon)Y_0 + \epsilon$, where now $a(\epsilon) = 1 + \epsilon/\alpha$. $\square$

Proof of Theorem 6. Set $c = 3^{1/p}/2$ and define the two independent random variables

$$X \sim \begin{cases} 2^{1/p} & \text{w.p. } 1/2, \\ -2^{1/p} & \text{w.p. } 1/2, \end{cases} \qquad \text{and} \qquad Y_0 \sim \begin{cases} c & \text{w.p. } 2/3, \\ -2c & \text{w.p. } 1/3. \end{cases}$$

By construction, we have $\mathbb{E}_{\mathbb{P}_0}[X] = \mathbb{E}_{\mathbb{P}_0}[Y_0] = 0$ and $\rho_0(X) = \rho_0(Y_0) = \alpha$.

Next, set $Y_\epsilon = a(\epsilon)Y_0 + \epsilon$ with $a(\epsilon) = (1 + \epsilon/\alpha)^{1/p}$ and $\epsilon \in (-\alpha, \alpha)$, implying that $\mathbb{E}_{\mathbb{P}_0}[Y_\epsilon] = \epsilon$ while still $\rho_0(Y_\epsilon) = \alpha$. To prove that ρ_0 is randomization-receptive for $p \geq 1$, we will argue that

$$f(\epsilon) = \rho\left(X \oplus_{\frac{1}{2}} Y_\epsilon\right) - \min\{\rho(X), \rho(Y)\} = \rho\left(X \oplus_{\frac{1}{2}} Y_\epsilon\right) - \frac{1}{2}\rho(X) - \frac{1}{2}\rho(Y_\epsilon)$$

is strictly negative if $p \geq 1$ and $\epsilon < 0$ for sufficiently small values of ϵ. To this end, observe that $\rho(Y_\epsilon)$ can be written as $\rho(Y_\epsilon) = -\epsilon + \frac{\alpha}{3}[2a(\epsilon)c]^p$ and

$$\rho\left(X \oplus_{\frac{1}{2}} Y_\epsilon\right) = -\frac{\epsilon}{2} + \frac{\alpha}{2}\left(\frac{1}{2}\left[2^{1/p} + \frac{\epsilon}{2}\right]^p + \frac{1}{3}\left[2a(\epsilon)c - \frac{\epsilon}{2}\right]^p\right)$$

for all ϵ in a neighborhood of 0, which implies that

$$f(\epsilon) = \frac{\alpha}{4}\left[2^{1/p} + \frac{\epsilon}{2}\right]^p + \frac{\alpha}{6}\left[2a(\epsilon)c - \frac{\epsilon}{2}\right]^p - \frac{\alpha}{2} - \frac{\alpha}{6}[2a(\epsilon)c]^p.$$

Note that $a(0) = 1$ implies $f(0) = 0$. Moreover, differentiating with respect to ϵ yields

$$f'(\epsilon) = \frac{\alpha p}{8}\left[2^{1/p} + \frac{\epsilon}{2}\right]^{p-1} + \frac{\alpha p}{6}\left[2a(\epsilon)c - \frac{\epsilon}{2}\right]^{p-1}\left[2a'(\epsilon)c - \frac{1}{2}\right] - \frac{\alpha p}{6}\left[2a(\epsilon)c\right]^{p-1}\left[2a'(\epsilon)c\right].$$

For $\epsilon = 0$ we then obtain

$$f'(0) = \frac{\alpha p}{4}\left[2^{-1/p} - 3^{-1/p}\right] > 0$$

for all $p \geq 1$. Thus, for any $p \geq 1$ and $\alpha > 0$ there is $\epsilon(p, \alpha) < 0$ such that $\rho(X \oplus_{\frac{1}{2}} Y) < \min\{\rho(X), \rho(Y)\}$ for the random variable $Y = Y_{\epsilon(p,\alpha)}$, that is, ρ_0 is randomization-receptive. $\qquad\square$

Proof of Theorem 7. The proof follows the same lines as the proof of Theorem 6 if we replace the random variable Y_ϵ with $Y_\epsilon = a(\epsilon)Y_0 + \epsilon$, where now $a(\epsilon) = 1 + \epsilon/\alpha$. $\qquad\square$

B   Proofs: Randomization under Distributional Ambiguity

Proof of Theorem 12. The proof proceeds in two steps. First we exploit ambiguity aversion to argue that $\rho_0(X) \geq \sup_{\mathbb{P} \in \mathcal{P}_0} \varrho_0(F_X^{\mathbb{P}})$, and then we use ambiguity monotonicity to show that $\rho_0(X) \leq \sup_{\mathbb{P} \in \mathcal{P}_0} \varrho_0(F_X^{\mathbb{P}})$.

Step 1: Select $X \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ and $\mathbb{Q} \in \mathcal{P}_0$, and denote by $F_X^{\mathbb{Q}} \in \mathcal{D}$ the distribution function of X under $\mathbb{Q}$. By Lemma 3, there exists a random variable $Y \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ with $F_Y^{\mathbb{P}} = F_X^{\mathbb{Q}}$ for every $\mathbb{P} \in \mathcal{P}_0$. This means that $\{F_Y^{\mathbb{P}} : \mathbb{P} \in \mathcal{P}_0\} = \{F_X^{\mathbb{Q}}\} \subseteq \{F_X^{\mathbb{P}} : \mathbb{P} \in \mathcal{P}_0\}$, which in turn implies that $\rho_0(X) \geq \rho_0(Y) = \varrho_0(F_X^{\mathbb{Q}})$, where the inequality follows from the ambiguity aversion of ρ_0, while the equality holds due to the definition of ϱ_0 and because Y is unambiguous. As the choice of X and $\mathbb{Q}$ was arbitrary, we may conclude that $\rho_0(X) \geq \sup_{\mathbb{Q} \in \mathcal{P}_0} \varrho_0(F_X^{\mathbb{Q}})$ for every $X \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$.

Step 2: Select again $X \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ and set $s = \sup_{\mathbb{P} \in \mathcal{P}_0} \varrho_0(F_X^{\mathbb{P}})$. By using translation invariance twice, which holds because ρ_0 is monetary, the degenerate random variable $Y = \rho_0(s)$ can be shown to satisfy

$$\rho_0(Y) = \rho_0(\rho_0(s)) = \rho_0(0 + \rho_0(s)) = \rho_0(0) - \rho_0(s) = \rho_0(0) - \rho_0(0 + s) = s.$$

Thus, we find $\varrho_0(F_Y^{\mathbb{P}}) = \rho_0(Y) = s \geq \varrho_0(F_X^{\mathbb{P}})$ for all $\mathbb{P} \in \mathcal{P}_0$, where the first equality holds due to the definition of ϱ_0 and because Y is unambiguous, and the inequality follows from the definition

of s. Hence, X is weakly preferred to Y under each fixed probability measure in the ambiguity set. By ambiguity monotonicity, we thus have $\rho_0(X) \leq \rho_0(Y) = s = \sup_{\mathbb{P} \in \mathcal{P}_0} \varrho_0(F_X^{\mathbb{P}})$. $\qquad\square$

Proof of Theorem 13. The presence of a maximally ambiguous random variable V allows us to model the rainbow urn game portrayed in Example 7 on the basis of the ambiguous probability space $(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$. Specifically, the reward associated with the pure action "name color k" can be represented by the random variable $X_k = 1 - 2 \cdot \mathbf{1}_{\{(k-1)/K \leq W < k/K\}}$ (for $k < K$) or $X_k = 1 - 2 \cdot \mathbf{1}_{\{(K-1)/K \leq W \leq 1\}}$ (for $k = K$), with $W = V \bmod 1$. The distribution function of X_k under $\mathbb{P} \in \mathcal{P}_0$ is thus given by

$$F_{X_k}^{\mathbb{P}}(x) = \begin{cases} 1 & \text{if } x \geq 1, \\ \mathbb{P}(\frac{k-1}{K} \leq W < \frac{k}{K}) & \text{if } -1 \leq x < 1, \\ 0 & \text{if } x < -1, \end{cases}$$

for $k < K$ and

$$F_{X_k}^{\mathbb{P}}(x) = \begin{cases} 1 & \text{if } x \geq 1, \\ \mathbb{P}(\frac{K-1}{K} \leq W \leq 1) & \text{if } -1 \leq x < 1, \\ 0 & \text{if } x < -1. \end{cases}$$

for $k = K$. As V is maximally ambiguous, there exists $\mathbb{P}_k \in \mathcal{P}_0$ with $W = V = \frac{k-1}{K}$ $\mathbb{P}_k$-almost surely. Thus, $X_k = -1$ $\mathbb{P}_k$-almost surely, implying that the distribution function of X_k under $\mathbb{P}_k$ satisfies $F_{X_k}^{\mathbb{P}_k}(x) = 1$ if $x \geq -1$; $= 0$ if $x < -1$. We conclude that $F_{X_k}^{\mathbb{P}_k}(x) \geq F_{X_k}^{\mathbb{P}}(x)$ for all $x \in \mathbb{R}$ and for all $\mathbb{P} \in \mathcal{P}_0$, that is, $F_{X_k}^{\mathbb{P}_k}$ first order stochastically dominates every other distribution function $F_{X_k}^{\mathbb{P}}$, $\mathbb{P} \in \mathcal{P}_0$. Thus, the risk of the pure strategy X_k can be expressed as

$$\rho_0(X_k) = \sup_{\mathbb{P} \in \mathcal{P}_0} \varrho_0(F_{X_k}^{\mathbb{P}}) = \varrho_0(F_{X_k}^{\mathbb{P}_k}) = \rho_0(-1),$$

where the first equality is due to Theorem 12, the second equality holds because any law invariant convex risk measure on a non-atomic probability space is monotonic with respect to first-order stochastic dominance (Bäuerle and Müller 2006, Theorem 4.2), and the last equality exploits again Theorem 12 and the fact that $F_{X_k}^{\mathbb{P}_k}$ coincides with the distribution function of the degenerate constant random variable -1. The optimal value of problem (3) over $\mathcal{X}_0 = \{X_k : k = 1, \ldots, K\}$ is thus given by $\rho_0(-1)$.

Next, we construct the optimization problem (4) over randomized strategies in the usual manner, and we exploit the presence of a randomization device to define a random variable $U \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathcal{P})$

with $U(\omega_0, u) = u$ for all $(\omega_0, u) \in \Omega$. By construction, U is *un*ambiguous and follows the standard uniform distribution irrespective of $\mathbb{P} \in \mathcal{P}$. The reward associated with the randomized action "name any color k with probability $\frac{1}{K}$" can thus be represented through the random variable

$$X_{1:K} = \sum_{k=1}^{K-1} \mathbf{1}_{\{(k-1)/K \leq U < k/K\}} X_k + \mathbf{1}_{\{(K-1)/K \leq U \leq 1\}} X_K.$$

Using a similar reasoning as in Example 2, one can further demonstrate that the distribution function of the randomized strategy $X_{1:K}$ coincides with the equally weighted convex combination of the distribution functions of all pure strategies X_k, that is,

$$F_{X_{1:K}}^{\mathbb{P}}(x) = \frac{1}{K} \sum_{k=1}^{K} F_{X_k}^{\mathbb{P}}(x) = \begin{cases} 0 & \text{if } x < -1, \\ \frac{1}{K} & \text{if } -1 \leq x < 1, \\ 1 & \text{if } x \geq 1, \end{cases}$$

where the second equality follows from the observation that

$$\frac{1}{K} \sum_{k=1}^{K-1} \mathbb{P}\left(\frac{k-1}{K} \leq W < \frac{k}{K}\right) + \frac{1}{K}\mathbb{P}\left(\frac{K-1}{K} \leq W \leq 1\right) = \frac{1}{K}\mathbb{P}(0 \leq W \leq 1) = \frac{1}{K}$$

irrespective of $\mathbb{P} \in \mathcal{P}$. Thus, as the number K of colors tends to infinity, $F_{X_{1:K}}^{\mathbb{P}}$ converges weakly to $F \in \mathcal{D}$ defined through $F(x) = 0$ if $x < 1$; $= 1$ if $x \geq 1$. For any fixed $\mathbb{Q} \in \mathcal{P}$ we then find

$$\lim_{K\uparrow\infty} \rho(X_{1:K}) = \lim_{K\uparrow\infty} \sup_{\mathbb{P}\in\mathcal{P}} \varrho_0(F_{X_{1:K}}^{\mathbb{P}}) = \lim_{K\uparrow\infty} \varrho_0(F_{X_{1:K}}^{\mathbb{Q}}) = \varrho_0(F) = \rho(1),$$

where the first equality follows from Proposition 4, the second equality holds because $F_{X_{1:K}}^{\mathbb{P}}$ is constant across all $\mathbb{P} \in \mathcal{P}$, and the third equality follows from the Lebesgue property of ρ_0. The last equality exploits again Proposition 4 and the fact that F coincides with the distribution function of the degenerate constant random variable 1. By translation invariance, the optimal value of problem (4) thus converges to $\rho(1) = \rho_0(-1) - 2$ as K grows. Hence, for sufficiently large K there is a randomized strategy whose risk is strictly smaller than the risk of the best pure strategy.

To complete the proof, we demonstrate how the existence of the risk reducing randomized strategy $X_{1:K}$ implies the existence of $X, Y \in L_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ and $\theta \in [0, 1]$ such that $\rho(X \oplus_\theta Y) < \min\{\rho(X), \rho(Y)\}$. To this end, note that for any $k = 1, \ldots, K$, the randomized strategy $X_{1:k}$ has the same distribution as some pure strategy $\bar{X}_{1:k} \in \mathcal{L}_\infty(\Omega_0, \mathcal{F}_0, \mathcal{P}_0)$ as well as the mixture $\bar{X}_{1:k-1} \oplus_{(k-1)/k} X_k$. Therefore, there must exist a k for which $\rho(\bar{X}_{1:k-1} \oplus_{(k-1)/k} X_k) <$

$\min\{\rho(X_{1:k-1}), \rho(X_k)\}$. Indeed, if this was not the case, we would have that

$$1 > \rho(X_{1:K}) = \rho(\bar{X}_{1:K-1} \oplus_{(K-1)/K} X_K) \geq \min\{\rho(X_K), \rho(X_{1:K-1})\}$$

$$\geq \min\{\rho(X_K), \rho(X_{K-1}), \rho(X_{1:K-2})\} \geq \ldots \geq \min_{k=1,\ldots,K} \rho(X_k) = 1,$$

where the equations from the first row follow from the fact that $\rho(X_{1:K}) < 1$, the assumed law invariance of ρ and the assumption that $\rho(\bar{X}_{1:K-1} \oplus_{(K-1)/K} X_K) \not< \min\{\rho(X_{1:K-1}), \rho(X_K)\}$, respectively. These steps are then repeated recursively in the second row for all $\rho(X_{1:k}) \geq \min\{\rho(X_k), \rho(X_{1:k-1})\}$ down to $\rho(X_{1:1}) = \rho(X_1)$. We thus obtain the contradiction $1 > 1$, which completes our proof. $\square$