Monotone Additive Statistics*

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

August 31, 2023

Abstract

The expectation is an example of a descriptive statistic that is monotone with respect to stochastic dominance, and additive for sums of independent random variables. We provide a complete characterization of such statistics, and explore a number of applications to models of individual and group decision-making. These include a representation of stationary monotone time preferences, extending the work of Fishburn and Rubinstein (1982) to time lotteries. This extension offers a new perspective on risk attitudes toward time, as well as on the aggregation of multiple discount factors.

1 Introduction

How should a random quantity be summarized by a single number? In Bayesian statistics, point estimators capture an entire posterior distribution. In finance, risk measures quantify the risk in a distribution of returns. And in economics, certainty equivalents characterize an agent's preference for uncertain outcomes.

We use the term *descriptive statistic*, or simply *statistic*, to refer to a map that assigns a number to each bounded random variable. We study statistics that are monotone with respect to first-order stochastic dominance, and additive for sums of independent random variables. An example of a monotone additive statistic is the expectation. The median is monotone but not additive, while the variance is additive, but not monotone.

Monotonicity is a well studied property of statistics (see, e.g., Bickel and Lehmann, 1975a,b), and holds, for example, for certainty equivalents of monotone preferences over

*We would like to thank Kim Border, Sebastian Ebert, Giacomo Lanzani, George Mailath, Meg Meyer, Pietro Ortoleva, Doron Ravid and Weijie Zhong for helpful comments and suggestions.

†Princeton University. Email: xmu@princeton.edu.

‡Caltech. Email: luciano@caltech.edu.

§Yale University. Email: philipp.strack@yale.edu. Philipp Strack was supported by a Sloan fellowship.

¶Caltech. Email: tamuz@caltech.edu. Omer Tamuz was supported by a grant from the Simons Foundation (#419427), a Sloan fellowship, a BSF award (#2018397) and a National Science Foundation CAREER award (DMS-1944153).

lotteries. Additivity is a stronger assumption. We focus on this property because of its conceptual simplicity and because it serves as a baseline assumption in many settings. As we argue, additivity corresponds to stationarity in the context of preferences over time lotteries (see §3). In the context of choices over monetary gambles it corresponds to invariance to background risk (§4.1), or to a form of separability across independent decision problems (§4.7).

Beyond the expectation, an additional example of a monotone additive statistic is the map K_a that, given $a \in \mathbb{R}$, assigns to each random variable X the value

$$K_a(X) = \frac{1}{a} \log \mathbb{E}\left[\mathrm{e}^{aX}\right]. \tag{1}$$

In the fields of probability and statistics, this function is known as the (normalized) cumulant generating function evaluated at a. In finance it is called the entropic risk measure. In economics, it corresponds to the certainty equivalent of an expected utility maximizer who exhibits constant absolute risk aversion (CARA) over gambles. For bounded random variables, the essential minimum and maximum provide further examples of such statistics; as we explain later, they are the limits of K_a as a approaches $\pm\infty$. The expectation is equal to K_0, the limit of K_a as a approaches 0.

Our main result establishes that these examples, and their weighted averages, are the only monotone additive statistics. That is, we show that if a statistic Φ is monotone, additive and normalized so that it satisfies $\Phi(c) = c$ for every constant c, then it is of the form

$$\Phi(X) = \int K_a(X)\,\mathrm{d}\mu(a)$$

for some probability measure μ over the extended real line. This result provides a simple representation of a natural family of statistics, which one may a priori have expected to be much richer.

Our first application is to time preferences. The starting point for our analysis is the work by Fishburn and Rubinstein (1982), who study preferences over dated rewards—a monetary reward, together with the time at which it will be received. They show that exponential discounting of time arises from a set of axioms, of which the most substantial, *stationarity*, postulates that preferences between two dated rewards are unaffected by the addition of a common delay.

We extend the analysis of Fishburn and Rubinstein (1982) to *time lotteries*, which consist of a monetary reward x and a random time T at which it will be received. In this setting, we too introduce a stationarity axiom that requires preferences to be invariant with respect to random independent delays. As we argue in the main text, this stationarity axiom captures a dynamic consistency assumption, together with the idea that preferences do not depend on calendar time.

We show that a monotone and stationary preference over time lotteries admits a representation of the form

$$u(x)\mathrm{e}^{-r\Phi(T)},$$

where Φ is a monotone additive statistic (Theorem 3). Thus, $\Phi(T)$ is the certainty equivalent of the random time T, i.e. the deterministic time that is as desirable as T. Over deterministic dated rewards, the above representation coincides with standard discounted utility. General time lotteries are reduced to deterministic ones through the certainty equivalent Φ. By our main representation theorem, it takes the form $\Phi(T) = \int K_a(T)\, \mathrm{d}\mu(a)$. In this context, each $K_a(T)$ is the certainty equivalent of T under an expected discounted preference with discount rate $-a$. The different certainty equivalents are then averaged according to the measure μ.

In this representation it is as if the decision maker had in mind not one but multiple discount factors. Thus, Φ can be interpreted as the certainty equivalent of a decision maker who is uncertain about the correct discount factor, or as the aggregated certainty equivalent of a group of different discounting agents.

Our representation theorem for monotone and stationary time preferences has implications for understanding the relation between stationarity and risk attitudes toward time. How people choose among prospects that involve risk over time has been studied both theoretically and experimentally (Chesson and Viscusi, 2003; Onay and Öncüler, 2007; Ebert, 2020; DeJarnette et al., 2020; Ebert, 2021). A basic paradox these papers highlight is that many subjects display risk aversion over the time dimension, even though the standard theory of expected discounted utility predicts that people are risk-seeking with respect to time lotteries. Our analysis shows that expected discounted utility is only one of many ways to extend exponential discounting from dated rewards to time lotteries without violating stationarity. In particular, monotone and stationary time preferences can accommodate risk aversion over time, as well as more nuanced preferences that display risk-averse or risk-seeking behavior depending on the choice at hand.

We further apply the characterization of monotone stationary preferences to the problem of aggregating heterogeneous time preferences. It is well known that directly averaging exponential discounting utilities leads to present bias (see Jackson and Yariv, 2014, 2015). Based on this observation, the literature concludes that within expected discounted utility, it is impossible to aggregate individual preferences into a social preference unless the latter is dictatorial.

We show that this difficulty is not due to stationarity, but rather to an insistence on the idea that the social preference should conform to expected discounted utility. When preferences are allowed to belong to the more general class of monotone stationary preferences, then a social preference obtained by averaging the certainty equivalents of the individuals satisfies Pareto efficiency and stationarity. Moreover, every Paretian and stationary social preference is obtained in this way (Theorem 4).

Monotone additive statistics also find applications to models of choice over monetary gambles. It is well known that an expected utility agent whose preferences are invariant to independent background risks must have CARA preferences. This invariance property makes CARA utility functions useful modeling tools when the analyst does not observe the agents' wealth level or the additional risks they face (see, e.g., Barseghyan, Molinari, O'Donoghue, and Teitelbaum, 2018). Beyond expected utility, monotone preferences that are invariant to background risks have certainty equivalents that are monotone additive statistics, and thus, by our representation theorem, are weighted averages of CARA certainty equivalents, where the mixing measure μ is over the coefficient of absolute risk aversion. Hence, in this representation, the decision maker entertains multiple utility functions, each defining a different certainty equivalent. Each lottery is evaluated by averaging over certainty equivalents.

An interesting feature of preferences represented by monotone additive statistics is that they can display behavior that is not uniformly risk-averse nor risk-seeking, such as that of an agent buying both lottery tickets and insurance (Friedman and Savage, 1948), all while maintaining invariance to background risk. At the same time, a potential difficulty for this class of preferences is that their defining parameter, the measure μ over the coefficient of risk aversion, is infinite-dimensional. To narrow down the parameter space, we focus on those preferences that also satisfy *betweenness*, a well-known weakening of the independence axiom that has been extensively studied in the literature (see Dekel, 1986; Gul, 1991). We show that a preference represented by a monotone additive statistic Φ satisfies betweenness if and only if it is of the form

$$\Phi(X) = \beta K_{-a\beta}(X) + (1-\beta)K_{a(1-\beta)}(X).$$

The parameter $\beta \in [0,1]$ controls the the relative weights of the risk-averse and risk-seeking components, with increased β making the decision maker more risk-averse. The parameter a is a scale parameter. This is a simple two-parameter family, but it is rich enough to accommodate preferences that are neither risk-averse nor risk-seeking, while maintaining stationarity.

Our final application concerns group decision-making under risk. We consider a firm that employs multiple agents, each of whom makes decisions following an individual preference relation, which can be seen as a decision rule prescribed by the firm. We show that in order for the agents' independent choices to not violate stochastic dominance when combined, it is sufficient and necessary that their preferences are represented by the same monotone additive statistic. Thus, these are the only preferences with the property that decentralized decisions cannot result in stochastically dominated outcomes for the organization.

1.1 Related Literature

A large literature in statistics studies descriptive statistics of probability distributions. A representative example is the work of Bickel and Lehmann (1975a,b), who study location statistics using an axiomatic, non-parametric approach that is similar to ours. This literature has however focused on different properties, and, to the best of our knowledge, does not contain a similar characterization of additivity and monotonicity. The mathematics literature has studied additive statistics as homomorphisms from the convolution semigroup to the real numbers (see Ruzsa and Székely, 1988; Mattner, 1999, 2004), but without imposing monotonicity.

In finance and actuarial sciences, $-K_a(X)$ is also known as an *entropic risk measure*, and is used to assess the riskiness of a financial position X. It is a canonical example of a coherent risk measure (see Föllmer and Schied, 2002, 2011; Föllmer and Knispel, 2011). In this literature, Goovaerts, Kaas, Laeven, and Tang (2004) study additive statistics that are monotone with respect to all entropic risk measures, i.e. those with the property that $K_a(X) \geq K_a(Y)$ for all $a \in \mathbb{R}$ implies $\Phi(X) \geq \Phi(Y)$, and show that they must be weighted averages of entropic risk measures, as in our main representation. In contrast, we show that this condition is implied by monotonicity and additivity of Φ.

In an earlier paper, Pomatto, Strack, and Tamuz (2020) show that on the domain of random variables that have all moments, the only monotone additive statistic is the expectation. The techniques used there involve fat-tailed random variables, which are used to rule out all other monotone additive statistics.[1] This precludes any study of risk aversion. In contrast, in this paper we primarily study the domain of bounded random variables, which allows for richer preferences with a variety of risk attitudes.

Monotone additive statistics also relate to what we called *additive divergences* in Mu, Pomatto, Strack, and Tamuz (2021). An additive divergence is a map defined over Blackwell experiments that satisfies monotonicity with respect to the Blackwell order and additivity for product experiments. While some of the techniques used in the two papers are similar, the main mathematical argument is fundamentally different. The last step of the proof of our Theorem 1 is the same type of Riesz Representation Theorem argument used in the previous paper. However, because the Blackwell order has different properties from first-order stochastic dominance, the remainder of the proof is different, with the previous

[1] The same phenomenon is studied more in depth in Mu, Pomatto, Philipp, and Tamuz (2023) and Fritz, Mu, and Tamuz (2020). In the latter paper it is shown that the expectation remains the unique monotone additive statistic on the domain L^p, for any $p \geq 1$, while there are no monotone additive statistics on L^p with $p < 1$, or on the domain of all random variables.

paper having no analogue of Theorem 7, which is the main technical step in the proof of Theorem 1. This new technique is also needed for the proof of Theorem 2, which characterizes monotone additive statistics beyond bounded random variables.

DeJarnette, Dillenberger, Gottlieb, and Ortoleva (2020) study preferences over time lotteries that display risk aversion. One class of preferences they propose is a generalization of expected discounted utility (GEDU) that for a random prize X delivered at a random time T takes the form $\mathbb{E}\left[\phi(u(X)\mathrm{e}^{-rT})\right]$ for some strictly increasing transformation ϕ. The curvature of ϕ determines the attitude towards risk. While GEDU satisfies stationarity for deterministic X and T, stationarity does not in general hold once random times T are considered, even with respect to adding a deterministic delay. The only intersection between our model and GEDU are preferences represented by K_a, corresponding to a point-mass mixing measure μ. These preferences have the standard EDU representation, but perhaps with a negative discount rate, as we explain in §3.3.[2]

Applied to choice under risk, our representation also bears resemblance to cautious expected utility theory (Cerreia-Vioglio, Dillenberger, and Ortoleva, 2015), in which a gamble is evaluated according to the minimum certainty equivalent across a family of utility functions. The two representations are conceptually related, as both involve uncertainty about the correct utility function. Our axioms are however different in that we study preferences that are invariant to adding an independent gamble, while Cerreia-Vioglio, Dillenberger, and Ortoleva (2015) consider the effect of mixing with another gamble.

Decision criteria that aggregate multiple certainty equivalents have appeared before in the literature. Myerson and Zambrano (2019) advocate the maximization of a sum of certainty equivalents as an effective rule for risk sharing. Chambers and Echenique (2012) formalize and characterize this rule as a social welfare functional.

The remainder of the paper is organized as follows. In §2 we introduce monotone additive statistics and state our main result. In §3 we apply this result to time lotteries, and in §4 we apply it to monetary gambles. Finally, §5 provides an overview of the proof of our main result. The appendix and online appendix contain omitted proofs for the results in the main text.

2 Monotone Additive Statistics

We denote by L^∞ the collection of bounded real random variables, defined over a nonatomic probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We will identify each $c \in \mathbb{R}$ with the corresponding constant

[2]When the prize X is held constant, a GEDU preference reduces to an expected utility preference over random times. In contrast, our prizes are always deterministic, and the stationary preferences over random times are represented by monotone additive statistics, which are not expected utility unless the mixing measure μ is a point mass (see Proposition 7 in §F of the online appendix).

random variable taking value $X(\omega) = c$ at each $\omega \in \Omega$.

We say that a map $\Phi\colon L^\infty \to \mathbb{R}$ is a *statistic* if it satisfies (i) $\Phi(X) = \Phi(Y)$ whenever $X, Y \in L^\infty$ have the same distribution, and (ii) $\Phi(c) = c$ for every $c \in \mathbb{R}$; that is, Φ assigns c to the constant random variable c. We are interested in statistics that satisfy monotonicity with respect to first-order stochastic dominance and additivity for sums of independent random variables. Formally, Φ is

- *additive* if $\Phi(X+Y) = \Phi(X) + \Phi(Y)$ whenever X and Y are independent, and
- *monotone* if $X \geq_1 Y$ implies $\Phi(X) \geq \Phi(Y)$, where $\geq_1$ denotes first-order stochastic dominance.

Since, by assumption, the value $\Phi(X)$ depends only the distribution of the random variable X, monotonicity is equivalent to the requirement that $\Phi(X) \geq \Phi(Y)$ whenever $X \geq Y$ almost surely. This equivalence is based on the well-known fact that $X \geq_1 Y$ if and only if there are random variables $\tilde{X}, \tilde{Y}$ such that X and $\tilde{X}$ are identically distributed, Y and $\tilde{Y}$ are identically distributed, and $\tilde{X} \geq \tilde{Y}$ almost surely.[3]

We denote by $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ the extended real numbers. Given $X \in L^\infty$ and $a \in \overline{\mathbb{R}} \setminus \{0, \pm\infty\}$, we consider the statistic

$$K_a(X) = \frac{1}{a} \log \mathbb{E}\left[\mathrm{e}^{aX}\right]. \tag{2}$$

The value $K_a(X)$ is the certainty equivalent of X for a CARA utility function with coefficient of risk aversion $-a$. In probability and statistics, $K_a(X)$ is known as the *(normalized) cumulant generating function* of X, evaluated at a. If X and Y are independent, then $\mathbb{E}\left[\mathrm{e}^{a(X+Y)}\right] = \mathbb{E}\left[\mathrm{e}^{aX}\right]\mathbb{E}\left[\mathrm{e}^{aY}\right]$, and hence K_a is additive. It is also monotone.

We additionally define $K_0(X), K_\infty(X), K_{-\infty}(X)$ to be the expectation, the essential maximum, and the essential minimum of X, respectively.[4] This choice of notation makes $a \mapsto K_a(X)$ a continuous function from $\overline{\mathbb{R}}$ to $\mathbb{R}$, for any X. Our main result is a representation theorem for monotone additive statistics:

Theorem 1. *$\Phi\colon L^\infty \to \mathbb{R}$ is a monotone additive statistic if and only if there exists a (unique) Borel probability measure μ on $\overline{\mathbb{R}}$ such that for every $X \in L^\infty$*

$$\Phi(X) = \int_{\overline{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu(a). \tag{3}$$

[3]An alternative, equivalent definition for a statistic is to let the domain of Φ be the set of probability distributions on $\mathbb{R}$ with bounded support. In this domain, additivity would be defined with respect to convolution. We choose to have the domain consist of random variables, as this approach offers some notational advantages.

[4]The essential maximum and minimum are the maximum and minimum of the support: $\max[X] = \sup\{a : \mathbb{P}[X \leq a] < 1\}$ and $\min[X] = \inf\{a : \mathbb{P}[X \leq a] > 0\}$.

Each K_a satisfies monotonicity and additivity, and it is immediate that these two properties are preserved under convex combinations. Theorem 1 says that the one-parameter family $\{K_a\}$ forms the extreme points of the set of monotone additive statistics; every such statistic must be a weighted average obtained by mixing over this family. In §5 we provide an overview of the proof of Theorem 1.

Theorem 1 can be extended to other domains of random variables. We consider the set L_M of random variables X for which $K_a(X)$, as defined in (2), is finite for all $a \in \mathbb{R}$. The domain L_M contains those unbounded random variables whose distributions have sub-exponential tails, as in the case of the normal distribution.

Theorem 2. *$\Phi\colon L_M \to \mathbb{R}$ is a monotone additive statistic if and only if it admits a (unique) representation of the form (3) where the measure μ has compact support in $\mathbb{R}$.*

The extension of Theorem 1 to the larger domain L_M adds to the applicability of our representation, as it includes distributions with unbounded support, such as Gaussian or Poisson, for which the function K_a has closed-form expressions. For example, Theorem 2 implies that when applied to a Gaussian random variable Z, a monotone additive statistic Φ defined on L_M takes the simple mean-variance form $\Phi(Z) = \mathbb{E}[Z] + c\mathrm{Var}[Z]/2$, where $c \in \mathbb{R}$ is the mean of the measure μ characterizing Φ.

A few additional remarks are in order. First, Theorems 1 and 2 answer an open question in the mathematical finance literature on risk measures posed by Goovaerts, Kaas, Laeven, and Tang (2004), who asked if entropic risk measures are the only additive risk measures. Second, a possible strengthening of our additivity condition requires $\Phi(X+Y) = \Phi(X) + \Phi(Y)$ to hold for all pairs of random variables, rather than just the independent ones. As is well known, the only statistic that is additive in this more restrictive sense is the expectation (see, for example, de Finetti, 1970). A different strengthening is additivity with respect to uncorrelated random variables. It follows from the analysis of Chambers and Echenique (2020) that on a finite probability space the expectation is, again, the only monotone statistic that is additive for uncorrelated random variables.

3 Monotone Stationary Time Preferences

Next, we apply monotone additive statistics to the study of time preferences. We consider decision problems where an agent is asked to choose between time lotteries that pay a fixed reward at a future random time, as in the case of a driver choosing between different routes, where some routes are more likely than others to face heavy traffic, or a company choosing between projects with different random completion times. We argue that in this context additivity is connected to a notion of stationarity, according to which a choice between future rewards is not affected by the addition of an independent delay. In this section we study preferences over time lotteries that are monotone and stationary, characterize them using monotone additive statistics, discuss the risk attitudes they can model, and apply them to the problem of aggregating heterogeneous time preferences.

3.1 Domain and Axioms

A *time lottery* is a monetary reward received by a decision maker at a future, random time. Formally, it consists of a pair (x, T), where $x \in \mathbb{R}_{++}$ is a positive payoff and $T \in L^\infty_+$ is the random time at which it realizes.[5] Thus, time is non-negative and continuous. Our primitive is a complete and transitive binary relation $\succeq$ on the domain $\mathbb{R}_{++} \times L^\infty_+$. We denote by $\sim$ the indifference relation induced by $\succeq$. To avoid notational confusion, in the rest of this section x and y always denote monetary payoffs, t, s and d denote deterministic times, and T, S, and D denote random times.

We say that a preference relation $\succeq$ on $\mathbb{R}_{++} \times L^\infty_+$ is a *monotone stationary time preference* (henceforth, MSTP) if it satisfies the following axioms:

Axiom 3.1 (More is Better)**.** *If $x > y$ then $(x, T) \succ (y, T)$.*

Axiom 3.2 (Earlier is Better)**.** *If $s > t$ then $(x, t) \succ (x, s)$, and if $S \geq_1 T$ then $(x, T) \succeq (x, S)$.*

Axiom 3.3 (Stationarity)**.** *If $(x, T) \succeq (y, S)$ then $(x, T + D) \succeq (y, S + D)$ for any D that is independent from T and S.*

Axiom 3.4 (Continuity)**.** *For any (y, S), the sets $\{(x, t) \colon (x, t) \succeq (y, S)\}$ and $\{(x, t) \colon (x, t) \preceq (y, S)\}$ are closed in $\mathbb{R}_{++} \times \mathbb{R}_+$.*

The first two Axioms 3.1 and 3.2 are standard conditions that directly generalize those in Fishburn and Rubinstein (1982), who studied preferences over dated rewards $\{(x, t)\}$ with a deterministic time t. They require the decision maker to prefer higher payoffs, and to prefer earlier times. Axiom 3.4 is a standard continuity assumption that does not require a choice of topology over random times. The most substantive of our axioms is stationarity. In §3.4 we discuss this axiom in depth and motivate it using the notions of time invariance and dynamic consistency (Halevy, 2015).

It is worthwhile to note that we implicitly assume agents to be indifferent with respect to the timing of resolution of uncertainty. We think of the choice as being made at time 0, and we do not distinguish between situations where the realization of the random time T is revealed immediately, gradually until time T, or only at time T. Modeling preferences over the timing of resolution of uncertainty would require enlarging the choice domain beyond time lotteries.

[5] Per standard notation, L^∞_+ denotes the set of non-negative bounded random variables.

3.2 Representation

Our next result characterizes monotone stationary time preferences:

Theorem 3. *A preference relation $\succeq$ over time lotteries is an MSTP if and only if there exist a monotone additive statistic Φ, a constant $r > 0$, and a continuous and increasing function $u\colon \mathbb{R}_{++} \to \mathbb{R}_{++}$ such that $\succeq$ is represented by*

$$V(x,T) = u(x) \cdot \mathrm{e}^{-r\Phi(T)}. \tag{4}$$

As in Fishburn and Rubinstein (1982), the parameter r can be normalized to be any arbitrary positive constant by applying a monotone transformation to the representation V. We will often set r appropriately to simplify the form of the representation. In contrast, the monotone additive statistic Φ is uniquely determined by the preference.

Over the domain of deterministic time lotteries (i.e. dated rewards), V coincides with an exponentially discounted utility representation with discount rate r. For general time lotteries, $\Phi(T)$ is the certainty equivalent of T, i.e. the unique deterministic time that satisfies $(x,T) \sim (x, \Phi(T))$. The monotonicity and continuity axioms ensure that such a certainty equivalent exists, and it is an implication of stationarity that $\Phi(T)$ is independent of the reward x. As we further show in the proof of Theorem 3, the monotonicity and stationarity axioms formally translate into the certainty equivalent Φ being a monotone additive statistic.

Proposition 6 in the appendix shows that the representation in Theorem 1 extends to the domain of non-negative bounded random variables. Thus every MSTP can be represented in the following form:

$$V(x,T) = u(x) \cdot \mathrm{e}^{-r\int K_a(T)\,\mathrm{d}\mu(a)}. \tag{5}$$

We recover expected discounted utility when μ is a point mass concentrated on a point $-a < 0$, in which case Φ takes the form

$$\Phi(T) = K_{-a}(T) = \frac{1}{-a}\,\log \mathbb{E}\left[\mathrm{e}^{-aT}\right].$$

This certainty equivalent, with the normalization $r = a$, yields the familiar representation $V(x,T) = u(x)\mathbb{E}[\mathrm{e}^{-aT}]$. For a general measure μ, the statistic $\Phi(T) = \int K_a(T)\,\mathrm{d}\mu(a)$ aggregates different discount rates by mixing over their corresponding certainty equivalents.

The key feature of the representation (5) is that the average is not over discount factors, but instead over certainty equivalents induced by the discount factors. The resulting representation is behaviorally distinct from expected discounted utility whenever μ is not a point mass. Indeed, as we formally prove in §F in the online appendix, this representation satisfies the independence axiom if and only if μ is a point mass.

3.3 Implications for Risk Attitudes toward Time

Theorem 3 demonstrates that there are many ways to extend discounted utility to the domain of time lotteries, while maintaining stationarity. As is well known, standard expected discounted utility preferences are *risk-seeking* over time, in the sense that a decision maker prefers receiving a reward at a random time T rather than at the deterministic expected time $t = \mathbb{E}[T]$. But other monotone additive statistics lead to stationary time preferences that are not risk-seeking. As an example, for every $a > 0$ the statistic

$$\Phi(T) = K_a(T) = \frac{1}{a}\log \mathbb{E}\left[\mathrm{e}^{aT}\right] \tag{6}$$

leads, with the normalization $r = a$, to the representation

$$V(x,T) = \frac{u(x)}{\mathbb{E}\left[\mathrm{e}^{aT}\right]}, \tag{7}$$

which is in fact *risk-averse* over time. Under this preference, the decision maker applies a *negative* discount rate $-a$ within the monotone additive statistic Φ, and yet is impatient. These two aspects are compatible because in the representation $u(x)\mathrm{e}^{-r\Phi(T)}$ the statistic Φ controls the risk attitude, while the decision maker still prefers receiving prizes earlier rather than later, since Φ appears with a negative coefficient.

Another key distinctive property of monotone stationary time preferences is their flexibility in allowing for risk attitudes that are not uniform across time lotteries. To illustrate this point, consider two decision problems with a fixed common reward $x = \$1000$, where in the first problem the choice is between

(I) receiving the reward after 1 day for sure, versus

(II) receiving the reward immediately with 99% probability and after 100 days with 1% probability.

In the second decision problem the choice is between

(I') receiving the reward after 99 days for sure, versus

(II') receiving the reward immediately with 1% probability and after 100 days with 99% probability.

In both problems, the times at which the safe options I and I' deliver the prize are equal to the expected delay of the lotteries II and II', and thus a decision maker who is globally risk-averse or risk-seeking toward time must either choose the safe options or the risky options in both problems. Nevertheless, it does not seem unreasonable for a person to choose I over II in order to avoid the risk of a long delay, but also choose II' to I', since the time lottery offers at least a chance of avoiding an otherwise very long delay.[6]

Preferences based on monotone additive statistics are not necessarily globally risk-averse or risk-seeking, and can accommodate the aforementioned behavior. For example, the statistic

$$\Phi(T) = \frac{1}{2}K_1(T) + \frac{1}{2}K_{-1}(T) = \frac{1}{2}\log \mathbb{E}\left[\mathrm{e}^{T}\right] - \frac{1}{2}\log \mathbb{E}\left[\mathrm{e}^{-T}\right]$$

leads the decision maker to choose the safe option I in the first problem and the risky option II' in the second.

Empirically, both risk-averse and risk-seeking behavior over time lotteries are observed. For example, the experiment by Ebert (2021) finds that there are risk-seeking and risk-averse subjects: "Overall, therefore, and in contrast to the evidence on wealth risk preferences, there is substantial heterogeneity in preferences toward delay risk." Moreover, DeJarnette, Dillenberger, Gottlieb, and Ortoleva (2020) find that even the same subject often exhibits both risk aversion and risk seeking depending on the choice at hand. In their experiment, out of 5 different choices over time lotteries, only 2.9% of subjects are always risk-seeking and only 12.4% are always risk-averse. Thus 84.7% of subjects exhibit behavior that is sometimes risk-seeking and sometimes risk-averse.[7]

In §4.5 below we provide a detailed analysis of the risk attitudes of preferences represented by monotone additive statistics, including a characterization of those statistics that give rise to mixed risk attitudes, as in the above example.

3.4 Stationarity, Time Invariance and Dynamic Consistency

In the absence of risk, it was shown by Halevy (2015) that stationarity can be understood as the implication of two more basic principles: that preferences are not affected by calendar time, and that the decision maker is dynamically consistent. As we next argue, Axiom 3.3 can be motivated by applying similar reasoning to time lotteries.

We consider an enlarged framework where the decision maker is endowed with a profile $(\succeq_t)$ of preferences over time lotteries, with $\succeq_t$ representing the preference the decision maker expresses at time t. Formally, $\succeq_t$ is a preference over $\mathbb{R}_{++} \times L^\infty_+$, where in the context of $\succeq_t$ the pair (x, T) represents a payoff of x received at time $t + T$. Adapting the definitions from Halevy (2015) to our setting, we define *time invariance* and *dynamic consistency* below:[8]

[6] We are grateful to Weijie Zhong for suggesting this example to us.

[7] See Table 1 in DeJarnette et al. (2020). These percentages are for a treatment with maximal delay of 12 weeks across all questions. DeJarnette et al. (2020) also measured risk preferences for time lotteries with a shorter maximal delay of 5 weeks. In this case an even higher number of 86.8% percent of subjects is neither purely risk-seeking nor purely risk-averse across all choices.

[8] These definitions are slightly different from his, and in particular his (deterministic) dynamic consistency

Definition. *The collection of preferences* $(\succeq_t)$ *satisfies* time invariance *if all the preferences* $\succeq_t$ *are equal.*

Intuitively, if the agent chooses (x, T) over (y, S) at some time t then she makes the same choice at all other times.

Definition. *The collection* $(\succeq_t)$ *satisfies* deterministic dynamic consistency *if, for every pair of time lotteries* (x, T) *and* (y, S)*, and every* $d, t \in \mathbb{R}_+$ *it holds that*

$$(x, T) \succeq_{t+d} (y, S) \textit{ implies } (x, T + d) \succeq_t (y, S + d).$$

That is, the decision maker does not reverse her choice between time t and time $t + d$. Time invariance together with deterministic dynamic consistency imply stationarity with respect to deterministic delays, namely $(x, T) \succeq_t (y, S)$ implies $(x, T + d) \succeq_t (y, S + d)$.

Our next definition proposes a generalization of dynamic consistency to a choice between (x, T) and (y, S) made after a random delay D. What we call *weak dynamic consistency* requires that if, at the random time $t + D$, the decision maker always prefers (x, T) over (y, S), then she would not revert her choice if asked to make the decision at time t for her future self. In general, the realization of the delay D could affect the distributions of S and T faced by the decision maker. Weak dynamic consistency considers only the case where the decision maker always faces the same choice independent of the delay, which mathematically corresponds to D being independent of both S and T.

Definition. *The collection* $(\succeq_t)$ *satisfies* weak dynamic consistency *if, for every pair of time lotteries* (x, T) *and* (y, S)*, every* $t \in \mathbb{R}_+$*, and every* $D \in L^\infty_+$ *independent of* S, T *it holds that*

$$(x, T) \succeq_{t+D} (y, S) \textit{ a.s.} \quad \textit{implies } (x, T + D) \succeq_t (y, S + D).$$

The condition $(x, T) \succeq_{t+D} (y, S)$ a.s. in this definition means that for almost every realization d of D it holds that $(x, T) \succeq_{t+d} (y, S)$. As we record in the next claim, our stationarity axiom is immediately implied by time invariance and weak dynamic consistency.

Claim 1. *Suppose the collection* $(\succeq_t)$ *satisfies time invariance, so that* $\succeq_t = \succeq$ *for every* t*, and also satisfies weak dynamic consistency. Then the preference* $\succeq$ *satisfies stationarity.*

Indeed, by time invariance $(x, T) \succeq_t (y, S)$ implies $(x, T) \succeq_{t+d} (y, S)$ for every realization d of D. Thus by weak dynamic consistency $(x, T) \succeq_t (y, S)$ implies $(x, T + D) \succeq_t (y, S + D)$ whenever D is independent of S, T. It is easy to see that conversely, stationarity also implies weak dynamic consistency under the assumption of time invariance.

axiom is slightly stronger, as his implication is in both directions.

In the above weak dynamic consistency axiom we considered the case where D is independent of S and T, which means that at the delayed time $t + D$ the agent always chooses between the same two time lotteries. A stronger dynamic consistency axiom would impose the same condition, but for an arbitrary delay D that need not be independent of S and T. To make this dependency more explicit we write S_d, T_d for random variables that have the conditional distributions of S, T when conditioning on $D = d$.

Definition. *The collection* $(\succeq_t)$ *satisfies* strong dynamic consistency *if, for every pair of time lotteries* (x, T) *and* (y, S)*, every* $t \in \mathbb{R}_+$*, and every* $D \in L^\infty_+$ *it holds that*

$$(x, T_D) \succeq_{t+D} (y, S_D) \text{ a.s. } \quad \text{implies } (x, T + D) \succeq_t (y, S + D).$$

The distributions of T_D and S_D depend on the value of the delay D. Thus, different values d of D correspond to different decision problems, each involving a choice between (x, T_d) and (y, S_d). Strong dynamic consistency requires that if in each such problem the decision maker always prefers the first option, then she must also prefer the first option from an ex-ante perspective.

Intuitively, strong dynamic consistency requires consistency at different times across different decision problems, while weak dynamic consistency only requires it over the same decision problem. For instance, imagine a traveler who must choose between a train and a flight, which involve travel times S and T respectively, and who does not know the specific day of the month D when they will need to travel. Dynamic consistency compares a traveler who must buy their ticket at the start of the month to one who can make the decision on the day of travel. Weak dynamic consistency applies when the distributions of travel times S and T are not dependent on the day of the month. In this case, it is reasonable to expect the two to make the same choice.

However, if travel times S_D and T_D do depend on the day D, such as during the holiday season, the ex-ante decision problem becomes more difficult. The traveler must take into account the conditional travel times for each day, which might not be cognitively feasible and can result in a violation of strong dynamic consistency.

Under time invariance, strong dynamic consistency immediately implies a strong stationarity axiom:

Axiom 3.5 (Strong Stationarity)**.** *For every pair of time lotteries* (x, T)*,* (y, S) *and every* $D \in L^\infty_+$ *not necessarily independent, if* $(x, T_D) \succeq (y, S_D)$ *a.s. then* $(x, T + D) \succeq (y, S + D)$*.*

Together with monotonicity and continuity, strong stationarity constrains the preference over time to be represented by K_a for some $a \in \overline{\mathbb{R}}$, rather than a general monotone additive statistic Φ under (non-strong) stationarity, as in Theorem 3.

Proposition 1. *An MSTP satisfies strong stationarity if and only if it is represented by*

$$V(x,T) = u(x) \cdot \mathrm{e}^{-rK_a(T)}$$

for some $a \in \overline{\mathbb{R}}$, $r > 0$, *and* $u \colon \mathbb{R}_{++} \to \mathbb{R}_{++}$.

In other terms, the preference over time is either risk-neutral, expected discounted utility, the discounted maximum or minimum, or the negatively discounted preference described in (7). Proposition 1 follows from the fact that strong stationarity, in combination with monotonicity and continuity, implies the classic independence axiom as we discuss in §F of the online appendix. Weak dynamic consistency does not imply the independence axiom and thus allows for a richer set of time preferences.

3.5 Aggregation of Preferences over Time Lotteries

In this section we apply monotone stationary time preferences to collective decision problems. A company making a choice among projects with different expected completion dates, a public agency choosing which research projects to fund, or a family deciding which highway to take, are all examples of social decisions where the alternatives at hand can be seen as time lotteries. In such situations, even if individuals share the same views about the desirability of the possible outcomes, there still exists a need to compromise between different degrees of patience and risk tolerance.

We model this type of problem by studying a group of individuals where each agent, denoted by i, is equipped with a preference relation $\succeq_i$ over time lotteries. These preferences may display different degrees of patience. Following the approach in social choice, we ask how individual preferences can be aggregated into a social preference relation $\succeq$ that is aligned to the individual preferences by the Pareto principle. In this context, the Pareto principle requires that if all individuals agree that one time lottery is better than another, so should the social preference:

Axiom 3.6 (Pareto)**.** *If* $(x,T) \succeq_i (y,S)$ *for every* i, *then* $(x,T) \succeq (y,S)$.

We first consider the case where each preference admits a standard expected discounted utility representation $u(x)\mathbb{E}[\mathrm{e}^{-r_i T}]$, where $u \colon \mathbb{R}_{++} \to \mathbb{R}_{++}$ is a utility function that is increasing, continuous, and common to all agents, and $r_i > 0$ is agent i's discount rate. For simplicity, we follow here the literature on experts' aggregation and focus on the case where agents, and later the social planner, share the same utility function u but may have different discount rates (Weitzman, 2001; Chambers and Echenique, 2018).

As implied by the next result, dictatorship is the only admissible aggregation procedure satisfying the Pareto axiom if one insists that the social preference must conform to expected discounted utility.

Proposition 2. *Let* $(\succeq_1, \ldots, \succeq_n, \succeq)$ *be expected discounted utility preferences over time lotteries, where each* $\succeq_i$ *is represented by* $u(x)\mathbb{E}[\mathrm{e}^{-r_i T}]$ *and* $\succeq$ *is represented by* $u(x)\mathbb{E}[\mathrm{e}^{-rT}]$*. Then, the Pareto axiom is satisfied if and only if* $\succeq_i = \succeq$ *for some agent* i*.*

As the proof shows, this impossibility result is a simple consequence of Harsanyi's theorem (Harsanyi, 1955). Similar impossibility results have been obtained in the setting of preferences over consumption streams (Gollier and Zeckhauser, 2005; Zuber, 2011; Jackson and Yariv, 2014, 2015; Feng and Ke, 2018; Chambers and Echenique, 2018).

The next result offers a solution to this impossibility result. It shows that Paretian aggregation and stationarity are compatible, and do not necessarily result in a dictatorship, if we allow preferences to belong to the larger class of MSTPs.

Theorem 4. *Let* $(\succeq_1, \ldots, \succeq_n, \succeq)$ *be MSTPs, where each* $\succeq_i$ *is represented by* $u(x)\mathrm{e}^{-r_i \Phi_i(T)}$ *and* $\succeq$ *is represented by* $u(x)\mathrm{e}^{-r\Phi(T)}$ *for some monotone additive statistics* (Φ_i) *and* Φ*. Suppose the utility function* u *additionally satisfies* $\lim_{x \to 0} u(x) = 0$ *or* $\lim_{x \to \infty} u(x) = \infty$*.*

Then, the Pareto axiom is satisfied if and only if there exists a probability vector $(\lambda_1, \ldots, \lambda_n)$ *such that*

$$r = \sum_{i=1}^{n} \lambda_i r_i \quad \text{and} \quad r\Phi = \sum_{i=1}^{n} \lambda_i r_i \Phi_i.$$

Thus, under the Pareto axiom, the certainty equivalent Φ of the social preference must be an average of the individual certainty equivalents. The theorem implies that in the special case where individuals have expected discounted utility preferences, we can aggregate preferences without violating stationarity by allowing the social preference to be an MSTP. The key insight is that a linear aggregation of certainty equivalents preserves both stationarity and the Pareto axiom, and that this is the unique way of preserving these properties. This approach complements alternative solutions that have been proposed in the literature to resolve the tension between Paretian aggregation and stationarity.[9]

In Theorem 4, the mild richness assumption on the utility function u implies that for any individual i and any pair of random times S and T, there exist payoffs x and y such that $(x, T) \sim_i (y, S)$. This indifference is used in our argument to deduce the if-and-only-if characterization stated in the result. For a general utility function u whose range may be bounded away from 0 and ∞, our proof still shows that the social certainty equivalent Φ is an average of the individual ones. But we need the additional assumption on u to conclude the same for $r\Phi$, which constrains the social discount rate r.

[9] For example, Feng and Ke (2018) define a different notion of Pareto efficiency that takes into account the preferences of individuals across generations. They show that a standard expected discounted social preference can satisfy this weaker Pareto axiom so long as it is more patient than all the individuals. Chambers and Echenique (2018) study a number of representations that weaken stationarity and generalize expected discounted utility.

4 Preferences Over Gambles

In the theory of risk, CARA utility functions form a restrictive but useful class of expected utility preferences. Their usefulness stems from the analytical tractability of the exponential form, as well as from their invariance properties.

CARA utility functions are invariant to changes in wealth, so that a prospect X is preferred to Y if and only if $X + w$ is preferred to $Y + w$ for all wealth levels w. They are more generally invariant to the addition of background risks: if X is preferred to Y then $X + W$ is preferred to $Y + W$ for every independent random variable W.

This property makes CARA utility functions a good approximation whenever stakes are small. In addition, they are used in empirical settings in which wealth is unknown. For example, when estimating risk preferences from insurance choices, the CARA family "*has the advantage that it implies a household's prior wealth w, which frequently is unobserved, is irrelevant to the household's decisions.*" (Barseghyan, Molinari, O'Donoghue, and Teitelbaum, 2018). The stronger property of invariance to background risk is also important, since households' additional background risks—arising from, say, investments in the stock market or health conditions—may be unobservable.

The invariance properties of CARA utility functions are conceptually distinct from the assumption that preferences obey the axioms of expected utility. In this section, we apply monotone additive statistics to study the general class of preferences that are monotone with respect to stochastic dominance and are invariant to background risk.

4.1 Background-risk Invariant Preferences

We consider a complete and transitive preference relation $\succeq$ over L^∞, interpreted here as the space of monetary gambles. We assume that for every gamble X there exists a unique certainty equivalent $\Phi(X)$ such that $\Phi(X) \sim X$. If the preference $\succeq$ is monotone with respect to first-order stochastic dominance then so is Φ. We say that $\succeq$ is *invariant to background risk* when it has the property that $X \succeq Y$ if and only if $X + Z \succeq Y + Z$ for Z independent of X and Y.

As we now explain, a preference $\succeq$ is monotone and invariant to background risk if and only if its certainty equivalent is a monotone additive statistic. Indeed, invariance implies that $X + Y \sim \Phi(X) + Y$ for any two independent random variables X and Y. Likewise, $Y + \Phi(X) \sim \Phi(Y) + \Phi(X)$. Combining the two indifferences yields

$$X + Y \sim \Phi(X) + \Phi(Y).$$

So, the certainty equivalent of $X + Y$ is given by the sum $\Phi(X) + \Phi(Y)$, and thus Φ is an additive. The converse is immediate to verify.

By Theorem 1, the certainty equivalent Φ of such a preference is a weighted average

$$\Phi(X) = \int K_a(X)\,\mathrm{d}\mu(a)$$

of the certainty equivalents of multiple CARA expected utility agents, where μ is a probability measure over the coefficient of absolute risk aversion. In §F of the online appendix we show that this representation violates the expected utility axioms unless μ is a point mass.

4.2 Risk Aversion

In this section we characterize risk-averse and risk-seeking behavior for preferences that are represented by monotone additive statistics. A preference relation $\succeq$ over gambles is risk-averse if its certainty equivalent Φ satisfies $\Phi(X) \leq \mathbb{E}[X]$ for every gamble X, and risk-seeking if the opposite inequality holds. In the domain of time lotteries, since the decision maker prefers lower waiting times, risk aversion corresponds to the opposite inequality $\Phi(T) \geq \mathbb{E}[T]$ for every random time T.

Risk aversion translates into a property of the support of the corresponding mixing measure μ:

Proposition 3. *A monotone additive statistic satisfies $\Phi(X) \leq \mathbb{E}[X]$ for every $X \in L^\infty$ if and only if*

$$\Phi(X) = \int_{\overline{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu(a)$$

for a Borel probability measure μ supported on $[-\infty, 0]$. Symmetrically, $\Phi(X) \geq \mathbb{E}[X]$ for every X if and only if the measure μ is supported on $[0, \infty]$.

In other words, a risk-averse decision maker ranks gambles by aggregating the certainty equivalents of risk-averse CARA utility functions. For the setting of time lotteries, the second part of the result shows that risk aversion, which corresponds to $\Phi(T) \geq \mathbb{E}[T]$, happens if and only if the measure μ is supported on $[0, \infty]$.[10] Thus, risk aversion toward time lotteries occurs whenever the decision maker aggregates the certainty equivalents of preferences that are represented by $u(x)/\mathbb{E}[\mathrm{e}^{aT}]$, across different $a > 0$. Mixed risk attitude, as discussed in §3.3, occurs when μ assigns positive mass to both negative and positive values of a.

A corollary of Proposition 3 is that an additive statistic Φ is monotone with respect to *second-order* (or any higher-order) stochastic dominance if and only if $\Phi(X) = \int K_a(X)\,\mathrm{d}\mu(a)$ for a probability measure μ supported on $[-\infty, 0]$. To see this, note that

[10] A small twist is that in the time setting, risk aversion only requires $\Phi(T) \geq \mathbb{E}[T]$ for every non-negative bounded random variable T. But our proof shows that Proposition 3 holds without change on the smaller domain L^∞_+. This comment also applies to Theorem 5 below.

monotonicity in higher-order stochastic dominance implies risk aversion and thus constrains the support of μ. Conversely, for each $a \leq 0$, the statistic $K_a(X) = \frac{1}{a} \log \mathbb{E}[e^{aX}]$ satisfies higher-order monotonicity because the function e^{ax} has derivatives of all orders that alternate signs. By linearity, $\int K_a(X) \, d\mu(a)$ is also higher-order monotone whenever μ is supported on $[-\infty, 0]$.

4.3 Higher-Order Risk Aversion

An important property of risk preferences is whether they display first- or second-order risk aversion, which captures the willingness to pay to avoid small risks (see Segal and Spivak, 1990). Formally, a preference with certainty equivalent Φ exhibits k-th order risk aversion if the risk premium for a mean zero lottery X vanishes at the order k, i.e., $\Phi(\varepsilon X)$ is of order ε^k. First-order risk aversion plays a key role in explaining commonly observed decisions such as the insurance of small risks and the demand for full insurance.[11]

We consider risk-averse preferences represented by monotone additive statistics, which as we established in Proposition 3 correspond to μ being supported on $[-\infty, 0]$. Our next result identifies the conditions under which invariance to background risk implies first- or second-order risk aversion.

Proposition 4. *Let $\succeq$ be a risk-averse preference represented by a monotone additive statistic Φ with mixing measure μ. Then*

(i) $\succeq$ exhibits first-order risk aversion if and only if $\mu(\{-\infty\}) > 0$.

(ii) $\succeq$ exhibits second-order risk aversion if and only if $\int |a| d\mu < \infty$.

If $\int |a| d\mu = \infty$ but μ does not have any mass at $-\infty$, then the preference $\succeq$ is neither first- nor second-order risk-averse. Note that by Theorem 2, if one allows for unbounded random variables with sub-exponential tails (such as normal random variables), then the mixing measure μ must have compact support, which implies second-order risk aversion by the above result. Thus, while one might not a priori expect invariance to background risk to be incompatible with first-order risk aversion, this follows as an implication of our characterization of monotone additive statistics on the domain L_M.

4.4 Mixed Risk Aversion

As pointed out in the classical work of Friedman and Savage (1948), it is not uncommon to observe behavior that is neither risk-averse nor risk-seeking, such as that of a person who buys both lottery tickets and insurance. For concreteness, in analogy with our discussion

[11] For example, Borch (1974) states that the prediction that people do not fully insure implied by second order risk-aversion is "against all observation."

in the time domain, consider a decision maker faced with the following two choices. In the first, the choice is between

(I) facing a risk of losing \$100 with probability 1%, or

(II) paying \$1 and being fully insured against that risk.

In the second decision problem the choice is between

(I') paying \$1 dollar for a lottery ticket that yields \$100 with probability 1%, or

(II') not participating in the lottery.

Under expected utility, a utility function that exhibits concavity and convexity across different regions of its domain can rationalize the choices of buying insurance in the first problem and buying the lottery ticket in the second one.[12] However, no such preference can predict such behavior at all wealth levels, let alone be invariant to background risk.

On the other hand, preferences that are represented by monotone additive statistics can accommodate such behavior while at the same time remain invariant to background risk. This is the case, for example, for a preference whose certainty equivalent $\Phi(X)$ takes the form $\Phi(X) = \frac{1}{2}K_{-a}(X) + \frac{1}{2}K_a(X)$, with a mixing measure that puts equal weights on two coefficients of risk aversion a and $-a$. See §4.6 below for additional examples.

4.5 Comparative Risk Attitudes

We now proceed to compare the risk attitudes expressed by different monotone additive statistics. For two preference relations $\succeq_1$ and $\succeq_2$ over gambles, with corresponding certainty equivalents Φ_1 and Φ_2, the preference $\succeq_1$ is *more risk-averse* than $\succeq_2$ if $\Phi_1(X) \leq \Phi_2(X)$ for every gamble $X \in L^\infty$. That is, if the first decision maker assigns to every gamble a lower certainty equivalent. The next proposition characterizes comparative risk aversion for preferences represented by monotone additive statistics:

Theorem 5. *Let $\succeq_1$ and $\succeq_2$ be represented by monotone additive statistics with mixing measures μ_1 and μ_2, respectively. Then $\succeq_1$ is more risk-averse than $\succeq_2$ if and only if*

(i) For every $b > 0$, $\int_{[b,\infty]} \frac{a-b}{a} \, \mathrm{d}\mu_1(a) \leq \int_{[b,\infty]} \frac{a-b}{a} \, \mathrm{d}\mu_2(a)$.

(ii) For every $b < 0$, $\int_{[-\infty,b]} \frac{a-b}{a} \, \mathrm{d}\mu_1(a) \geq \int_{[-\infty,b]} \frac{a-b}{a} \, \mathrm{d}\mu_2(a)$.

[12] This approach, first put forward by Friedman and Savage (1948), has been criticized by Markowitz (1952) for implying a number of implausible predictions. Further, Machina (1982) argued that attitudes towards gambling do not change drastically in response to a change in wealth levels.

The condition that $\succeq_1$ is more risk-averse than $\succeq_2$ is, by definition, equivalent to having the mixing measures μ_1 and μ_2 satisfy $\int f \mathrm{d}\mu_1 \leq \int f \mathrm{d}\mu_2$ for all functions f of the form $f(a) = K_a(X)$, as we vary X. Since $K_a(X)$ increases in the parameter a, then a sufficient condition for $\succeq_1$ being more risk-averse is that μ_2 first-order stochastically dominates μ_1. Intuitively, first-order stochastic dominance suffices because we can think of each Φ_i as the average certainty equivalent of an agent with a random CARA preference drawn from μ_i. So if μ_2 dominates μ_1 then agent 2 is more risk-seeking than agent 1.

First-order stochastic dominance is, however, only a sufficient condition. The reason is that the functions of the form $K_{(\cdot)}(X)$, as we vary X, do not span in their cone the collection of all increasing functions, and hence define a strictly finer stochastic order over the mixing measures. Theorem 5 characterizes this stochastic order in simpler terms, by showing that to check $\int K_a(X)\mathrm{d}\mu_1(a) \leq \int K_a(X)\mathrm{d}\mu_2(a)$ for every gamble X it suffices to check $\int g\mathrm{d}\mu_1 \leq \int g\mathrm{d}\mu_2$ only for the increasing functions g of the form $g(a) = \frac{a-b}{a}1_{a\geq b}$ or $g(a) = -\frac{a-b}{a}1_{a\leq b}$. In other words, the convex cone generated by the set of normalized cumulant generating functions is equal to the convex cone generated by a simple one-parameter family of test functions, a result that might be of independent interest.

For a concrete example that the order characterized by Theorem 5 is strictly finer than first-order stochastic dominance, consider μ_1 to be a point mass at $a = 2$ and μ_2 to have $1/4$ mass at $a = 1$ and $3/4$ mass at $a = 3$. Clearly, neither one first-order dominates the other. Condition (ii) in Theorem 5 is trivially satisfied, whereas condition (i) reduces to $\frac{1}{2}(2-b)^+ \leq \frac{1}{4}(1-b)^+ + \frac{1}{4}(3-b)^+$, which holds because the function $(a-b)^+ = \max\{a-b, 0\}$ is convex in a.

4.6 Betweenness

A disadvantage of the class of preferences represented by monotone additive statistics is that it is large, with the entire measure μ as an infinite-dimensional parameter of the preference. In this section we identify a small subset of such preferences that is indexed by only two parameters, and yet retains enough flexibility to accommodate interesting risk attitudes such as mixed risk aversion.

To this end we study preferences that satisfy the *betweenness* axiom. This well-known property, first studied by Dekel (1986) and Chew (1989), requires that the decision maker's preference over probability distributions displays indifference curves that are straight lines. In comparison, the standard independence axiom (which we study in §F of the online appendix) would additionally require the indifference curves to be parallel to each other.

Given two random variables X and Y, we denote by $X_\lambda Y$ a random variable whose distribution is equal to X with probability $\lambda \in [0,1]$ and equal to Y with probability $1-\lambda$. Formally, $X_\lambda Y$ is any random variable whose distribution is a convex combination that assigns weight λ to the distribution of X and weight $1-\lambda$ to the distribution of Y.

Axiom 4.1 (Betweenness)**.** *For all X, Y and all $\lambda \in (0,1)$, $X \sim Y$ if and only if $X_\lambda Y \sim Y$.*

The betweenness axiom characterizes the following class of preferences:

Proposition 5. *Suppose a preference $\succeq$ on L^∞ is represented by a monotone additive statistic $\Phi(X) = \int_{\mathbb{R}} K_a(X)\,\mathrm{d}\mu(a)$. Then $\succeq$ satisfies the betweenness axiom if and only if*

$$\Phi(X) = \beta K_{-a\beta}(X) + (1-\beta) K_{a(1-\beta)}(X)$$

for some $\beta \in [0,1]$ and $a \in [0,\infty)$.

This family of preferences is much smaller, as it is parameterized by only two numbers. It retains the properties of monotonicity, invariance to background risk, as well as the tractability of the CARA representation. Yet it is versatile enough to describe the kind of mixed risk attitude that leads to buying both insurance and lottery tickets.

The risk-attitude parameter β weights the levels of risk aversion/seeking, with $\beta = 1$ corresponding to pure CARA risk aversion and $\beta = 0$ corresponding to pure CARA risk seeking. For internal β, the preference exhibits mixed risk aversion as guaranteed by the previous Proposition 3. Moreover, a simple calculation shows that for any $\beta \in (0,1)$, such a preference would buy both insurance and lottery tickets of the kind described in §4.4 whenever those gambles entail a small probability of a large loss or gain.[13]

The parameter a is a scale parameter. It can be understood as the scale at which the preference deviates from risk neutrality. For gambles whose sizes are much smaller than $1/a$, the preference is very close to being risk-neutral. While for gambles that vary by much more than $1/a$, behavior will be far from risk-neutral. Changing a amounts to changing units, e.g., increasing a by a factor of 100 is equivalent to measuring money in terms of cents rather than dollars.

4.7 Combined Choices over Gambles

In large organizations, risky prospects are not always chosen through a deliberate, centralized process. Rather, they are combinations of independent choices, often carried out with limited coordination among the different actors.

Consider, for example, a bank that employs two workers. The first is a trader who must choose between two contracts, the Lean Hog futures X and X'. The second is an

[13] It can be shown that if $\beta \neq 0.5$, then lottery tickets and insurance as described in §4.4 are preferred if and only if the probability of gain/loss (0.01 in the example) is smaller than $\min(\beta, 1-\beta)$, and the corresponding gain/loss amount (100 in the example) is sufficiently large. If $\beta = 0.5$, then the same holds for any probability of gain/loss < 0.5, and for any gain/loss amount.

administrator who must choose between two insurance policies Y or Y' for the bank's building. Assuming the first worker chooses X and the second Y, the resulting revenue for the bank is given by the random variable $X + Y$. When the agents face choice problems that belong to independent domains, so that X and X' are stochastically independent from Y and Y', it is natural to ask to what extent coordination is necessary for the organization.

In this section we make this question precise by asking under what conditions the agents' combined choices respect first-order stochastic dominance. Our result shows this is true if and only if individual preferences are identical and represented by a monotone additive statistic. Thus, this is the only class of preferences with the property that choices over independent domains can be decentralized without obvious harm to the organization.

We study the following model. We are given two preference relations $\succeq_1$ and $\succeq_2$ over L^∞, the set of bounded gambles, that are complete and transitive (our result immediately generalizes to three or more agents). As in the example above, we think of each preference relation as describing the choices of a different agent, so that $X \succeq_i X'$ if agent i chooses X over X'. These preferences can be interpreted as being endogenous or as the result of exogenous incentives; for example, the bank trader's preferences could be driven by her contract with the employer.

Our main axiom requires that whenever the two agents face independent decision problems, their choices, when combined, do not violate stochastic dominance:

Axiom 4.2 (Consistency of Combined Choices). *Suppose X, X' are independent of Y, Y'. If $X \succ_1 X'$ and $Y \succ_2 Y'$, then $X' + Y'$ does not strictly dominate $X + Y$ in first-order stochastic dominance.*

If we interpret $\succeq_1$ and $\succeq_2$ as decision-making rules that are determined by the organization, then Axiom 4.2 requires such rules to never result in an outcome that is stochastically dominated. That collective choices should not violate stochastic dominance is clearly a desirable requirement for a rational organization. A similar axiom was first introduced by Rabin and Weizsäcker (2009) in the context of a model of narrow framing.

In addition to this axiom, we assume individual preference relations $\succeq_i$ satisfy basic continuity and monotonicity assumptions:

Axiom 4.3 (Continuity). *If $X \succ Y$ then there exists $\varepsilon > 0$ such that $X \succ Y + \varepsilon$ and $X - \varepsilon \succ Y$.*

Axiom 4.4 (Responsiveness). *$X + \varepsilon \succ_i X$ for every $\varepsilon > 0$.*

We next show that under these axioms, the two preference relations must be represented by monotone additive statistics. Moreover, the statistic must be the same for both agents.

Theorem 6. *Two preference $\succeq_1, \succeq_2$ on L^∞ satisfy Axioms 4.2, 4.3, and 4.4 if and only if there exists a monotone additive statistic that represents both $\succeq_1$ and $\succeq_2$.*

Thus, when individual choices are not coordinated, their combination will, in general, lead to violations of stochastic dominance, even when agents' choices concern independent decision problems. The theorem singles out preferences represented by monotone additive statistics as the only class of preferences that are robust to this lack of coordination.

Theorem 6 admits an alternative interpretation, closely related to the work of Rabin and Weizsäcker (2009) on narrow framing. In their paper, a decision maker faces multiple decisions and engages in "narrow bracketing" by choosing separately, in each problem, according to a fixed preference relation $\succeq$ over gambles. This is a special case of our model where $\succeq = \succeq_1 = \succeq_2$. They show that the decision maker's combined choices result in dominated outcomes whenever $\succeq$ is not wealth invariant (i.e. if $X \succ Y$ but $Y + c \succ X + c$ for some X, Y and $c \in \mathbb{R}$), but leave open the question of characterizing the class of preferences, beyond expected utility, that satisfy Axiom 4.2. Theorem 6 provides a complete characterization of those preferences over gambles for which narrow framing does not lead to dominated choices.

5 Overview of the Proof of Theorem 1

Our approach to the proof of Theorem 1 is via a stochastic order known as the *catalytic stochastic order* (see Fritz, 2017, and references therein). Given $X, Y \in L^\infty$, we say that X dominates Y in the catalytic stochastic order on L^∞ if there exists a $Z \in L^\infty$, independent of X and Y, such that $X + Z$ dominates $Y + Z$ in first-order stochastic dominance.

The applicability of this order to our problem is immediate. If X dominates Y in the catalytic stochastic order then

$$\Phi(X + Z) \geq \Phi(Y + Z)$$

for some Z, independent of X and Y. If Φ is also additive, then $\Phi(X + Z) = \Phi(X) + \Phi(Z)$ and $\Phi(Y + Z) = \Phi(Y) + \Phi(Z)$, and so we have that $\Phi(X) \geq \Phi(Y)$. Thus, any monotone additive Φ is monotone with respect to this order.

Clearly, if $X \geq_1 Y$ then X also dominates Y in the catalytic stochastic order, as one can take $Z = 0$. A priori, one may conjecture that this is also a necessary condition. But as Figure 1 shows, it is easy to give examples of two random variables X and Y that are not ranked with respect to first-order stochastic dominance, but are ranked with respect to the catalytic stochastic order.[14] The random variable X equals 1 with probability 1/3 and 0 with probability 2/3, while Y is uniformly distributed on $[-\frac{3}{5}, \frac{2}{5}]$. As the figure shows, their c.d.f.s are not ranked, and hence they are not ranked in terms of first-order stochastic dominance.[15]

[14] We are indebted to the late Kim Border for helping us construct this example.

[15] Pomatto, Strack, and Tamuz (2020) give examples of random variables X and Y that are not ranked

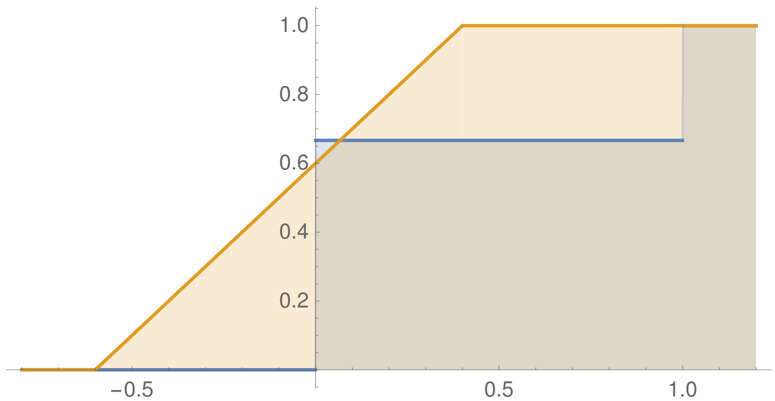

Figure 1: The c.d.f.s of X (blue) and Y (orange).

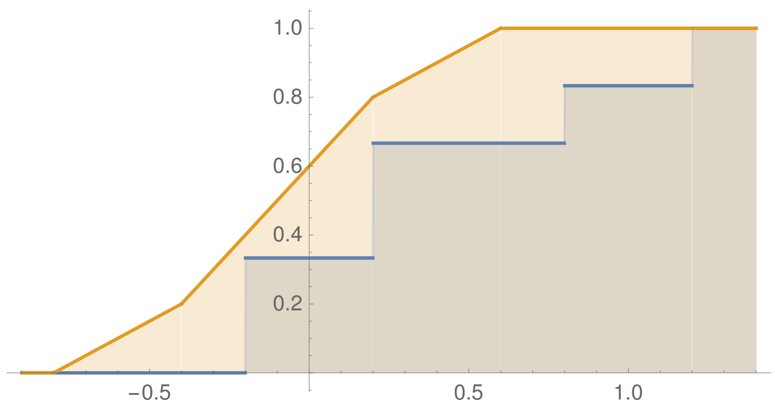

Figure 2: The c.d.f.s of $X + Z$ (blue) and $Y + Z$ (orange).

However, if we let Z assign probability half to $\pm\frac{1}{5}$, then $X + Z >_1 Y + Z$. Intuitively, since the c.d.f. of $X + Z$ is the average of the two translations (by $\pm\frac{1}{5}$) of the c.d.f. of X, and since the same holds for the c.d.f. of Y, the result of adding Z is the disappearance of the small "kink" in which the ranking of the c.d.f.s is reversed. This is depicted in Figure 2.

Every monotone additive statistic Φ provides an obstruction to dominance in the catalytic stochastic order. That is, if $\Phi(X) < \Phi(Y)$ then it is impossible that $X+Z \geq_1 Y+Z$ for some independent Z, since monotonicity would imply that $\Phi(X + Z) \geq \Phi(Y + Z)$, and additivity would then imply that $\Phi(X) \geq \Phi(Y)$. In particular, considering the statistic K_a yields that $K_a(X) \geq K_a(Y)$ for all $a \in \overline{\mathbb{R}}$ is necessary for there to exist some Z that

in stochastic dominance, but are ranked after adding an *unbounded* independent Z. In fact, they show that this is possible whenever $\mathbb{E}[X] > \mathbb{E}[Y]$. As we explain below, this result no longer holds when Z is required to be bounded.

makes X stochastically dominate Y.[16] The following result shows that the statistics K_a are, in a sense, the only obstructions:

Theorem 7. *Let $X, Y \in L^\infty$ satisfy $K_a(X) > K_a(Y)$ for all $a \in \overline{\mathbb{R}}$. Then there exists a c.d.f. H such that any independent $Z \in L^\infty$ with c.d.f. H satisfies $X + Z \geq_1 Y + Z$.*

To prove Theorem 7 we explicitly construct H as a truncated Gaussian c.d.f. with appropriately chosen parameters. The idea behind the proof is as follows. Denote by F and G the c.d.f.s of X and Y, respectively, and suppose that they are supported on $[-N, N]$. Let $h(x) = \frac{1}{\sqrt{2\pi V}}\mathrm{e}^{-\frac{x^2}{2V}}$ be the density of a Gaussian Z. Then the c.d.f.s of $X + Z$ and $Y + Z$ are given by the convolutions $F * h$ and $G * h$, and their difference is equal to

$$\begin{aligned}
[G * h - F * h](y) &= \int_{-N}^{N} [G(x) - F(x)] \cdot h(y - x)\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi V}}\mathrm{e}^{-\frac{y^2}{2V}} \cdot \int_{-N}^{N} \underbrace{[G(x) - F(x)] \cdot \mathrm{e}^{\frac{y}{V}\cdot x}}_{(*)} \cdot \underbrace{\mathrm{e}^{-\frac{x^2}{2V}}}_{(**)}\,\mathrm{d}x
\end{aligned}$$

If we denote $a = \frac{y}{V}$, then by integration by parts, the integral of just $(*)$ is equal to $\frac{1}{a}\left(\mathbb{E}\left[\mathrm{e}^{aX}\right] - \mathbb{E}\left[\mathrm{e}^{aY}\right]\right)$, which is positive by the assumption that $K_a(X) > K_a(Y)$ and is in fact bounded away from zero. The term $(**)$ can be made arbitrarily close to 1—uniformly on the integral domain $[-N, N]$—by making V large. This implies that $[G*h - F*h](y) \geq 0$ for all y, and we further show that the inequality still holds if we modify H by truncating its tails, ensuring that it is in L^∞.

Theorem 7 leads to the following lemma, which is a key component of the proof of Theorem 1:

Lemma 1. *Let $\Phi \colon L^\infty \to \mathbb{R}$ be a monotone additive statistic. If $K_a(X) \geq K_a(Y)$ for all $a \in \overline{\mathbb{R}}$ then $\Phi(X) \geq \Phi(Y)$.*

Proof. Suppose $K_a(X) \geq K_a(Y)$ for all $a \in \overline{\mathbb{R}}$. Let $\hat{X}$, $\hat{Y}$ and Z in L be such that: $\hat{X}$ has the same c.d.f. as $X + \varepsilon$, $\hat{Y}$ has the same c.d.f. as Y, and Z has the c.d.f. obtained by applying Theorem 7 to $\hat{X}$ and $\hat{Y}$. We can indeed apply the theorem, since $K_a(\hat{X}) = K_a(X) + \varepsilon > K_a(Y) = K_a(\hat{Y})$ for all a. Hence, $\hat{X} + Z \geq_1 \hat{Y} + Z$. Thus, by monotonicity of Φ, $\Phi(\hat{X} + Z) \geq \Phi(\hat{Y} + Z)$, and by additivity $\Phi(\hat{X}) \geq \Phi(\hat{Y})$. This means that $\Phi(X) + \varepsilon = \Phi(\hat{X}) \geq \Phi(\hat{Y}) = \Phi(Y)$ for all $\varepsilon > 0$, and hence $\Phi(X) \geq \Phi(Y)$. $\square$

[16] In fact, except for the trivial case where X and Y have the same distribution, this necessary condition can be strengthened to a strict inequality $K_a(X) > K_a(Y)$ for all $a \in \mathbb{R}$. This is because $X + Z \geq_1 Y + Z$ implies the strict inequality $K_a(X + Z) > K_a(Y + Z)$ for finite a whenever $X + Z$ and $Y + Z$ have different distributions. Thus, Theorem 7 below implies that for distributions with different minima and maxima, the condition $K_a(X) > K_a(Y)$ for all $a \in \mathbb{R}$ is both necessary and sufficient for dominance in the catalytic stochastic order.

Once we have established Lemma 1, the remainder of the proof uses functional analysis techniques (in particular the Riesz Representation Theorem) to deduce the integral representation in Theorem 1. See §A in the appendix for the complete proof.

An alternative proof of Lemma 1 can be given based on a different stochastic order, known as the *large numbers order*. Given two random variables X and Y, let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be i.i.d. copies of X and Y, respectively. We say that X dominates Y in large numbers if

$$X_1 + \cdots + X_n \geq_1 Y_1 + \cdots + Y_n$$

for all n large enough. Using large-deviations techniques, it was shown by Aubrun and Nechita (2008) that if $K_a(X) > K_a(Y)$ for all $a \in \overline{\mathbb{R}}$, then X dominates Y in large numbers. This implies Lemma 1 since, by the additivity of Φ, $\Phi(X) \geq \Phi(Y)$ holds if and only if $n\Phi(X) = \Phi(X_1 + \cdots + X_n) \geq \Phi(Y_1 + \cdots + Y_n) = n\Phi(Y)$.

Compared to this alternative argument, our proof of Lemma 1 based on Theorem 7 is self-contained and more elementary. More importantly, (an analogue of) the catalytic stochastic order established in Theorem 7 is essential for studying monotone additive statistics defined on a domain of unbounded random variables, for which the large numbers order is difficult to characterize as far as we know.[17] This generalization of Theorem 7 is presented in Lemma 7 in the online appendix, as a key step toward the proof of Theorem 2.

[17] One particular challenge is that the large numbers order require a uniform comparison between the tail probabilities of $X_1 + \cdots + X_n$ versus those of $Y_1 + \cdots + Y_n$, for a fixed large n. For a given threshold of the tail, large-deviations theory can be used to show the desired comparison when n is large enough. But making the required n uniform across all thresholds becomes nontrivial when the random variables X and Y are unbounded.

References

C. D. Aliprantis and K. Border. *Infinite dimensional analysis: A hitchhiker's guide*. Springer, 2006.

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

L. Barseghyan, F. Molinari, T. O'Donoghue, and J. C. Teitelbaum. Estimating risk preferences in the field. *Journal of Economic Literature*, 56(2):501–564, 2018.

P. Bickel and E. Lehmann. Descriptive statistics for nonparametric models I. Introduction. *Annals of Statistics*, 3(5):1038–1044, 1975a.

P. Bickel and E. Lehmann. Descriptive statistics for nonparametric models II. Location. *Annals of Statistics*, 3(5):1045–1069, 1975b.

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

K. H. Borch. *The mathematical theory of insurance: An annotated selection of papers on insurance published 1960-1972*. Lexington Books, 1974.

S. Cerreia-Vioglio, D. Dillenberger, and P. Ortoleva. Cautious expected utility and the certainty effect. *Econometrica*, 83(2):693–728, 2015.

C. P. Chambers and F. Echenique. When does aggregation reduce risk aversion? *Games and Economic Behavior*, 76(2):582–595, 2012.

C. P. Chambers and F. Echenique. On multiple discount rates. *Econometrica*, 86(4): 1325–1346, 2018.

C. P. Chambers and F. Echenique. Spherical preferences. *Journal of Economic Theory*, 189:105086, 2020.

H. W. Chesson and W. K. Viscusi. Commonalities in time and ambiguity aversion for long-term risks. *Theory and Decision*, 54(1):57–71, 2003.

S. H. Chew. Axiomatic utility theories with the betweenness property. *Annals of Operations Research*, 19(1):273–298, 1989.

J. H. Curtiss. A note on the theory of moment generating functions. *The Annals of Mathematical Statistics*, 13(4):430–433, 1942.

B. de Finetti. *Theory of probability*. Wiley, 1970.

P. DeJarnette, D. Dillenberger, D. Gottlieb, and P. Ortoleva. Time lotteries and stochastic impatience. *Econometrica*, 88(2):619–656, 2020.

E. Dekel. An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom. *Journal of Economic Theory*, 40:304–318, 1986.

S. Ebert. Decision making when things are only a matter of time. *Operations Research*, 2020.

S. Ebert. Prudent discounting: Experimental evidence on higher-order time risk preferences. *International Economic Review*, 2021.

T. Feng and S. Ke. Social discounting and intergenerational pareto. *Econometrica*, 86(5): 1537–1567, 2018.

P. C. Fishburn and A. Rubinstein. Time preference. *International Economic Review*, 23: 677–694, 1982.

H. Föllmer and T. Knispel. Entropic risk measures: Coherence vs. convexity, model ambiguity and robust large deviations. *Stochastics and Dynamics*, 11(02n03):333–351, 2011.

H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and stochastics*, 6(4):429–447, 2002.

H. Föllmer and A. Schied. *Stochastic finance: An introduction in discrete time*. Walter de Gruyter, 2011.

M. Friedman and L. J. Savage. The utility analysis of choices involving risk. *Journal of political Economy*, 56(4):279–304, 1948.

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

T. Fritz, X. Mu, and O. Tamuz. Monotone homomorphisms on convolution semigroups, 2020. Working Paper.

C. Gollier and R. Zeckhauser. Aggregation of heterogeneous time preferences. *Journal of Political Economy*, 113(4):878–896, 2005.

M. J. Goovaerts, R. Kaas, R. J. Laeven, and Q. Tang. A comonotonic image of independence for additive risk measures. *Insurance: Mathematics and Economics*, 35(3):581–594, 2004.

F. Gul. A theory of disappointment aversion. *Econometrica*, 59(3):667–686, 1991.

Y. Halevy. Time consistency: Stationarity and time invariance. *Econometrica*, 83(1): 335–352, 2015.

J. C. Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of political economy*, 63(4):309–321, 1955.

M. O. Jackson and L. Yariv. Present bias and collective dynamic choice in the lab. *American Economic Review*, 104(12):4184–4204, 2014.

M. O. Jackson and L. Yariv. Collective dynamic choice: the necessity of time inconsistency. *American Economic Journal: Microeconomics*, 7(4):150–78, 2015.

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

M. J. Machina. “expected utility” analysis without the independence axiom. *Econometrica*, 50(2):277–323, 1982.

H. Markowitz. The utility of wealth. *Journal of political Economy*, 60(2):151–158, 1952.

L. Mattner. What are cumulants? *Documenta Mathematica*, 4:601–622, 1999.

L. Mattner. Cumulants are universal homomorphisms into Hausdorff groups. *Probability theory and related fields*, 130(2):151–166, 2004.

X. Mu, L. Pomatto, P. Strack, and O. Tamuz. From Blackwell dominance in large samples to Rényi divergences and back again. *Econometrica*, 89(1):475–506, 2021.

X. Mu, L. Pomatto, S. Philipp, and O. Tamuz. Background risk and small-stakes risk aversion. 2023. Working Paper.

R. B. Myerson and E. Zambrano. *Probability models for economic decisions*. MIT Press, 2019.

S. Onay and A. Öncüler. Intertemporal choice under timing risk: An experimental approach. *Journal of Risk and Uncertainty*, 34(2):99–121, 2007.

L. Pomatto, P. Strack, and O. Tamuz. Stochastic dominance under independent noise. *Journal of Political Economy*, 128(5):1877–1900, 2020.

M. Rabin and G. Weizsäcker. Narrow bracketing and dominated choices. *American Economic Review*, 99(4):1508–43, 2009.

I. Ruzsa and G. J. Székely. *Algebraic probability theory*. John Wiley & Sons Inc, 1988.

U. Segal and A. Spivak. First order versus second order risk aversion. *Journal of Economic Theory*, 51(1):111–125, 1990.

M. L. Weitzman. Gamma discounting. *American Economic Review*, 91(1):260–271, 2001.

L. Zhou. Harsanyi's utilitarianism theorems: general societies. *Journal of Economic Theory*, 72(1):198–207, 1997.

S. Zuber. Can social preferences be both stationary and Paretian? *Annals of Economics and Statistics*, 101/102:347–360, 2011.

Appendix

The appendix contains the omitted proofs for most of the results that have been explicitly stated in the main text, in the order in which they appeared. The only exceptions are Theorem 2 on the larger domain L_M, Proposition 1 on strong stationarity and Proposition 5 on betweenness, whose proofs are relegated to the online appendix.

In the proofs we often use the notation $K_X(a) = K_a(X)$, so that K_X is a map from $\overline{\mathbb{R}}$ to $\mathbb{R}$. The following facts are standard:

Lemma 2. *Let $X, Y \in L^\infty$.*

1. *$K_X : \overline{\mathbb{R}} \to \mathbb{R}$ is well defined, non-decreasing and continuous.*
2. *If $K_X = K_Y$ then X and Y have the same distribution.*

Proof. See Curtiss (1942). □

A Proof of Theorem 1

We follow the proof outlined in §5 of the main text and first establish Theorem 7.

A.1 Proof of Theorem 7

First, we can add the same constant b to both X and Y so that $\min[Y + b] = -N$ and $\max[X + b] = N$ for some $N > 0$. Since translating both X and Y leaves the existence of an appropriate Z unchanged (and also does not affect $K_X > K_Y$), we henceforth assume without loss of generality that $\min[Y] = -N$, and $\max[X] = N$. Since $K_X > K_Y$, we know that $\min[X] > -N$ and $\max[Y] < N$.

Denote the c.d.f.s of X and Y by F and G, respectively. Let $\sigma(x) = G(x) - F(x)$. Note that σ is supported on $[-N, N]$ and bounded in absolute value by 1. Moreover, by choosing $\varepsilon > 0$ sufficiently small, we have that $\min[X] > -N + \varepsilon$ and $\max[Y] < N - \varepsilon$. So $\sigma(x)$ is positive on $[-N, -N + \varepsilon]$ and on $[N - \varepsilon, N]$. In fact, there exists $\delta > 0$ such that $\sigma(x) \geq \delta$ whenever $x \in [-N + \frac{\varepsilon}{4}, -N + \frac{\varepsilon}{2}]$ and $x \in [N - \frac{\varepsilon}{2}, N - \frac{\varepsilon}{4}]$. We also fix a large constant A such that

$$\mathrm{e}^{\frac{\varepsilon A}{4}} \geq \frac{8N}{\varepsilon \delta}.$$

Define

$$M_\sigma(a) = \int_{-N}^{N} \sigma(x) \mathrm{e}^{ax} \, \mathrm{d}x.$$

Note that for $a \neq 0$, integration by parts shows $M_\sigma(a) = \frac{1}{a}\left(\mathbb{E}\left[\mathrm{e}^{aX}\right] - \mathbb{E}\left[\mathrm{e}^{aY}\right]\right)$, and that $M_\sigma(0) = \mathbb{E}[X] - \mathbb{E}[Y]$. Therefore, since $K_X > K_Y$, we have that M_σ is strictly positive everywhere. Since $M_\sigma(a)$ is clearly continuous in a, it is in fact bounded away from zero on any compact interval.

We will use these properties of σ to construct a truncated Gaussian density h such that

$$[\sigma * h](y) = \int_{-N}^{N} \sigma(x) h(y - x) \, \mathrm{d}x \geq 0$$

for each $y \in \mathbb{R}$. If we let Z be a random variable independent from X and Y, whose distribution has density function h, then $\sigma * h = (G - F) * h$ is the difference between the c.d.f.s of $Y + Z$ and $X + Z$. Thus $[\sigma * h](y) \geq 0$ for all y would imply $X + Z \geq_1 Y + Z$.

To do this, we write $h(x) = \mathrm{e}^{-\frac{x^2}{2V}}$ for all $|x| \leq T$, where V is the variance and T is the truncation point to be chosen.[18] We will show that given the above constants N and A, $[\sigma * h](y) \geq 0$ holds for each y when V is sufficiently large and $T \geq AV + N$.

First consider the case where $y \in [-AV, AV]$. In this region, $|y - x| \leq T$ is automatically satisfied when $x \in [-N, N]$. So we can compute the convolution $\sigma * h$ as follows:

$$\int \sigma(x) h(y - x) \, \mathrm{d}x = \mathrm{e}^{-\frac{y^2}{2V}} \cdot \int_{-N}^{N} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V} \cdot x} \cdot \mathrm{e}^{-\frac{x^2}{2V}} \, \mathrm{d}x. \tag{8}$$

Note that $\frac{y}{V}$ in the exponent belongs to the compact interval $[-A, A]$. So for our fixed choice of A, the integral $M_\sigma(\frac{y}{V}) = \int_{-N}^{N} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V} \cdot x} \, \mathrm{d}x$ is uniformly bounded away from zero when y varies in the current region. Thus,

$$\begin{aligned} \int_{-N}^{N} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V} \cdot x} \cdot \mathrm{e}^{-\frac{x^2}{2V}} \, \mathrm{d}x &= M_\sigma\left(\frac{y}{V}\right) - \int_{-N}^{N} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V} \cdot x} \cdot (1 - \mathrm{e}^{-\frac{x^2}{2V}}) \, \mathrm{d}x \\ &\geq M_\sigma\left(\frac{y}{V}\right) - 2N \cdot \mathrm{e}^{AN} \cdot (1 - \mathrm{e}^{\frac{-N^2}{2V}}), \end{aligned} \tag{9}$$

which is positive when V is sufficiently large. So the right-hand side of (8) is positive.

[18] In general we need a normalizing factor to ensure h integrates to one, but this multiplicative constant does not affect the argument.

Next consider the case where $y \in (AV, T+N-\varepsilon]$; the case where $-y$ is in this range can be treated symmetrically. Here the convolution can be written as

$$[\sigma * h](y) = \int_{\max\{-N, y-T\}}^{N} \sigma(x) \cdot \mathrm{e}^{\frac{-(y-x)^2}{2V}} \, \mathrm{d}x.$$

We break the range of integration into two sub-intervals: $I_1 = [\max\{-N, y-T\}, N-\varepsilon]$ and $I_2 = [N-\varepsilon, N]$. On I_1 we have $\sigma(x) = G(x) - F(x) \geq -1$. As long as $AV \geq N - \varepsilon$, we have $\mathrm{e}^{\frac{-(y-x)^2}{2V}} \leq \mathrm{e}^{\frac{-(y-N+\varepsilon)^2}{2V}}$ for $y > AV$ and $x \leq N - \varepsilon$, and thus

$$\int_{x \in I_1} \sigma(x) \cdot \mathrm{e}^{\frac{-(y-x)^2}{2V}} \, \mathrm{d}x \geq -2N \cdot \mathrm{e}^{\frac{-(y-N+\varepsilon)^2}{2V}}.$$

On I_2 we have $\sigma(x) \geq 0$ by our choice of ε, and furthermore $\sigma(x) \geq \delta$ when $x \in [N - \frac{\varepsilon}{2}, N - \frac{\varepsilon}{4}]$. Thus

$$\int_{x \in I_2} \sigma(x) \cdot \mathrm{e}^{\frac{-(y-x)^2}{2V}} \, \mathrm{d}x \geq \frac{\varepsilon}{4} \cdot \delta \cdot \mathrm{e}^{\frac{-(y-N+\frac{\varepsilon}{2})^2}{2V}} \geq 2N \cdot \mathrm{e}^{\frac{-(y-N+\frac{\varepsilon}{2})^2}{2V} - \frac{\varepsilon A}{4}},$$

where the second inequality holds by the choice of A. Observe that when $y > AV$ and V is large, the exponent $\frac{-(y-N+\frac{\varepsilon}{2})^2}{2V} - \frac{\varepsilon A}{4}$ is larger than $\frac{-(y-N+\varepsilon)^2}{2V}$. Summing the above two inequalities then yields the desired result that $[\sigma * h](y) \geq 0$.

Finally, if $y \in (T+N-\varepsilon, T+N]$, then the range of integration in computing $[\sigma * h](y)$ is from $x = y - T$ to $x = N$, where $\sigma(x)$ is always positive. So the convolution is positive. And if $y > T + N$, then clearly the convolution is zero. These arguments symmetrically apply to $-y \in (T+N-\varepsilon, T+N]$ and $-y > T+N$. We therefore conclude that $[\sigma * h](y) \geq 0$ for all y, completing the proof.

A.2 Integral Representation

For fixed X, $K_X(a) = K_a(X)$ is a function of a, from $\overline{\mathbb{R}}$ to $\mathbb{R}$. Let $\mathcal{L}$ denote the set of functions $\{K_X : X \in L^\infty\}$. If Φ is a monotone additive statistic and $K_X = K_Y$, then X and Y have the same distribution and $\Phi(X) = \Phi(Y)$. Thus there exists some functional $F \colon \mathcal{L} \to \mathbb{R}$ such that $\Phi(X) = F(K_X)$. It follows from the additivity of Φ and the additivity of K_a that F is additive: $F(K_X + K_Y) = F(K_X) + F(K_Y)$.[19] Moreover, F is monotone in the sense that $F(K_X) \geq F(K_Y)$ whenever $K_X \geq K_Y$ (i.e., $K_X(a) \geq K_Y(a)$ for all $a \in \overline{\mathbb{R}}$); this follows from Lemma 1 which in turn is proved by Theorem 7 (see §5 in the main text).

The rest of this proof is a functional analysis exercise analogous to the proof of Theorem 2 in Mu, Pomatto, Strack, and Tamuz (2021), but for completeness we provide the details below. The main goal is to show that the monotone additive functional F on $\mathcal{L}$ can be

[19] We note that $\mathcal{L}$ is closed under addition. This is because $K_X + K_Y = K_{X'} + K_{Y'}$ whenever X', Y' are independently distributed random variables with the same distribution as X, Y. Such random variables X', Y' exist as the probability space is non-atomic, see for example Proposition 9.1.11 in Bogachev (2007). Thus, for $K_X, K_Y \in \mathcal{L}$ we can find X', Y' so that $K_X + K_Y = K_{X'} + K_{Y'} = K_{X'+Y'} \in \mathcal{L}$.

extended to a positive linear functional on the entire space of continuous functions $\mathcal{C}(\overline{\mathbb{R}})$. We first equip $\mathcal{L}$ with the sup-norm of $\mathcal{C}(\overline{\mathbb{R}})$ and establish a technical claim.

Lemma 3. *$F\colon \mathcal{L} \to \mathbb{R}$ is 1-Lipschitz:*

$$|F(K_X) - F(K_Y)| \leq \|K_X - K_Y\|.$$

Proof. Let $\|K_X - K_Y\| = \varepsilon$. Then $K_{X+\varepsilon} = K_X + \varepsilon \geq K_Y$. Hence, by Lemma 1, $F(K_Y) \leq F(K_{X+\varepsilon})$, and so

$$F(K_Y) - F(K_X) \leq F(K_{X+\varepsilon}) - F(K_X) = F(K_\varepsilon) = \Phi(\varepsilon) = \varepsilon.$$

Symmetrically we have $F(K_X) - F(K_Y) \leq \varepsilon$, as desired. $\square$

Lemma 4. *Any monotone additive functional F on $\mathcal{L}$ can be extended to a positive linear functional on $\mathcal{C}(\overline{\mathbb{R}})$.*

Proof. First consider the rational cone spanned by $\mathcal{L}$:

$$\mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}) = \{qL : q \in \mathbb{Q}_+, L \in \mathcal{L}\}.$$

Define $G\colon \mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}) \to \mathbb{R}$ as $G(qL) = qF(L)$, which is an extension of F. The functional G is well defined: If $\frac{m}{n}K_1 = \frac{r}{n}K_2$ for $K_1, K_2 \in \mathcal{L}$ and $n, m, r \in \mathbb{N}$, then, using the fact that $\mathcal{L}$ is closed under addition, we obtain $mF(K_1) = F(mK_1) = F(rK_2) = rF(K_2)$, hence $\frac{m}{n}F(K_1) = \frac{r}{n}F(K_2)$. G is also additive, because

$$G\left(\frac{m}{n}K_1\right) + G\left(\frac{r}{n}K_2\right) = \frac{m}{n}F\left(K_1\right) + \frac{r}{n}F\left(K_2\right) = \frac{1}{n}F(mK_1 + rK_2) = G\left(\frac{m}{n}K_1 + \frac{r}{n}K_2\right).$$

In the same way we can show G is positively homogeneous over $\mathbb{Q}_+$ and monotone.

Moreover, G is Lipschitz: Lemma 3 implies

$$\left|G\left(\frac{m}{n}K_1\right) - G\left(\frac{r}{n}K_2\right)\right| = \frac{1}{n}\left|F(mK_1) - F\left(rK_2\right)\right| \leq \frac{1}{n}\left\|mK_1 - rK_2\right\| = \left\|\frac{m}{n}K_1 - \frac{r}{n}K_2\right\|.$$

Thus G can be extended to a Lipschitz functional H defined on the closure of $\mathrm{Cone}_{\mathbb{Q}}(\mathcal{L})$ with respect to the sup norm. In particular, H is defined on the convex cone spanned by $\mathcal{L}$:

$$\mathrm{Cone}(\mathcal{L}) = \{\lambda_1 K_1 + \cdots + \lambda_k K_k : k \in \mathbb{N} \text{ and for each } 1 \leq i \leq k, \lambda_i \in \mathbb{R}_+, K_i \in \mathcal{L}\}.$$

It is immediate to verify that the properties of additivity, positive homogeneity (now over $\mathbb{R}_+$), and monotonicity extend, by continuity, from G to H.

Consider the vector subspace $\mathcal{V} = \mathrm{Cone}(\mathcal{L}) - \mathrm{Cone}(\mathcal{L}) \subset \mathcal{C}(\overline{\mathbb{R}})$ and define $I\colon \mathcal{V} \to \mathbb{R}$ as

$$I(g_1 - g_2) = H(g_1) - H(g_2)$$

for all $g_1, g_2 \in \mathrm{Cone}(\mathcal{L})$. The functional I is well defined and linear (because H is additive and positively homogeneous). Moreover, by monotonicity of H, $I(f) \geq 0$ for any non-negative function $f \in \mathcal{V}$.

The lemma then follows from the next theorem of Kantorovich (1937), a generalization of the Hahn-Banach Theorem. It applies not only to $\mathcal{C}(\overline{\mathbb{R}})$ but to any Riesz space (see Theorem 8.32 in Aliprantis and Border, 2006).

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

The "majorization" condition $g \geq f$ is satisfied because every function in $\mathcal{C}(\overline{\mathbb{R}})$ is bounded and $\mathcal{V}$ contains all of the constant functions. □

The integral representation in Theorem 1 now follows from Lemma 4 by the Riesz-Markov-Kakutani Representation Theorem.

A.3 Uniqueness of Mixing Measure

We complete the proof of Theorem 1 by showing that the mixing measure μ is unique:

Lemma 5. *Suppose μ and ν are two Borel probability measures on $\overline{\mathbb{R}}$ such that*

$$\int_{\overline{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu(a) = \int_{\overline{\mathbb{R}}} K_a(X)\,\mathrm{d}\nu(a).$$

for all $X \in L^\infty$.[20] *Then $\mu = \nu$.*

Proof. We first show $\mu(\{\infty\}) = \nu(\{\infty\})$. For any $\varepsilon > 0$, consider the Bernoulli random variable X_ε that takes value 1 with probability ε. It is easy to see that as ε decreases to zero, $K_a(X_\varepsilon)$ also decreases to zero for each $a < \infty$ whereas $K_\infty(X_\varepsilon) = \max[X_\varepsilon] = 1$. Since $K_a(X_\varepsilon)$ is uniformly bounded in $[0, 1]$, the Dominated Convergence Theorem implies

$$\lim_{\varepsilon \to 0} \int_{\overline{\mathbb{R}}} K_a(X_\varepsilon)\,\mathrm{d}\mu(a) = \mu(\{\infty\}).$$

A similar identity holds for the measure ν, so $\mu(\{\infty\}) = \nu(\{\infty\})$ follows from the assumption that $\int_{\overline{\mathbb{R}}} K_a(X_\varepsilon)\,\mathrm{d}\mu(a) = \int_{\overline{\mathbb{R}}} K_a(X_\varepsilon)\,\mathrm{d}\nu(a)$.

We can symmetrically apply the above argument to the Bernoulli random variable that takes value 1 with probability $1 - \varepsilon$. Thus $\mu(\{-\infty\}) = \nu(\{-\infty\})$ holds as well.

Next, for each $n \in \mathbb{N}_+$ and real number $b > 0$, define a random variable $X_{n,b}$ by

$$\mathbb{P}[X_{n,b} = n] = \mathrm{e}^{-bn}$$
$$\mathbb{P}[X_{n,b} = 0] = 1 - \mathrm{e}^{-bn}.$$

[20] The proof shows that it suffices to require such equality for non-negative integer-valued X.

Then $K_a(X_{n,b}) = \frac{1}{a}\log\left[(1-\mathrm{e}^{-bn}) + \mathrm{e}^{(a-b)n}\right]$, and so

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}K_a(X_{n,b}) &= \lim_{n\to\infty}\frac{1}{n}\frac{1}{a}\log\left[1-\mathrm{e}^{-bn}+\mathrm{e}^{(a-b)n}\right]\\ &= \begin{cases}0 & \text{if } a<b\\ \frac{a-b}{a} & \text{if } a\geq b.\end{cases}\end{aligned}$$

This result holds also for $a = 0, \pm\infty$.

Note that $\frac{1}{n}K_a(X_{n,b})$ is uniformly bounded in $[0,1]$ for all values of n, b, a, since $K_a(X_{n,b})$ is bounded between $\min[X_{n,b}] = 0$ and $\max[X_{n,b}] = n$. Thus, by the Dominated Convergence Theorem,

$$\lim_{n\to\infty}\int_{\overline{\mathbb{R}}}\frac{1}{n}K_a(X_{n,b})\,\mathrm{d}\mu(a) = \int_{[b,\infty]}\frac{a-b}{a}\,\mathrm{d}\mu(a), \tag{10}$$

and similarly for ν. It follows that for all $b > 0$,

$$\int_{[b,\infty]}\frac{a-b}{a}\,\mathrm{d}\mu(a) = \int_{[b,\infty]}\frac{a-b}{a}\,\mathrm{d}\nu(a).$$

As $\mu(\{\infty\}) = \nu(\{\infty\})$, we in fact have

$$\int_{[b,\infty)}\frac{a-b}{a}\,\mathrm{d}\mu(a) = \int_{[b,\infty)}\frac{a-b}{a}\,\mathrm{d}\nu(a).$$

This common integral is denoted by $f(b)$.

We now define a measure $\hat{\mu}$ on $(0,\infty)$ by the condition $\frac{\mathrm{d}\hat{\mu}(a)}{\mathrm{d}\mu(a)} = \frac{1}{a}$; note that $\hat{\mu}$ is a positive measure, but need not be a probability measure. Then

$$f(b) = \int_{[b,\infty)}\frac{a-b}{a}\,\mathrm{d}\mu(a) = \int_{[b,\infty)}(a-b)\,\mathrm{d}\hat{\mu}(a) = \int_b^\infty \hat{\mu}([x,\infty))\,\mathrm{d}x,$$

where the last step uses Tonelli's Theorem. Hence $\hat{\mu}([b,\infty])$ is the negative of the left derivative of $f(b)$ (this uses the fact that $\hat{\mu}([b,\infty])$ is left continuous in b). In the same way, if we define $\hat{\nu}$ by $\frac{\mathrm{d}\hat{\nu}(a)}{\mathrm{d}\nu(a)} = \frac{1}{a}$, then $\hat{\nu}([b,\infty])$ is also the negative of the left derivative of $f(b)$. Therefore $\hat{\mu}$ and $\hat{\nu}$ are the same measure on $(0,\infty)$, which implies that μ and ν coincide on $(0,\infty)$.

By a symmetric argument (with $n - X_{n,b}$ in place of $X_{n,b}$), we deduce that μ and ν also coincide on $(-\infty, 0)$. Finally, since they are both probability measures, μ and ν must have the same mass at 0, if any. So $\mu = \nu$. □

B Applications to Time Lotteries

B.1 Monotone Additive Statistics for Non-Negative Random Variables

In our applications to time lotteries the random times are non-negative (bounded) random variables. We accordingly prove a version of Theorem 1 that applies to this smaller domain.

Proposition 6. $\Phi\colon L^\infty_+ \to \mathbb{R}$ *is a monotone additive statistic if and only if there exists a unique Borel probability measure* μ *on* $\overline{\mathbb{R}}$ *such that for every* $X \in L^\infty$

$$\Phi(X) = \int_{\overline{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu(a). \tag{11}$$

Proof. It suffices to show that a monotone additive statistic defined on L^∞_+ can be extended to a monotone additive statistic defined on L^∞. Suppose Φ is defined on L^∞_+. Then for any bounded random variable X, we can define

$$\Psi(X) = \min[X] + \Phi(X - \min[X]),$$

where we note that $X - \min[X]$ is a non-negative random variable.

Clearly Ψ is a statistic that depends only on the distribution of X (as Φ does), and $\Psi(c) = c + \Phi(0) = c$ for constants c. When X is non-negative, the additivity of Φ gives $\Phi(X) = \Phi(\min[X]) + \Phi(X - \min[X]) = \min[X] + \Phi(X - \min[X])$, so Ψ is an extension of Φ. Moreover, Ψ is additive because $\min[X+Y] = \min[X] + \min[Y]$, and $\Phi(X + Y - \min[X+Y]) = \Phi(X - \min[X]) + \Phi(Y - \min[Y])$ by the additivity of Φ. Finally, to show Ψ is monotone, suppose X and Y are bounded random variables satisfying $X \geq_1 Y$. Then we can choose a sufficiently large n such that $X + n$ and $Y + n$ are both non-negative, and $X + n \geq_1 Y + n$. Since Φ is monotone for non-negative random variables, $\Phi(X+n) \geq \Phi(Y+n)$. Thus $\Psi(X+n) \geq \Psi(Y+n)$ by the fact that Ψ extends Φ, and $\Psi(X) \geq \Psi(Y)$ by the additivity of Ψ. This proves that Ψ is a monotone additive statistic on L^∞ that extends Φ. □

B.2 Proof of Theorem 3

It is straightforward to check that the representation satisfies the axioms, so we focus on the other direction of deriving the representation from the axioms. In the first step, we fix any reward $x > 0$. Then by monotonicity in time and continuity, for each (x, T) there exists a (unique) deterministic time $\Phi_x(T)$ such that $(x, \Phi_x(T)) \sim (x, T)$. Clearly, when T is a deterministic time, $\Phi_x(T)$ is simply T itself. Note also that if S first-order stochastically dominates T, then

$$(x, \Phi_x(T)) \sim (x, T) \succeq (x, S) \sim (x, \Phi_x(S)),$$

so that $\Phi_x(S) \geq \Phi_x(T)$. We next show that for any T and S that are independent, $\Phi_x(T+S) = \Phi_x(T) + \Phi_x(S)$. Indeed, by stationarity, $(x, \Phi_x(T)) \sim (x, T)$ implies $(x, \Phi_x(T) + S) \sim (x, T+S)$ and $(x, \Phi_x(S)) \sim (x, S)$ implies $(x, \Phi_x(T) + \Phi_x(S)) \sim (x, \Phi_x(T) + S)$. Taken together, we have

$$(x, \Phi_x(T) + \Phi_x(S)) \sim (x, T + S).$$

Since $\Phi_x(T) + \Phi_x(S)$ is a deterministic time, the definition of Φ_x gives $\Phi_x(T) + \Phi_x(S) = \Phi_x(T + S)$ as desired. It follows that each $\Phi_x \colon L^\infty_+ \to \mathbb{R}$ is a monotone additive statistic.

In the second step, note that our preference $\succeq$ induces a preference on $\mathbb{R}_{++} \times \mathbb{R}_+$ consisting of deterministic dated rewards. By Theorem 2 in Fishburn and Rubinstein (1982), for any given $r > 0$ we can find a continuous and strictly increasing utility function $u \colon \mathbb{R}_{++} \to \mathbb{R}_{++}$ such that for deterministic times $t, s \geq 0$

$$(x, t) \succeq (y, s) \quad \text{if and only if} \quad u(x) \cdot \mathrm{e}^{-rt} \geq u(y) \cdot \mathrm{e}^{-rs}.$$

By definition, $(x, T) \sim (x, \Phi_x(T))$ for any random time T. Thus we obtain that the decision maker's preference is represented by

$$(x, T) \succeq (y, S) \quad \text{if and only if} \quad u(x) \cdot \mathrm{e}^{-r\Phi_x(T)} \geq u(y) \cdot \mathrm{e}^{-r\Phi_y(S)}.$$

It remains to show that for all $x, y > 0$, Φ_x and Φ_y are the same statistic. For this we choose deterministic times t and s such that $(x, t) \sim (y, s)$, i.e., $u(x) \cdot \mathrm{e}^{-rt} = u(y) \cdot \mathrm{e}^{-rs}$. For any random time T, stationarity implies $(x, t + T) \sim (y, s + T)$, so that

$$u(x) \cdot \mathrm{e}^{-r\Phi_x(t+T)} = u(y) \cdot \mathrm{e}^{-r\Phi_y(s+T)}.$$

Using the additivity of Φ_x and Φ_y, we can divide the above two equalities and obtain $\Phi_x(T) = \Phi_y(T)$ as desired. Since this holds for all T and all $x, y > 0$, we can write $\Phi_x(T) = \Phi(T)$ for a single monotone additive statistic Φ. This completes the proof.

B.3 Proof of Proposition 2

Define, for every $t \geq 0$, $v_i(t) = \mathrm{e}^{-r_i t}$ and $v(t) = \mathrm{e}^{-rt}$. We have that for any two random times S and T, $(1, S) \succeq_i (1, T)$ if and only if $\mathbb{E}[v_i(S)] \geq \mathbb{E}[v_i(T)]$, and $(1, S) \succeq (1, T)$ if and only if $\mathbb{E}[v(S)] \geq \mathbb{E}[v(T)]$. Thus it follows from the Pareto axiom that for any two random times S and T, $\mathbb{E}[v_i(S)] \geq \mathbb{E}[v_i(T)]$ for all i implies $\mathbb{E}[v(S)] \geq \mathbb{E}[v(T)]$.

By Harsanyi's Theorem (Zhou, 1997, Theorem 2) there exist (λ_i) in $\mathbb{R}_+$ and $\alpha \in \mathbb{R}$ such that for every t, $v(t) = \sum_i \lambda_i v_i(t) + \alpha$. By letting $t \to \infty$ we obtain $0 = \alpha$ and by setting $t = 0$ it follows that $1 = \sum_i \lambda_i$. Further plugging in $t = 1$ and $t = 2$, we obtain

$$\sum_{i=1}^n \lambda_i \mathrm{e}^{-2r_i} = \mathrm{e}^{-2r} = \left(\mathrm{e}^{-r}\right)^2 = \left(\sum_{i=1}^n \lambda_i \mathrm{e}^{-r_i}\right)^2.$$

But the Cauchy-Schwartz inequality gives

$$\sum_{i=1}^n \lambda_i \mathrm{e}^{-2r_i} = \left(\sum_{i=1}^n \lambda_i \mathrm{e}^{-2r_i}\right) \cdot \left(\sum_{i=1}^n \lambda_i\right) \geq \left(\sum_{i=1}^n \lambda_i \mathrm{e}^{-r_i}\right)^2.$$

Thus equality holds, which implies that $r_i = r_j$ for any two agents such that $\lambda_i > 0$ and $\lambda_j > 0$. From $\mathrm{e}^{-r} = \sum_{i=1}^n \lambda_i \mathrm{e}^{-r_i}$ we conclude that $r = r_i$ for any agent i with $\lambda_i > 0$.

B.4 Proof of Theorem 4

We first prove that the proposed representation for the social preference relation $\succeq$ satisfies the Pareto axiom. If $(x, T) \succeq_i (y, S)$ for every i, then $u(x)\mathrm{e}^{-r_i\Phi_i(T)} \geq u(y)\mathrm{e}^{-r_i\Phi_i(S)}$, which can be rewritten as

$$r_i(\Phi_i(S) - \Phi_i(T)) \geq \log \frac{u(y)}{u(x)}.$$

Summing across i using the weights λ_i we obtain

$$\sum_{i=1}^{n} \lambda_i r_i(\Phi_i(S) - \Phi_i(T)) \geq \log \frac{u(y)}{u(x)} \sum_{i=1}^{n} \lambda_i = \log \frac{u(y)}{u(x)}.$$

Since $r\Phi = \sum_{i=1}^{n} \lambda_i r_i \Phi_i$, it follows that $r(\Phi(S) - \Phi(T)) \geq \log \frac{u(y)}{u(x)}$ as well, which is equivalent to $u(x)\mathrm{e}^{-r\Phi(T)} \geq u(y)\mathrm{e}^{-r\Phi(S)}$. Thus $(x, T) \succeq (y, S)$ as desired.

The rest of this proof shows that the Pareto axiom implies $r\Phi = \sum_{i=1}^{n} \lambda_i r_i \Phi_i$ and $r = \sum_i \lambda_i r_i$ for some non-negative weights λ_i that sum to 1. To that end we first show that Φ itself is an average of (Φ_i). Note that if $\Phi_i(T) \leq \Phi_i(S)$ for every i, then $(1, T) \succeq_i (1, S)$ for every i and thus, by the Pareto axiom, $(1, T) \succeq (1, S)$ and $\Phi(T) \leq \Phi(S)$ also hold.

We say that a collection of monotone additive statistics $(\Phi_1, \ldots, \Phi_n, \Phi)$ have the *Pareto property* if $\Phi_i(T) \leq \Phi_i(S)$ for every i implies $\Phi(T) \leq \Phi(S)$. We now show:

Lemma 6. *Let $(\Phi_1, \ldots, \Phi_n, \Phi)$ be monotone additive statistics defined on L^∞_+, and suppose that they satisfy the Pareto property. Then there exists a probability vector $(\beta_1, \ldots, \beta_n)$ such that $\Phi = \sum_{i=1}^{n} \beta_i \Phi_i$.*

Proof. Let $(\mu_1, \ldots, \mu_n, \mu)$ be the mixing measures on $\overline{\mathbb{R}}$ that correspond to the monotone additive statistics $(\Phi_1, \ldots, \Phi_n, \Phi)$. Define the linear functionals $(I_1, \ldots, I_n, I)$ on $\mathcal{C}(\overline{\mathbb{R}})$ as $I_i(f) = \int_{\overline{\mathbb{R}}} f \, \mathrm{d}\mu_i$ and $I(f) = \int_{\overline{\mathbb{R}}} f \, \mathrm{d}\mu$.

We call a set of functions $\mathcal{D} \subseteq \mathcal{C}(\overline{\mathbb{R}})$ a *Pareto domain* if for every $f, g \in \mathcal{D}$,

$$I_i(f) \geq I_i(g) \quad i = 1, \ldots, n \implies I(f) \geq I(g).$$

The Pareto property implies $\mathcal{L}_+ = \{K_X : X \in L^\infty_+\}$ is a Pareto domain. Define, as in the proof of Theorem 1, $\mathcal{L} = \{K_X : X \in L^\infty\}$ as well as the rational cone spanned by $\mathcal{L}$:

$$\mathrm{cone}_{\mathbb{Q}}(\mathcal{L}) = \{qL : q \in \mathbb{Q}_+, L \in \mathcal{L}\} = \bigcup_{n=1}^{\infty} \frac{1}{n}\mathcal{L}$$

We show that $\mathcal{L}$ and $\mathrm{cone}_{\mathbb{Q}}(\mathcal{L})$ are both Pareto domains. Given $X, Y \in L^\infty$, let c be a large positive constant such that $X + c \geq 0$ and $Y + c \geq 0$. If $I_i(K_X) \geq I_i(K_Y)$ for all i then $I_i(K_X + c) \geq I_i(K_Y + c)$ for all i since each I_i is linear. Thus, by the Pareto property and the linearity of I, $I(K_X + c) \geq I(K_Y + c)$ and $I(K_X) \geq I(K_Y)$. This shows $\mathcal{L}$ is a

Pareto domain. As for $\mathrm{cone}_{\mathbb{Q}}(\mathcal{L})$, observe that $I_i(\frac{1}{m}K_X) \geq I_i(\frac{1}{n}K_Y)$ for all i is equivalent to $I_i(nK_X) \geq I_i(mK_Y)$ for all i, which implies $I(nK_X) \geq I(mK_Y)$ since $\mathcal{L}$ is a Pareto domain and is closed under addition. This shows $I(\frac{1}{m}K_X) \geq I(\frac{1}{n}K_Y)$ as desired.

Next we show that the closure of $\mathrm{cone}_{\mathbb{Q}}(\mathcal{L})$ (with respect to the usual sup norm) is also a Pareto domain. Let f, g be in the closure, such that $I_i(f) \geq I_i(g)$ for all i. Pick sequences (f_k) and (g_k) in $\mathrm{cone}_{\mathbb{Q}}(\mathcal{L})$ converging to f and g. Define $\varepsilon_{i,k} = |I_i(f) - I_i(f_k)| + |I_i(g) - I_i(g_k)|$ and $\varepsilon_k = \max_{1 \leq i \leq n} \varepsilon_{i,k}$. Then from $I_i(f) \geq I_i(g)$ we deduce $I_i(f_k) \geq I_i(g_k) - \varepsilon_k = I_i(g_k - \varepsilon_k)$ for every i. Note that $g_k - \varepsilon_k$ belongs to $\mathrm{cone}_{\mathbb{Q}}(\mathcal{L})$ since the latter contains all the constant functions and is closed under addition. Thus by the fact that $\mathrm{cone}_{\mathbb{Q}}(\mathcal{L})$ is a Pareto domain, $I_i(f_n) \geq I_i(g_n - \varepsilon_n)$ for every i implies $I(f_k) \geq I(g_k - \varepsilon_k) = I(g_k) - \varepsilon_k$ for every k. Continuity of the functionals (I_i) yields $\varepsilon_k \to 0$. Continuity of I thus yields $I(f) \geq I(g)$.

This proves that the closure of $\mathrm{cone}_{\mathbb{Q}}(\mathcal{L})$ is a Pareto domain. Since the subset of a Pareto domain is a Pareto domain, we conclude that $\mathrm{cone}(\mathcal{L})$ (i.e. the cone generated by $\mathcal{L}$) is a Pareto domain as well.

Now define $\mathcal{V} = \mathrm{cone}(\mathcal{L}) - \mathrm{cone}(\mathcal{L})$ to be the vector space generated by the cone. It is immediate to verify, using the linearity of the integral, that $\mathcal{V}$ is a Pareto domain as well. In particular, for any $f \in \mathcal{V}$, $I_i(f) \leq 0$ for every i implies $I(f) \leq 0$. Corollary 5.95 in Aliprantis and Border (2006) thus implies there exist non-negative scalars $\beta_1, \ldots, \beta_n$ such that $I = \sum_{i=1}^n \beta_i I_i$ on $\mathcal{V}$. So $I(K_X) = \sum_{i=1}^n \beta_i I_i(K_X)$ for every $X \in L^\infty$, which implies $\Phi(X) = \sum_{i=1}^n \beta_i \Phi_i(X)$. For constant X this implies $\sum_i \beta_i = 1$, proving the lemma. $\square$

To show that $r\Phi$ is also an average of $(r_i\Phi_i)$, we make use of the Pareto axiom when applied to time lotteries with different rewards. For any $S, T \in L^\infty_+$, the Pareto axiom says that if rewards x, y are such that $r_i\Phi_i(S) - r_i\Phi_i(T) \geq \log(u(y)/u(x))$ for all i, then $r\Phi(S) - r\Phi(T) \geq \log(u(y)/u(x))$ also holds. By the richness assumption on u, we can choose x, y with

$$\log(u(y)/u(x)) = \min_{1 \leq i \leq n} \{r_i\Phi_i(S) - r_i\Phi_i(T)\}.$$

Therefore the Pareto axiom implies that for any $S, T \in L^\infty_+$,

$$r\Phi(S) - r\Phi(T) \geq \min_{1 \leq i \leq n} \{r_i\Phi_i(S) - r_i\Phi_i(T)\}. \tag{12}$$

The conclusion that $r\Phi$ is an average of $(r_i\Phi_i)$ will follow from the condition (12) via an application of Farkas' Lemma. To rewrite this condition in linear algebra form, we let $m \leq n$ be the largest number of different Φ_i that are linearly independent (when viewed as functions on L^∞_+). Reordering if necessary, we can assume $\Phi_1, \ldots, \Phi_m$ are linearly independent, and every Φ_i is a (not necessarily positive) linear combination of those m. Thus we can find vectors $\gamma^1, \ldots, \gamma^n \in \mathbb{R}^m$ such that every $r_i\Phi_i$ can be rewritten as the following inner product (i.e., linear combination):

$$r_i\Phi_i = \gamma^i \cdot (\Phi_1, \ldots, \Phi_m).$$

Since Φ is an average of (Φ_i), there also exists $\gamma \in \mathbb{R}^m$ such that $r\Phi = \gamma \cdot (\Phi_1, \ldots, \Phi_m)$.

Consider the following set of vectors:

$$\mathcal{W} = \{w \in \mathbb{R}^m : \ \gamma \cdot w \geq \min_{1 \leq i \leq n} \gamma^i \cdot w\}.$$

Let $\mathcal{D}$ be all vectors of the form $(\Phi_1(S) - \Phi_1(T), \ldots, \Phi_m(S) - \Phi_m(T))$ for some $S, T \in L^\infty_+$. Condition (12) says that $\mathcal{D} \subseteq \mathcal{W}$. Note that $-\mathcal{D} = \mathcal{D}$, and $\mathcal{D}$ is closed under addition because every Φ_i is additive. Moreover, since the definition of $\mathcal{W}$ involve homogeneous inequalities, $\frac{1}{N}\mathcal{D} \subseteq \mathcal{W}$ for every positive integer N. From these properties we deduce that any vector of the form $q_1 w_1 + \cdots + q_k w_k$ with $q_j \in \mathbb{Q}$ and $w_j \in \mathcal{D}$ belongs to $\mathcal{W}$, because it can be written as $\frac{1}{N}w$ for some positive integer N and $w \in \mathcal{D}$. Since $\mathcal{W}$ is a closed set, the span of $\mathcal{D}$ (not just the rational span) is also contained in $\mathcal{W}$. Finally note that $\mathcal{D}$ spans the entirety of $\mathbb{R}^m$. This is because by setting $T = 0$, $\mathcal{D}$ in particular includes vectors of the form $(\Phi_1(S), \ldots, \Phi_m(S))$, and such vectors cannot all belong to a lower-dimensional subspace by the assumption that $\Phi_1, \ldots, \Phi_m$ are linearly independent.

Therefore, $\mathcal{D} = \mathcal{W} = \mathbb{R}^m$, which implies

$$\gamma \cdot w \geq \min_{1 \leq i \leq n} \gamma^i \cdot w \ \text{ for all } w \in \mathbb{R}^m. \tag{13}$$

For any $\varepsilon > 0$, this condition implies that there exists no $w \in \mathbb{R}^m$ such that $-\gamma^i \cdot w \leq -1 - \varepsilon$ for every i while $\gamma \cdot w \leq 1$. Let A be an $(n+1) \times m$ matrix whose first n rows are $-\gamma^1, \ldots, -\gamma^n$, and whose last row is γ. Let b be the $n+1$-dimensional vector $(-1-\varepsilon, \ldots, -1-\varepsilon, 1)$. Then $Aw \leq b$ has no solution $w \in \mathbb{R}^m$.

By Farkas' Lemma, there exists a non-negative $n+1$-dimensional vector $z = (z_1, \ldots, z_{n+1})$ such that $z'A = 0$ while $z \cdot b < 0$. The former implies $z_{n+1}\gamma = z_1\gamma^1 + \cdots + z_n\gamma^n$, while the latter implies $z_{n+1} < (1+\varepsilon)(z_1 + \cdots + z_n)$. Note that z_{n+1} cannot be zero, for otherwise we have a positive linear combination of $\gamma^1, \ldots, \gamma^n$ that gives the zero vector, leading to the impossible implication that a positive linear combination of $\Phi_1, \ldots, \Phi_n$ equals zero.

Thus we can write $\gamma = \alpha_1\gamma^1 + \cdots + \alpha_n\gamma^n$, with non-negative weights $\alpha_i = \frac{z_i}{z_{n+1}}$ whose sum is greater than $\frac{1}{1+\varepsilon}$. Consequently $r\Phi = \sum_{i=1}^n \alpha_i r_i \Phi_i$, which implies $r = \sum_{i=1}^n \alpha_i r_i$ and thus $\alpha_i \leq \frac{r}{r_i}$ in any such representation. Since ε is arbitrary, a compactness argument then yields that $\gamma = \sum_{i=1}^n \alpha_i \gamma^i$ for some non-negative weights α_i with $\sum_{i=1}^n \alpha_i \geq 1$.

We can also choose $\hat{b} = (1-\varepsilon, \ldots, 1-\varepsilon, -1)$ and deduce from (13) that $Aw \leq \hat{b}$ has no solution $w \in \mathbb{R}^m$. Then a similar analysis yields $\gamma = \hat{\alpha}_1\gamma^1 + \cdots + \hat{\alpha}_n\gamma^n$ for some weights $\hat{\alpha}_i \geq 0$ and $\sum_{i=1}^n \alpha_i < \frac{1}{1-\epsilon}$. Again by compactness, we can assume $\sum_{i=1}^n \hat{\alpha}_i \leq 1$. Finally, by suitably averaging between α_i and $\hat{\alpha}_i$, we can find non-negative weights (λ_i) whose sum is equal to 1, such that $\gamma = \sum_{i=1}^n \lambda_i \gamma^i$. So $r\Phi = \sum_{i=1}^n \lambda_i r_i \Phi_i$. Since Φ is also a convex combination of (Φ_i), it follows that $r = \sum_i \lambda_i r_i$, completing the proof.

C Preferences over Gambles

C.1 Proof of Proposition 3

The result can be derived as a corollary of Theorem 5, but we also provide a direct proof here. We focus on the "only if" direction because the "if" direction follows immediately from the monotonicity of $K_a(X)$ in a. Suppose μ is not supported on $[-\infty, 0]$, we will show that the resulting monotone additive statistic Φ does not always exhibit risk aversion. Since μ has positive mass on $(0, \infty]$, we can find $\varepsilon > 0$ such that μ assigns mass at least ε to $(\varepsilon, \infty]$. Now consider a gamble X which is equal to 0 with probability $\frac{n-1}{n}$ and equal to n with probability $\frac{1}{n}$, for some large positive integer n. Then $\mathbb{E}[X] = 1$ and $K_a(X) \geq \min[X] = 0$ for every $a \in \overline{\mathbb{R}}$. Moreover, for $a \geq \varepsilon$ we have

$$K_a(X) \geq K_\varepsilon(X) = \frac{1}{\varepsilon} \log \left(\frac{n-1}{n} + \frac{1}{n} \mathrm{e}^{\varepsilon n} \right) \geq \frac{n}{2}$$

whenever n is sufficient large. Thus

$$\Phi(X) = \int_{\overline{\mathbb{R}}} K_a(X) \, \mathrm{d}\mu(a) \geq \int_{[\varepsilon, \infty]} K_a(X) \, \mathrm{d}\mu(a) \geq \frac{n}{2} \varepsilon.$$

We thus have $\Phi(X) > 1 = \mathbb{E}[X]$ for all large n, showing that the preference represented by Φ sometimes exhibits risk seeking.

Symmetrically, if μ is not supported on $[0, \infty]$, then Φ must sometimes exhibit risk aversion (by considering X equal to 0 with probability $\frac{1}{n}$ and equal to n with probability $\frac{n-1}{n}$). This completes the proof.

C.2 Proof of Proposition 4

Fix $X \in L^\infty$ with $\mathbb{E}[X] = 0$ and $\min[X] < 0$. For $a \in (-\infty, 0)$ define $f_{a,X}(\varepsilon) = K_a(\varepsilon X) = \frac{1}{a} \log \mathbb{E}\left[\mathrm{e}^{a\varepsilon X}\right]$ for $\varepsilon \geq 0$. Then we have

$$f'_{a,X}(\varepsilon) = \frac{\partial}{\partial \varepsilon} K_a(\varepsilon X) = \frac{\mathbb{E}\left[X \mathrm{e}^{a\varepsilon X}\right]}{\mathbb{E}\left[\mathrm{e}^{a\varepsilon X}\right]}.$$

Thus $f'_{a,X}(0) = \mathbb{E}[X] = 0$, implying that

$$\lim_{\varepsilon \to 0} \frac{K_a(\varepsilon X)}{\varepsilon} = 0 \text{ whenever } a \in (-\infty, 0).$$

Note that $K_0(\varepsilon X) = \mathbb{E}[\varepsilon X] = 0$ for every ε, while $K_{-\infty}(\varepsilon X) = \min[\varepsilon X] = \varepsilon \min[X]$. So the above limit is also zero for $a = 0$ but equal to $\min[X] < 0$ for $a = -\infty$.

Since $\frac{K_a(\varepsilon X)}{\varepsilon}$ is uniformly bounded between $\min[X]$ and 0, we can apply the Dominated Convergence Theorem to deduce

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \Phi(\varepsilon X) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int K_a(\varepsilon X) \, \mathrm{d}\mu(a) = \int \left(\lim_{\varepsilon \to 0} \frac{K_a(\varepsilon X)}{\varepsilon} \right) \mathrm{d}\mu(a) = \min[X] \cdot \mu(\{-\infty\}).$$

This shows that Φ exhibits first-order risk aversion if and only if $\mu(\{-\infty\}) > 0$.

Below we assume $\mu(-\infty) = 0$ and use a similar method to study second-order risk aversion, focusing on the limit $\lim_{\varepsilon\to 0} \frac{K_a(\varepsilon X)}{\varepsilon^2}$ for $a \in (-\infty, 0)$. From the above formula for $f'_{a,X}(\varepsilon)$, we can further compute the second derivative as

$$f''_{a,X}(\varepsilon) = \frac{a\mathbb{E}\left[X^2 \mathrm{e}^{a\varepsilon X}\right] \cdot \mathbb{E}\left[\mathrm{e}^{a\varepsilon X}\right] - a\mathbb{E}\left[X\mathrm{e}^{a\varepsilon X}\right]^2}{\mathbb{E}\left[\mathrm{e}^{a\varepsilon X}\right]^2}.$$

In particular, $f''_{a,X}(0) = a\,\mathrm{Var}(X)$, and $\|f''_{a,X}(\varepsilon)| \leq |a| \max\{\max[X]^2, \min[X]^2\}$ for every ε.

By Taylor's Theorem, we can write $K_a(\varepsilon X) = f_{a,X}(\varepsilon)$ as

$$f_{a,X}(\varepsilon) = f_{a,X}(0) + f'_{a,X}(0) \cdot \varepsilon + \frac{1}{2} f''_{a,X}(\eta) \cdot \varepsilon^2 \text{ for some } \eta \in (0, \varepsilon).$$

Plugging in $f_{a,X}(0) = f'_{a,X}(0) = 0$, this yields $\frac{K_a(\varepsilon X)}{\varepsilon^2} = \frac{1}{2} f''_{a,X}(\eta)$ which converges to $\frac{1}{2} f''_{a,X}(0) = \frac{1}{2} a\,\mathrm{Var}(X)$ as $\varepsilon \to 0$. This ratio is also bounded in absolute value by the dominating function $\frac{1}{2}|a| \max\{\max[X]^2, \min[X]^2\}$.

When $\int |a| \,\mathrm{d}\mu(a)$ is finite, the dominating function is integrable. Thus by the Dominated Convergence Theorem,

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon^2} \Phi(\varepsilon X) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon^2} \int K_a(\varepsilon X) \,\mathrm{d}\mu(a) = \int \left(\lim_{\varepsilon\to 0} \frac{K_a(\varepsilon X)}{\varepsilon^2} \right) \mathrm{d}\mu(a) = \mathrm{Var}(X) \cdot \int \frac{a}{2} \,\mathrm{d}\mu(a).$$

This is a finite negative number, showing that Φ exhibits second-order risk aversion when $\int |a| \,\mathrm{d}\mu(a)$ is finite.

Finally suppose $\int |a| \,\mathrm{d}\mu(a) = \infty$, we will show $\lim_{\varepsilon\to 0} \frac{1}{\varepsilon^2} \Phi(\varepsilon X) = -\infty$ so that Φ is not second-order risk averse. Indeed, since $-\frac{K_a(\varepsilon X)}{\varepsilon^2}$ is non-negative for every $\varepsilon \geq 0$ and $a \leq 0$, we can apply Fatou's Lemma to deduce

$$\liminf_{\varepsilon\to 0} \frac{1}{\varepsilon^2} \int -K_a(\varepsilon X) \,\mathrm{d}\mu(a) \geq \int \left(\liminf_{\varepsilon\to 0} \frac{-K_a(\varepsilon X)}{\varepsilon^2} \right) \mathrm{d}\mu(a) = \mathrm{Var}(X) \cdot \int \frac{-a}{2} \,\mathrm{d}\mu(a) = \infty.$$

Thus $\lim_{\varepsilon\to 0} \frac{-\Phi(\varepsilon X)}{\varepsilon^2} = \infty$ as we desire to show.

C.3 Proof of Theorem 5

We first show that conditions (i) and (ii) are necessary for $\int_{\mathbb{R}} K_a(X) \,\mathrm{d}\mu_1(a) \leq \int_{\mathbb{R}} K_a(Y) \,\mathrm{d}\mu_2(a)$ to hold for every X. This part of the argument closely follows the proof of Lemma 5. Specifically, by considering the same random variables $X_{n,b}$ as defined there, we have the key equation (10). Since the limit on the left-hand side is smaller for μ_1 than for μ_2, we conclude that for every $b > 0$, $\int_{[b,\infty]} \frac{a-b}{a} \,\mathrm{d}\mu_1(a)$ on the right-hand side must be smaller than the corresponding integral for μ_2. Thus condition (i) holds, and an analogous argument shows condition (ii) also holds.

To complete the proof, it remains to show that when conditions (i) and (ii) are satisfied,

$$\int_{\overline{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu_1(a) \le \int_{\overline{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu_2(a)$$

holds for every X. Since μ_1 and μ_2 are both probability measures, we can subtract $\mathbb{E}[X]$ from both sides and arrive at the equivalent inequality

$$\int_{\overline{\mathbb{R}}_{\neq 0}} (K_a(X) - \mathbb{E}[X])\,\mathrm{d}\mu_1(a) \le \int_{\overline{\mathbb{R}}_{\neq 0}} (K_a(X) - \mathbb{E}[X])\,\mathrm{d}\mu_2(a). \tag{14}$$

Note that we can exclude $a = 0$ from the range of integration because $K_a(X) = \mathbb{E}[X]$ there. Below we show that condition (i) implies

$$\int_{(0,\infty]} (K_a(X) - \mathbb{E}[X])\,\mathrm{d}\mu_1(a) \le \int_{(0,\infty]} (K_a(X) - \mathbb{E}[X])\,\mathrm{d}\mu_2(a). \tag{15}$$

Similarly, condition (ii) gives the same inequality when the range of integration is $[-\infty, 0)$. Adding these two inequalities would yield the desired comparison in (14).

To prove (15), we let $L_X(a) = a \cdot K_a(X) = \log \mathbb{E}\left[\mathrm{e}^{aX}\right]$ be the cumulant generating function of X. It is well known that $L_X(a)$ is convex in a, with $L'_X(0) = \mathbb{E}[X]$ and $\lim_{a\to\infty} L'_X(a) = \max[X]$. Then the integral on the left-hand side of (15) can be calculated as follows:

$$\begin{aligned}\int_{(0,\infty]} (K_a(X) - \mathbb{E}[X])\,\mathrm{d}\mu_1(a) &= \int_{(0,\infty)} (K_a(X) - \mathbb{E}[X])\,\mathrm{d}\mu_1(a) + (\max[X] - \mathbb{E}[X]) \cdot \mu_1(\{\infty\}) \\ &= \int_{(0,\infty)} (L_X(a) - a\mathbb{E}[X])\,\mathrm{d}\frac{\mu_1(a)}{a} + (\max[X] - \mathbb{E}[X]) \cdot \mu_1(\{\infty\})\end{aligned}$$

Note that since the function $g(a) = L_X(a) - a\mathbb{E}[X]$ satisfies $g(0) = g'(0) = 0$, it can be written as

$$g(a) = \int_0^a g'(t)\,\mathrm{d}t = \int_0^a \int_0^t g''(b)\,\mathrm{d}b\,\mathrm{d}t = \int_0^a g''(b) \cdot (a - b)\,\mathrm{d}b.$$

Plugging back to the previous identity, we obtain

$$\begin{aligned}&\int_{(0,\infty]} (K_a(X) - \mathbb{E}[X])\,\mathrm{d}\mu_1(a) \\ &= \int_{(0,\infty)} \int_0^a L''_X(b) \cdot (a - b)\,\mathrm{d}b\,\mathrm{d}\frac{\mu_1(a)}{a} + (\max[X] - \mathbb{E}[X]) \cdot \mu_1(\{\infty\}) \\ &= \int_0^\infty L''_X(b) \int_{[b,\infty)} (a - b)\,\mathrm{d}\frac{\mu_1(a)}{a}\,\mathrm{d}b + (L'_X(\infty) - L'_X(0)) \cdot \mu_1(\{\infty\}) \\ &= \int_0^\infty L''_X(b) \int_{[b,\infty)} \frac{a-b}{a}\,\mathrm{d}\mu_1(a)\,\mathrm{d}b + \int_0^\infty L''_X(b) \cdot \mu_1(\{\infty\})\,\mathrm{d}b \\ &= \int_0^\infty L''_X(b) \int_{[b,\infty]} \frac{a-b}{a}\,\mathrm{d}\mu_1(a)\,\mathrm{d}b,\end{aligned}$$

where the last step uses $\frac{a-b}{a} = 1$ when $a = \infty > b$.

The above identity also holds when μ_1 is replaced by μ_2. We then see that (15) follows from condition (i) and $L''_X(b) \ge 0$ for all b. This completes the proof.

C.4 Proof of Theorem 6

The "if" direction is straightforward: if $\succeq_1$ and $\succeq_2$ are both represented by a monotone additive statistic Φ, then they satisfy responsiveness and continuity. In addition, combined choices are not stochastically dominated because if $X \succ_1 X'$ and $Y \succ_2 Y'$ then $\Phi(X) > \Phi(X')$ and $\Phi(Y) > \Phi(Y')$. Thus $\Phi(X+Y) > \Phi(X'+Y')$ and $X'+Y'$ cannot stochastically dominate $X+Y$.

Turning to the "only if" direction, we suppose $\succeq_1$ and $\succeq_2$ satisfy the axioms. We first show that these preferences are the same. Suppose for the sake of contradiction that $X \succeq_1 Y$ but $Y \succ_2 X$ for some X, Y. Then by continuity, there exists $\varepsilon > 0$ such that $Y \succ_2 X + \varepsilon$. By responsiveness, we also have $X \succeq_1 Y \succ Y - \frac{\varepsilon}{2}$. Thus $X \succ_1 Y - \frac{\varepsilon}{2}$, $Y \succ_2 X + \varepsilon$, but $X + Y$ is strictly stochastically dominated by $Y - \frac{\varepsilon}{2} + X + \varepsilon = X + Y + \frac{\varepsilon}{2}$, contradicting Axiom 4.2.

Henceforth we denote both $\succeq_1$ and $\succeq_2$ by $\succeq$. We next show that for any X and any $\varepsilon > 0$, $\max[X] + \varepsilon \succ X \succ \min[X] - \varepsilon$. To see why, suppose for contradiction that X is weakly preferred to $\max[X] + \varepsilon$ (the other case can be handled similarly). Then we obtain a contradiction to Axiom 4.2 by observing that $X \succ \max[X] + \frac{\varepsilon}{2}$, $\frac{\varepsilon}{4} \succ 0$ but $X + \frac{\varepsilon}{4} <_1 \max[X] + \frac{\varepsilon}{2} + 0$.

Given these upper and lower bounds for X, we can define $\Phi(X) = \sup\{c \in \mathbb{R} : \ c \preceq X\}$, which is well-defined and finite. By definition of the supremum and responsiveness, for any $\varepsilon > 0$ it holds that $\Phi(X) - \varepsilon \prec X \prec \Phi(X) + \varepsilon$. Thus by continuity, $\Phi(X) \sim X$ is the (unique) certainty equivalent of X.

It remains to show that Φ is a monotone additive statistic. For this we show that $X \sim Y$ implies $X + Z \sim Y + Z$ for any independent Z. Suppose for contradiction that $X + Z \succ Y + Z$. Then by continuity we can find $\varepsilon > 0$ such that $X + Z \succ Y + Z + \varepsilon$. By responsiveness, it also holds that $Y + \frac{\varepsilon}{2} \succ Y \sim X$. But the sum $(X + Z) + (Y + \frac{\varepsilon}{2})$ is stochastically dominated by $(Y + Z + \varepsilon) + X$, contradicting Axiom 4.2.

Therefore, from $X \sim \Phi(X)$ and $Y \sim \Phi(Y)$ we can apply the preceding result twice to obtain $X + Y \sim \Phi(X) + Y \sim \Phi(X) + \Phi(Y)$ whenever X, Y are independent, so that $\Phi(X + Y) = \Phi(X) + \Phi(Y)$ is additive. Finally, we show Φ is monotone. Consider any $Y \geq_1 X$, and suppose for contradiction that $X \succ Y$. Then there exists $\varepsilon > 0$ such that $X \succ Y + \varepsilon$. This leads to a contradiction since $X \succ Y + \varepsilon$, $\frac{\varepsilon}{2} \succ 0$, but $X + \frac{\varepsilon}{2}$ is stochastically dominated by $Y + \varepsilon + 0$.

This completes the proof that both preferences $\succeq_1$ and $\succeq_2$ are represented by the same certainty equivalent $\Phi(X)$, which is a monotone additive statistic.

Online Appendix

D Proof of Theorem 2

The proof is considerably more complex than the proof of Theorem 1, so we break it into several steps below.

D.1 Step 1: Catalytic Order on L_M

We first establish a generalization of Theorem 7 to unbounded random variables. For two random variables X and Y with c.d.f. F and G respectively, we say that X *dominates* Y *in both tails* if there exists a positive number N with the property that

$$G(x) > F(x) \quad \text{for all } |x| \geq N.$$

In particular, X needs to be unbounded from above, and Y unbounded from below.

Lemma 7. *Suppose $X, Y \in L_M$ satisfy $K_a(X) > K_a(Y)$ for every $a \in \mathbb{R}$. Suppose further that X dominates Y in both tails. Then there exists an independent random variable $Z \in L_M$ such that $X + Z \geq_1 Y + Z$.*

Proof. We will take Z to have a normal distribution, which does belong to L_M. Following the proof of Theorem 7, we let $\sigma(x) = G(x) - F(x)$, and seek to show that $[\sigma * h](y) \geq 0$ for every y when h is a Gaussian density with sufficiently large variance. By assumption, $\sigma(x)$ is strictly positive for $|x| \geq N$. Thus there exists $\delta > 0$ such that $\int_{N+1}^{N+2} \sigma(x)\,\mathrm{d}x > \delta$, as well as $\int_{-N-2}^{-N-1} \sigma(x)\,\mathrm{d}x > \delta$. We fix $A > 0$ that satisfies $\mathrm{e}^A \geq \frac{4N}{\delta}$.

Similar to (8), we have for $h(x) = \mathrm{e}^{-\frac{x^2}{2V}}$ that

$$\mathrm{e}^{\frac{y^2}{2V}} \int \sigma(x) h(y-x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V}\cdot x} \cdot \mathrm{e}^{-\frac{x^2}{2V}}\,\mathrm{d}x. \tag{16}$$

The variance V is to be determined below.

We first show that the right-hand side is positive if $V \geq (N+2)^2$ and $\frac{y}{V} \geq A$. Indeed, since $\sigma(x) > 0$ for $|x| \geq N$, this integral is bounded from below by

$$\begin{aligned}
& \int_{-N}^{N} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V}\cdot x} \cdot \mathrm{e}^{-\frac{x^2}{2V}}\,\mathrm{d}x + \int_{N+1}^{N+2} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V}\cdot x} \cdot \mathrm{e}^{-\frac{x^2}{2V}}\,\mathrm{d}x \\
\geq \;& -2N \cdot \mathrm{e}^{\frac{y}{V}\cdot N} + \delta \cdot \mathrm{e}^{\frac{y}{V}\cdot(N+1)} \cdot \mathrm{e}^{-\frac{(N+2)^2}{2V}} \\
= \;& \mathrm{e}^{\frac{y}{V}\cdot N} \cdot \left(-2N + \delta \cdot \mathrm{e}^{\frac{y}{V}} \cdot \mathrm{e}^{-\frac{(N+2)^2}{2V}}\right) \\
> \;& 0,
\end{aligned}$$

where the last inequality uses $\mathrm{e}^{\frac{y}{V}} \geq \mathrm{e}^{A} \geq \frac{4N}{\delta}$ and $\mathrm{e}^{-\frac{(N+2)^2}{2V}} \geq \mathrm{e}^{-\frac{1}{2}} > \frac{1}{2}$. By a symmetric argument, we can show that the right-hand side of (16) is also positive when $\frac{y}{V} \leq -A$.

It remains to consider the case where $\frac{y}{V} \in [-A, A]$. Here we rewrite the integral on the right-hand side of (16) as

$$\int_{-\infty}^{\infty} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V} \cdot x} \cdot \mathrm{e}^{-\frac{x^2}{2V}} \,\mathrm{d}x = M_\sigma(\frac{y}{V}) - \int_{-\infty}^{\infty} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V} \cdot x} \cdot (1 - \mathrm{e}^{-\frac{x^2}{2V}}) \,\mathrm{d}x,$$

where $M_\sigma(a) = \int_{-\infty}^{\infty} \sigma(x) \cdot \mathrm{e}^{ax} \,\mathrm{d}x = \frac{1}{a}\mathbb{E}\left[\mathrm{e}^{aX}\right] - \frac{1}{a}\mathbb{E}\left[\mathrm{e}^{aY}\right]$ is by assumption strictly positive for all a. By continuity, there exists some $\varepsilon > 0$ such that $M_{\sigma(a)} > \varepsilon$ for all $|a| \leq A$. So it only remains to show that when V is sufficiently large,

$$\int_{-\infty}^{\infty} \sigma(x) \cdot \mathrm{e}^{ax} \cdot (1 - \mathrm{e}^{-\frac{x^2}{2V}}) \,\mathrm{d}x < \varepsilon \quad \text{for all } |a| \leq A. \tag{17}$$

To estimate this integral, note that $M_\sigma(A) = \int_{-\infty}^{\infty} \sigma(x) \cdot \mathrm{e}^{Ax} \,\mathrm{d}x$ is finite. Since $\sigma(x) > 0$ for $|x|$ sufficiently large, we deduce from the Monotone Convergence Theorem that $\int_{-\infty}^{T} \sigma(x) \cdot \mathrm{e}^{Ax} \,\mathrm{d}x$ converges to $M_\sigma(A)$ as $T \to \infty$. In other words, $\int_{T}^{\infty} \sigma(x) \cdot \mathrm{e}^{Ax} \,\mathrm{d}x \to 0$. We can thus find a sufficiently large $T > N$ such that $\int_{T}^{\infty} \sigma(x) \cdot \mathrm{e}^{Ax} \,\mathrm{d}x < \frac{\varepsilon}{4}$, and likewise $\int_{-\infty}^{-T} \sigma(x) \cdot \mathrm{e}^{-Ax} \,\mathrm{d}x < \frac{\varepsilon}{4}$.

As $1 - \mathrm{e}^{-\frac{x^2}{2V}} \geq 0$ and $\mathrm{e}^{ax} \leq \mathrm{e}^{A|x|}$ when $|a| \leq A$, we deduce that

$$\int_{|x| \geq T} \sigma(x) \cdot \mathrm{e}^{ax} \cdot (1 - \mathrm{e}^{-\frac{x^2}{2V}}) \,\mathrm{d}x < \frac{\varepsilon}{2} \quad \text{for all } |a| \leq A.$$

Moreover, for this fixed T, we have $\mathrm{e}^{-\frac{T^2}{2V}} \to 1$ when V is large, and thus

$$\int_{|x| \leq T} \sigma(x) \cdot \mathrm{e}^{ax} \cdot (1 - \mathrm{e}^{-\frac{x^2}{2V}}) \,\mathrm{d}x < 2T\mathrm{e}^{AT}(1 - \mathrm{e}^{-\frac{T^2}{2V}}) < \frac{\varepsilon}{2} \quad \text{for all } |a| \leq A.$$

These estimates together imply that (17) holds for sufficiently large V. This completes the proof. □

D.2 Step 2: A Perturbation Argument

With Lemma 7, we know that if Φ is a monotone additive statistic defined on L_M, then $K_a(X) \geq K_a(Y)$ for all $a \in \mathbb{R}$ implies $\Phi(X) \geq \Phi(Y)$ *under the additional assumption that* X dominates Y in both tails (same proof as for Lemma 1). Below we deduce the same result without this extra assumption. To make the argument simpler, assume X and Y are unbounded both from above and from below; otherwise, we can add to them an independent Gaussian random variable without changing either the assumption or the conclusion. In doing so, we can further assume X and Y admit probability density functions.

We first construct a heavy right-tailed random variable as follows:

Lemma 8. *For any $Y \in L_M$ that is unbounded from above and admits densities, there exists $Z \in L_M$ such that $Z \geq 0$ and $\frac{\mathbb{P}[Z>x]}{\mathbb{P}[Y>x]} \to \infty$ as $x \to \infty$.*

Proof. For this result, it is without loss to assume $Y \geq 0$ because we can replace Y by $|Y|$ and only strengthen the conclusion. Let $g(x)$ be the probability density function of Y. We consider a random variable Z whose p.d.f. is given by $cxg(x)$ for all $x \geq 0$, where $c > 0$ is a normalizing constant to ensure $\int_{x \geq 0} cxg(x)\, \mathrm{d}x = 1$. Since the likelihood ratio between $Z = x$ and $Y = x$ is cx, it is easy to see that the ratio of tail probabilities also diverges. Thus it only remains to check $Z \in L_M$. This is because

$$\mathbb{E}\left[\mathrm{e}^{aZ}\right] = c \int_{x \geq 0} xg(x)\mathrm{e}^{ax}\, \mathrm{d}x,$$

which is simply c times the derivative of $\mathbb{E}\left[\mathrm{e}^{aY}\right]$ with respect to a. It is well-known that the moment generating function is smooth whenever it is finite. So this derivative is finite, and $Z \in L_M$. $\square$

In the same way, we can construct heavy left-tailed distributions:

Lemma 9. *For any $X \in L_M$ that is unbounded from below and admits densities, there exists $W \in L_M$, such that $W \leq 0$ and $\frac{\mathbb{P}[W \leq x]}{\mathbb{P}[X \leq x]} \to \infty$ as $x \to -\infty$.*

With these technical lemmata, we now construct "perturbed" versions of any two random variables X and Y to achieve dominance in both tails. For any random variable $Z \in L_M$ and every $\varepsilon > 0$, let Z_ε be the random variable that equals Z with probability ε, and 0 with probability $1 - \varepsilon$. Note that Z_ε also belongs to L_M.

Lemma 10. *Given any two random variables $X, Y \in L_M$ that are unbounded on both sides and admit densities. Let $Z \geq 0$ and $W \leq 0$ be constructed from the above two lemmata. Then for every $\varepsilon > 0$, $X + Z_\varepsilon$ dominates $Y + W_\varepsilon$ in both tails.*

Proof. For the right tail, we need $\mathbb{P}[X + Z_\varepsilon > x] > \mathbb{P}[Y + W_\varepsilon > x]$ for all $x \geq N$. Note that $W_\varepsilon \leq 0$, so $\mathbb{P}[Y + W_\varepsilon > x] \leq \mathbb{P}[Y > x]$. On other hand,

$$\mathbb{P}[X + Z_\varepsilon > x] \geq \mathbb{P}[X \geq 0] \cdot \mathbb{P}[Z_\varepsilon > x] = \mathbb{P}[X \geq 0] \cdot \varepsilon \cdot \mathbb{P}[Z > x].$$

Since by assumption X is unbounded from above, the term $\mathbb{P}[X \geq 0] \cdot \varepsilon$ is a strictly positive constant that does not depend on x. Thus for sufficiently large x, we have

$$\mathbb{P}[X \geq 0] \cdot \varepsilon \cdot \mathbb{P}[Z > x] > \mathbb{P}[Y > x]$$

by the construction of Z. This gives dominance in the right tail. The left tail is similar. $\square$

D.3 Step 3: Monotonicity w.r.t. K_a

The next result generalizes the key Lemma 1 to our current setting:

Lemma 11. *Let $\Phi\colon L_M \to \mathbb{R}$ be a monotone additive statistic. If $K_a(X) \geq K_a(Y)$ for all $a \in \mathbb{R}$ then $\Phi(X) \geq \Phi(Y)$.*

Proof. As discussed, we can without loss assume X, Y are unbounded on both sides, and admit densities. Let Z and W be constructed as above, then for each $\varepsilon > 0$, $X + Z_\varepsilon$ dominates $Y + W_\varepsilon$ in both tails, and $K_a(X + Z_\varepsilon) > K_a(X) \geq K_a(Y) > K_a(Y + W_\varepsilon)$ for every $a \in \mathbb{R}$, where the inequalities are strict as Z, W are not identically zero.

Thus the pair $X + Z_\varepsilon$ and $Y + W_\varepsilon$ satisfy the assumptions in Lemma 7. We can then find an independent random variable $V \in L_M$ (depending on ε), such that

$$X + Z_\varepsilon + V \geq_1 Y + W_\varepsilon + V.$$

Monotonicity and additivity of Φ then imply $\Phi(X)+\Phi(Z_\varepsilon) \geq \Phi(Y)+\Phi(W_\varepsilon)$, after canceling out $\Phi(V)$. The desired result $\Phi(X) \geq \Phi(Y)$ follows from the lemma below, which shows that our perturbations only slightly affect the statistic value. $\square$

Lemma 12. *For any $Z \in L_M$ with $Z \geq 0$, it holds that $\Phi(Z_\varepsilon) \to 0$ as $\varepsilon \to 0$. Similarly $\Phi(W_\varepsilon) \to 0$ for any $W \in L_M$ with $W \leq 0$.*

Proof. We focus on the case for Z_ε. Suppose for contradiction that $\Phi(Z_\varepsilon)$ does not converge to zero. Note that as ε decreases, Z_ε decreases in first-order stochastic dominance. So $\Phi(Z_\varepsilon) \geq 0$ also decreases, and non-convergence must imply there exists some $\delta > 0$ such that $\Phi(Z_\varepsilon) > \delta$ for every $\varepsilon > 0$. Let μ_ε be image measure of Z_ε. We now choose a sequence ε_n that decreases to zero very fast, and consider the measures

$$\nu_n = \mu_{\varepsilon_n}^{*n},$$

which is the n-th convolution power of μ_{ε_n}. Thus the sum of n i.i.d. copies of Z_{ε_n} is a random variable whose image measure is ν_n. We denote this sum by U_n.

For each n we choose ε_n sufficiently small to satisfy two properties: (i) $\varepsilon_n \leq \frac{1}{n^2}$, and (ii) it holds that

$$\mathbb{E}\left[\mathrm{e}^{nU_n} - 1\right] \leq 2^{-n}.$$

This latter inequality can be achieved because $\mathbb{E}\left[\mathrm{e}^{nU_n}\right] = \left(\mathbb{E}\left[\mathrm{e}^{nZ_{\varepsilon_n}}\right]\right)^n$, and as $\varepsilon_n \to 0$ we also have $\mathbb{E}\left[\mathrm{e}^{nZ_{\varepsilon_n}}\right] = 1 - \varepsilon_n + \varepsilon_n \mathbb{E}\left[\mathrm{e}^{nZ}\right] \to 1$ since $Z \in L_M$.

For these choices of ε_n and corresponding U_n, let $H_n(x)$ denote the c.d.f. of U_n, and define $H(x) = \inf_n H_n(x)$ for each $x \in \mathbb{R}$. Since $H_n(x) = 0$ for $x < 0$, the same is true for

$H(x)$. Also note that each $H_n(x)$ is a non-decreasing and right-continuous function in x, and so is $H(x)$.

We claim that $\lim_{x\to\infty} H(x) = 1$. Indeed, recall that U_n is the n-fold sum of Z_{ε_n}, which has mass $1-\varepsilon_n$ at zero. So U_n has mass at least $(1-\varepsilon_n)^n \geq (1-\frac{1}{n^2})^n \geq 1-\frac{1}{n}$ at zero. In other words, $H_n(0) \geq 1-\frac{1}{n}$. By considering the finitely many c.d.f.s $H_1(x), H_2(x), \ldots, H_{n-1}(x)$, we can find N such that $H_i(x) \geq 1-\frac{1}{n}$ for every $i < n$ and $x \geq N$. Together with $H_i(x) \geq H_i(0) \geq 1-\frac{1}{i} \geq 1-\frac{1}{n}$ for $i \geq n$, we conclude that $H_i(x) \geq 1-\frac{1}{n}$ whenever $x \geq N$, and so $H(x) \geq 1-\frac{1}{n}$. Since n is arbitrary, the claim follows. The fact that $H_n(x) \geq 1-\frac{1}{n}$ also shows that in the definition $H(x) = \inf_n H_n(x)$, the "inf" is actually achieved as the minimum.

These properties of $H(x)$ imply that it is the c.d.f. of some non-negative random variable U. We next show $U \in L_M$, i.e., $\mathbb{E}\left[\mathrm{e}^{aU}\right] < \infty$ for every $a \in \mathbb{R}$. Since $U \geq 0$, we only need to consider $a \geq 0$. To do this, we take advantage of the following identity based on integration by parts:

$$\mathbb{E}\left[\mathrm{e}^{aU_n}-1\right] = -\int_{x\geq 0}(\mathrm{e}^{ax}-1)\,\mathrm{d}(1-H_n(x)) = a\int_{x\geq 0}\mathrm{e}^{ax}(1-H_n(x))\,\mathrm{d}x.$$

Now recall that we chose U_n so that $\mathbb{E}\left[\mathrm{e}^{nU_n}-1\right] \leq 2^{-n}$. So $\mathbb{E}\left[\mathrm{e}^{aU_n}-1\right] \leq 2^{-n}$ for every positive integer $n \geq a$. It follows that the sum $\sum_{n=1}^{\infty}\mathbb{E}\left[\mathrm{e}^{aU_n}-1\right]$ is finite for every $a \geq 0$. Using the above identity, we deduce that

$$a\int_{x\geq 0}\mathrm{e}^{ax}\sum_{n=1}^{\infty}(1-H_n(x))\,\mathrm{d}x < \infty,$$

where we have switched the order of summation and integration by the Monotone Convergence Theorem. Since $H(x) = \min_n H_n(x)$, it holds that $1-H(x) \leq \sum_{n=1}^{\infty}(1-H_n(x))$ for every x. And thus

$$\mathbb{E}\left[\mathrm{e}^{aU}-1\right] = a\int_{x\geq 0}\mathrm{e}^{ax}(1-H(x))\,\mathrm{d}x < \infty$$

also holds. This proves $U \in L_M$.

We are finally in a position to deduce a contradiction. Since by construction the c.d.f. of U is no larger than the c.d.f. of each U_n, we have $U \geq_1 U_n$ and $\Phi(U) \geq \Phi(U_n)$ by monotonicity of Φ. But $\Phi(U_n) = n\Phi(Z_{\varepsilon_n}) > n\delta$ by additivity, so this leads to $\Phi(U)$ being infinite. This contradiction proves the desired result. □

D.4 Step 4: Functional Analysis

To complete the proof of Theorem 2, we also need to modify the functional analysis step in our earlier proof of Theorem 1. One difficulty is that for an unbounded random variable X, $K_a(X)$ takes the value ∞ as $a \to \infty$. Thus we can no longer think of $K_X(a) = K_a(X)$ as a real-valued continuous function on $\overline{\mathbb{R}}$.

We remedy this as follows. Note first that if Φ is a monotone additive statistic defined on L_M, then it is also monotone and additive when restricted to the smaller domain of bounded random variables. Thus Theorem 1 gives a probability measure μ on $\mathbb{R} \cup \{\pm\infty\}$ such that

$$\Phi(X) = \int_{\overline{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu(a)$$

for all $X \in L^\infty$. In what follows, μ is fixed. We just need to show that this representation also holds for $X \in L_M$.

As a first step, we show μ does not put any mass on $\pm\infty$. Indeed, if $\mu(\{\infty\}) = \varepsilon > 0$, then for any bounded random variable $X \geq 0$, the above integral gives $\Phi(X) \geq \varepsilon \cdot \max[X]$. Take any $Y \in L_M$ such that $Y \geq 0$ and Y is unbounded from above. Then monotonicity of Φ gives $\Phi(Y) \geq \Phi(\min\{Y, n\}) \geq \varepsilon \cdot n$ for each n. This contradicts $\Phi(Y)$ being finite. Similarly we can rule out any mass at $-\infty$.

The next lemma gives a way to extend the representation to certain unbounded random variables.

Lemma 13. *Suppose $Z \in L_M$ is bounded from below by 1 and unbounded from above, while $Y \in L_M$ is bounded from below and satisfies $\lim_{a\to\infty} \frac{K_a(Y)}{K_a(Z)} = 0$, then*

$$\Phi(Y) = \int_{(-\infty,\infty)} K_a(Y)\,\mathrm{d}\mu(a).$$

Proof. Given the assumptions, $K_a(Z) \geq 1$ for all $a \in \mathbb{R}$, with $\lim_{a\to\infty} K_a(Z) = \infty$. Let L_M^Z be the collection of random variables $X \in L_M$ such that X is bounded from below, and $\lim_{a\to\infty} \frac{K_a(X)}{K_a(Z)}$ exists and is finite. L_M^Z includes all bounded X (in which case $\lim_{a\to\infty} \frac{K_a(X)}{K_a(Z)} = 0$), as well as Y and Z itself. L_M^Z is also closed under adding independent random variables.

Now, for each $X \in L_M^Z$, we can define

$$K_{X|Z}(a) = \frac{K_a(X)}{K_a(Z)},$$

which reduces to our previous definition of $K_X(a)$ when Z is the constant 1. This function $K_{X|Z}(a)$ extends by continuity to $a = -\infty$, where its value is $\frac{\min[X]}{\min[Z]}$, as well as to $a = \infty$ by definition of L_M^Z. Thus $K_{X|Z}(\cdot)$ is a continuous function on $\overline{\mathbb{R}}$.

Since Φ induces an additive statistic when restricted to L_M^Z, and $K_{X|Z} + K_{Y|Z} = K_{X+Y|Z}$, we have an additive functional F defined on $\mathcal{L} = \{K_{X|Z} : X \in L_M^Z\}$, given by

$$F(K_{X|Z}) = \frac{\Phi(X)}{\Phi(Z)}.$$

Because $Z \geq 1$ implies $\Phi(Z) \geq 1$, F is well-defined, and $F(1) = 1$. By Lemma 11, F is also monotone in the sense that $K_{X|Z}(a) \geq K_{Y|Z}(a)$ for each $a \in \overline{\mathbb{R}}$ implies $F(K_{X|Z}) \geq F(K_{Y|Z})$.

Likewise we can show F is 1-Lipschitz. Note that $K_{X|Z}(a) \leq K_{Y|Z}(a) + \frac{m}{n}$ is equivalent to $K_a(X) \leq K_a(Y) + \frac{m}{n} K_a(Z)$ and equivalent to $K_a(X^{*n}) \leq K_a(Y^{*n} + Z^{*m})$, where we use the notation X^{*n} to denote the sum of n i.i.d. copies of X. If this holds for all a, then by Lemma 11 we also have $\Phi(X^{*n}) \leq \Phi(Y^{*n} + Z^{*m})$, and thus $\Phi(X) \leq \Phi(Y) + \frac{m}{n}\Phi(Z)$ by additivity. An approximation argument shows that for any *real* number $\varepsilon > 0$, $K_{X|Z}(a) \leq K_{Y|Z}(a) + \varepsilon$ for all a implies $\Phi(X) \leq \Phi(Y) + \varepsilon\Phi(Z)$. Thus the functional F is 1-Lipschitz.

Given these properties, we can exactly follow the proof of Theorem 1 to extend the functional F to be a positive linear functional on the space of all continuous functions over $\overline{\mathbb{R}}$ (the majorization condition is again satisfied by constant functions, as $K_{Z|Z} = 1$). Therefore, by the Riesz Representation Theorem, we obtain a probability measure μ_Z on $\overline{\mathbb{R}}$ such that for all $X \in L_M^Z$,

$$\frac{\Phi(X)}{\Phi(Z)} = \int_{\overline{\mathbb{R}}} \frac{K_a(X)}{K_a(Z)} \, \mathrm{d}\mu_Z(a).$$

In particular, for any X bounded from below such that $\lim_{a \to \infty} \frac{K_a(X)}{K_a(Z)} = 0$, it holds that

$$\Phi(X) = \int_{[-\infty,\infty)} K_a(X) \cdot \frac{\Phi(Z)}{K_a(Z)} \, \mathrm{d}\mu_Z(a),$$

where we are able to exclude ∞ from the range of integration (this is useful below).

If we define the measure $\hat{\mu}_Z$ by $\frac{\mathrm{d}\hat{\mu}_Z}{\mathrm{d}\mu_Z}(a) = \frac{\Phi(Z)}{K_a(Z)} \leq \Phi(Z)$, then since $K_a(X)$ is finite for $a < \infty$, we have

$$\Phi(X) = \int_{[-\infty,\infty)} K_a(X) \, \mathrm{d}\hat{\mu}_Z(a).$$

This in particular holds for all bounded X, so plugging in $X = 1$ gives that $\hat{\mu}_Z$ is a probability measure. But now we have two probability measures μ and $\hat{\mu}_Z$ on $\overline{\mathbb{R}}$ that lead to the same integral representation for bounded random variables, so Lemma 5 implies that $\hat{\mu}_Z$ coincides with μ and is supported on the standard real line. Plugging in $X = Y$ in the above display then yields the desired result. □

The next lemma further extends the representation:

Lemma 14. *For every $X \in L_M$ that is bounded from below,*

$$\Phi(X) = \int_{(-\infty,\infty)} K_a(X) \, \mathrm{d}\mu(a).$$

Proof. It suffices to consider X that is unbounded from above. Moreover, without loss we can assume $X \geq 0$,, since we can add any constant to X. Given the previous lemma, we just need to construct $Z \geq 1$ such that $\lim_{a \to \infty} \frac{K_a(X)}{K_a(Z)} = 0$. Note that $\mathbb{E}\left[\mathrm{e}^{aX}\right]$ strictly increases in a for $a \geq 0$. This means we can uniquely define a sequence $a_1 < a_2 < \cdots$ by the equation $\mathbb{E}\left[\mathrm{e}^{a_n X}\right] = \mathrm{e}^n$. This sequence diverges as $n \to \infty$. We then choose any increasing sequence b_n such that $b_n > n$ and $a_n b_n > 2n^2$.

Consider the random variable Z that is equal to b_n with probability $\mathrm{e}^{-\frac{a_n b_n}{2}}$ for each n, and equal to 1 with remaining probability. To see that $Z \in L_M$, we have

$$\mathbb{E}\left[\mathrm{e}^{aZ}\right] \leq \mathrm{e}^{a} + \sum_{n=1}^{\infty} \mathrm{e}^{-\frac{a_n b_n}{2}} \cdot \mathrm{e}^{a b_n} = \mathrm{e}^{a} + \sum_{n=1}^{\infty} \mathrm{e}^{(a-\frac{a_n}{2})\cdot b_n}.$$

For any fixed a, $\frac{a_n}{2}$ is eventually greater than $a+1$. This, together with the fact that $b_n > n$, implies the above sum converges.

Moreover, for any $a \in [a_n, a_{n+1})$, we have

$$\mathbb{E}\left[\mathrm{e}^{aZ}\right] \geq \mathbb{E}\left[\mathrm{e}^{a_n Z}\right] \geq \mathbb{P}[Z = b_n] \cdot \mathrm{e}^{a_n b_n} \geq \mathrm{e}^{\frac{a_n b_n}{2}} > \mathrm{e}^{n^2},$$

whereas $\mathbb{E}\left[\mathrm{e}^{aX}\right] \leq \mathbb{E}\left[\mathrm{e}^{a_{n+1} X}\right] \leq \mathrm{e}^{n+1}$. Thus

$$\frac{K_a(X)}{K_a(Z)} = \frac{\log \mathbb{E}\left[\mathrm{e}^{aX}\right]}{\log \mathbb{E}\left[\mathrm{e}^{aZ}\right]} \leq \frac{n+1}{n^2},$$

which converges to zero as a (and thus n) approaches infinity. □

D.5 Step 5: Wrapping Up

By a symmetric argument, the representation $\Phi(X) = \int_{(-\infty,\infty)} K_a(X)\,\mathrm{d}\mu(a)$ also holds for all X bounded from above. In the remainder of the proof, we will use an approximation argument to generalize this to all $X \in L_M$. We first show a technical lemma:

Lemma 15. *The measure μ is supported on a compact interval of $\mathbb{R}$.*

Proof. Suppose not, and without loss assume the support of μ is unbounded from above. We will construct a non-negative $Y \in L_M$ such that $\Phi(Y) = \infty$ according to the integral representation. Indeed, by assumption we can find a sequence $2 < a_1 < a_2 < \cdots$ such that $a_n \to \infty$ and $\mu([a_n, \infty)) \geq \frac{1}{n}$ for all large n. Let Y be the random variable that equals n with probability $\mathrm{e}^{-\frac{a_n \cdot n}{2}}$ for each n, and equals 0 with remaining probability. Then similar to the above, we can show $Y \in L_M$. Moreover, $\mathbb{E}\left[\mathrm{e}^{a_n Y}\right] \geq \mathrm{e}^{\frac{a_n \cdot n}{2}}$, implying that $K_{a_n}(Y) \geq \frac{n}{2}$. Since $K_a(Y)$ is increasing in a, we deduce that for each n,

$$\int_{[a_n,\infty)} K_a(Y)\,\mathrm{d}\mu(a) \geq K_{a_n}(Y) \cdot \mu([a_n, \infty)) \geq \frac{n}{2} \cdot \frac{1}{n} = \frac{1}{2}.$$

The fact that this holds for $a_n \to \infty$ contradicts the assumption that $\Phi(Y) = \int_{(-\infty,\infty)} K_a(Y)\,\mathrm{d}\mu(a)$ is finite. □

Thus we can take N sufficiently large so that μ is supported on $[-N, N]$. To finish the proof, consider any $X \in L_M$ that may be unbounded on both sides. For each positive integer n, let $X_n = \min\{X, n\}$ denote the truncation of X at n. Since $X \geq_1 X_n$, we have

$$\Phi(X) \geq \Phi(X_n) = \int_{[-N,N]} K_a(X_n)\,\mathrm{d}\mu(a)$$

Observe that for each $a \in [-N, N]$, $K_a(X_n)$ converges to $K_a(X)$ as $n \to \infty$. Moreover, the fact that $K_a(X_n)$ increases both in n and in a implies that for all a and all n,

$$|K_a(X_n)| \le \max\{|K_a(X_1)|, |K_a(X)|\} \le \max\{|K_{-N}(X_1)|, |K_N(X_1)|, |K_{-N}(X)|, |K_N(X)|\}.$$

As $K_a(X_n)$ is uniformly bounded, we can apply the Dominated Convergence Theorem to deduce

$$\Phi(X) \ge \lim_{n\to\infty} \int_{[-N,N]} K_a(X_n)\,\mathrm{d}\mu(a) = \int_{[-N,N]} K_a(X)\,\mathrm{d}\mu(a).$$

On the other hand, if we truncate the left tail and consider $X^{-n} = \max\{X, -n\}$, then a symmetric argument shows

$$\Phi(X) \le \lim_{n\to\infty} \int_{[-N,N]} K_a(X^{-n})\,\mathrm{d}\mu(a) = \int_{[-N,N]} K_a(X)\,\mathrm{d}\mu(a).$$

Therefore for all $X \in L_M$ it holds that

$$\Phi(X) = \int_{[-N,N]} K_a(X)\,\mathrm{d}\mu(a).$$

This completes the entire proof of Theorem 2.

E Proof of Proposition 5

Since the preference $\succeq$ is represented by Φ, the betweenness axiom is equivalent to the following:

$$\Phi(X) = \Phi(Y) \text{ if and only if } \Phi(X_\lambda Y) = \Phi(Y).$$

In this case, we say that the statistic Φ satisfies betweenness. We need to show that $\Phi(X)$ satisfies betweenness if and only if it is equal to $K_a(X)$ for some $a \in \mathbb{R}$ or equal to $\beta K_{-a\beta}(X) + (1-\beta)K_{a(1-\beta)}(X)$ for some $\beta \in (0,1)$ and $a \in (0,\infty)$.

We first show the "if" direction. Specifically, when $\Phi(X) = K_a(X)$ for some $a \in \mathbb{R}$, then the preference is CARA expected utility, which satisfies independence and thus betweenness. When $\Phi(X) = \beta K_{-a\beta}(X) + (1-\beta)K_{a(1-\beta)}(X)$, we can use the definition of K to rewrite it as

$$\Phi(X) = \frac{1}{a}\left(\log \mathbb{E}[\mathrm{e}^{a(1-\beta)X}] - \log \mathbb{E}[\mathrm{e}^{-a\beta X}]\right).$$

Thus $\Phi(X) = \Phi(Y)$ if and only if $\log \mathbb{E}\left[\mathrm{e}^{a(1-\beta)X}\right] - \log \mathbb{E}\left[\mathrm{e}^{-a\beta X}\right] = \log \mathbb{E}\left[\mathrm{e}^{a(1-\beta)Y}\right] - \log \mathbb{E}\left[\mathrm{e}^{-a\beta Y}\right]$, which in turn is equivalent to

$$\frac{\mathbb{E}\left[\mathrm{e}^{a(1-\beta)X}\right]}{\mathbb{E}\left[\mathrm{e}^{a(1-\beta)Y}\right]} = \frac{\mathbb{E}\left[\mathrm{e}^{-a\beta X}\right]}{\mathbb{E}\left[\mathrm{e}^{-a\beta Y}\right]}.$$

Since $\mathbb{E}\left[\mathrm{e}^{bX_\lambda Y}\right] = \lambda\mathbb{E}\left[\mathrm{e}^{bX}\right] + (1-\lambda)\mathbb{E}\left[\mathrm{e}^{bY}\right]$ for every $b \in \mathbb{R}$, it is not difficult to see that the above ratio equality holds if and only if it holds when X is replaced by $X_\lambda Y$. Hence $\Phi(X) = \Phi(Y)$ if and only if $\Phi(X_\lambda Y) = \Phi(Y)$, i.e. betweenness is satisfied.

Turning to the "only if" direction. We will characterize any monotone additive statistic Φ that satisfies a weaker form of betweenness:

Lemma 16. *Suppose Φ is a monotone additive statistic such that $\Phi(X) = c$ implies $\Phi(X_\lambda c) = c$ whenever c is a constant. Then either Φ takes the form described by Proposition 5, or $\Phi(X) = \beta \min[X] + (1-\beta)\max[X]$ for some $\beta \in [0,1]$.*

This result implies Proposition 5 because $\Phi(X) = \beta \min[X] + (1-\beta)\max[X]$ violates the original betweenness axiom. To see that, let $X = 0$ and choose any Y supported on ± 1. Then $X_\lambda Y$ and Y have the same minimum and maximum, so that $\Phi(X_\lambda Y) = \Phi(Y)$. But $\Phi(X) = \Phi(Y)$ cannot hold for *all* Y supported on ± 1.

The proof of Lemma 16 is in turn based on the following lemma which further relaxes betweenness:

Lemma 17. *Suppose $\Phi(X) = \int_{\mathbb{R}} K_a(X)\,\mathrm{d}\mu(a)$ has the property that $\Phi(X) = c$ implies $\Phi(X_\lambda c) \le c$. Then the measure μ restricted to $[0,\infty]$ is either the zero measure, or it is supported on a single point.*

Proof. It suffices to show that if μ puts positive mass on $(0,\infty]$, then that mass is supported on a single point and $\mu(\{0\}) = 0$. For this let $N > 0$ denote the essential maximum of the support of μ; that is, $N = \min\{x : \mu((x,\infty]) = 0\}$. We allow $N = \infty$ when the support of μ is unbounded from above, or when μ has a non-zero mass at ∞. For any positive real number $b < N$, consider the same random variable $X_{n,b}$ as in the proof of Lemma 5, given by

$$\mathbb{P}\left[X_{n,b} = n\right] = \mathrm{e}^{-bn}$$
$$\mathbb{P}\left[X_{n,b} = 0\right] = 1 - \mathrm{e}^{-bn}.$$

As shown in the proof of Lemma 5, $\frac{1}{n}K_a(X_{n,b})$ is uniformly bounded in $[0,1]$, and

$$\lim_{n\to\infty} \frac{1}{n} K_a(X_{n,b}) = \frac{(a-b)^+}{a}.$$

Thus if we let $c_n = \Phi(X_{n,b})$, then by the Dominated Convergence Theorem,

$$\lim_{n\to\infty} \frac{c_n}{n} = \lim_{n\to\infty} \frac{1}{n}\Phi(X_{n,b}) = \lim_{n\to\infty} \int_{\mathbb{R}} \frac{1}{n} K_a(X_{n,b})\,\mathrm{d}\mu(a) = \int_{(b,\infty]} \frac{a-b}{a}\,\mathrm{d}\mu(a).$$

Denote $\gamma = \int_{(b,\infty]} \frac{a-b}{a}\,\mathrm{d}\mu(a)$. This number γ is strictly positive because $b < N$ implies $\mu((b,\infty]) > 0$. We can also assume $\gamma < 1$, since otherwise μ must be the point mass at ∞.

Now, as $\Phi(X_{n,b}) = c_n$ we know by assumption that $\Phi(Y_{n,b}) \leq c_n$ for each n, where $Y_{n,b}$ is the mixture between $X_{n,b}$ and the constant c_n (in what follows λ is fixed as n varies):

$$\begin{aligned}
\mathbb{P}\left[Y_{n,b} = n\right] &= \lambda \mathrm{e}^{-bn} \\
\mathbb{P}\left[Y_{n,b} = 0\right] &= \lambda(1 - \mathrm{e}^{-bn}) \\
\mathbb{P}\left[Y_{n,b} = c_n\right] &= 1 - \lambda.
\end{aligned}$$

Using $\lim_{n\to\infty} c_n/n = \gamma$, we have

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} K_a(Y_{n,b}) &= \lim_{n\to\infty} \frac{1}{n}\frac{1}{a} \log\left[\lambda\left(1 - \mathrm{e}^{-bn} + \mathrm{e}^{(a-b)n}\right) + (1-\lambda)\mathrm{e}^{a\cdot c_n}\right] \\
&= \begin{cases} 0 & \text{if } a < 0 \\ (1-\lambda)\gamma & \text{if } a = 0 \\ \gamma & \text{if } 0 < a < \frac{b}{1-\gamma} \\ \frac{a-b}{a} & \text{if } a \geq \frac{b}{1-\gamma}. \end{cases}
\end{aligned}$$

Note that the cutoff point $a = \frac{b}{1-\gamma}$ is where $a - b = a\gamma$. When a is smaller than this, the dominant term in the bracketed sum above is $(1-\lambda)\mathrm{e}^{a\cdot c_n}$. Whereas for larger a, the dominant term becomes $\lambda \mathrm{e}^{(a-b)\cdot n}$.

Crucially, $\lim_{n\to\infty} \frac{1}{n} K_a(Y_{n,b}) \geq \frac{(a-b)^+}{a}$ holds for every a, with strict inequality for $a \in [0, \frac{b}{1-\gamma})$. Thus again by the Dominated Convergence Theorem,

$$\lim_{n\to\infty} \frac{c_n}{n} \geq \lim_{n\to\infty} \frac{1}{n}\Phi(Y_{n,b}) = \lim_{n\to\infty} \int_{\mathbb{R}} \frac{1}{n} K_a(Y_{n,b})\,\mathrm{d}\mu(a) \geq \int_{(b,\infty]} \frac{a-b}{a}\,\mathrm{d}\mu(a).$$

But we know that the far left is equal to the far right. So both inequalities hold equal, and in particular $\lim_{n\to\infty} \frac{1}{n} K_a(Y_{n,b}) = \frac{(a-b)^+}{a}$ holds μ-almost surely.

As discussed, $\lim_{n\to\infty} \frac{1}{n} K_a(Y_{n,b}) > \frac{(a-b)^+}{a}$ for any $a \in [0, \frac{b}{1-\gamma})$. So we can conclude that $\mu([0, \frac{b}{1-\gamma})) = 0$. This must hold for any $b \in (0, N)$ and corresponding γ. Letting b arbitrarily close to N thus yields $\mu([0, N)) = 0$ (since $\frac{b}{1-\gamma} > b$). It follows that when restricted to $[0, \infty]$ the measure μ is concentrated at the single point N, as we desire to show. □

Proof of Lemma 16. From Lemma 17, we know that the measure μ associated with Φ can only be supported on one point in all of $[0, \infty]$. By a symmetric argument, μ also has at most one point support in all of $[-\infty, 0]$. Thus either $\mu = \delta_a$ for some $a \in \overline{\mathbb{R}}$, or μ is supported on two points $\{a_1, a_2\}$ with $a_1 < 0 < a_2$. In the former case we are done, so below we study the latter case where μ has two-point support.

Suppose $\Phi(X) = \beta K_{a_1}(X) + (1-\beta)K_{a_2}(X)$ for some $\beta \in (0, 1)$ and $a_1 < 0 < a_2$. If $a_1 = -\infty$ while $a_2 < \infty$, then $\Phi(X) = \beta \min[X] + (1-\beta)K_{a_2}(X)$. Take any non-constant X

and let c denote $\Phi(X)$. Note that since $K_{a_2}(X) > \min[X]$, $c = \beta \min[X] + (1-\beta) K_{a_2}(X)$ lies strictly between $\min[X]$ and $K_{a_2}(X)$. Consider the mixture $X_\lambda c$, then $\min[X_\lambda c] = \min[X]$, whereas

$$K_{a_2}(X_\lambda c) = \frac{1}{a_2} \log \left(\lambda \mathbb{E}\left[\mathrm{e}^{a_2 X}\right] + (1-\lambda)\mathrm{e}^{a_2 c}\right) < \frac{1}{a_2} \log \mathbb{E}\left[\mathrm{e}^{a_2 X}\right] = K_{a_2}(X),$$

where the inequality uses $c < K_{a_2}(X) = \frac{1}{a_2} \log \mathbb{E}\left[\mathrm{e}^{a_2 X}\right]$ and $a_2 > 0$. We thus deduce that

$$\Phi(X_\lambda c) = \beta \min[X_\lambda c] + (1-\beta) K_{a_2}(X_\lambda c) < \beta \min[X] + (1-\beta) K_{a_2}(X) = c,$$

contradicting the betweenness axiom. A symmetric argument rules out the possibility that $a_1 > -\infty$ while $a_2 = \infty$.

Hence, either $a_1 = -\infty$ and $a_2 = \infty$, or $a_1 \in (-\infty, 0)$ and $a_2 \in (0, \infty)$. In the former case $\Phi(X)$ is an average of the minimum and the maximum, so we are again done. It remains to consider the latter case where a_1, a_2 are both finite. In this case we will show that $\beta = \frac{-a_1}{a_2 - a_1}$. Once this is shown, we can let $a = a_2 - a_1$ so that $a_1 = -a\beta$ and $a_2 = a(1-\beta)$. Thus $\Phi(X) = \beta K_{-a\beta}(X) + (1-\beta) K_{a(1-\beta)}(X)$ as desired.

Let us take an arbitrary non-constant X, and let

$$c = \Phi(X) = \frac{\beta}{a_1} \log \mathbb{E}\left[\mathrm{e}^{a_1 X}\right] + \frac{1-\beta}{a_2} \log \mathbb{E}\left[\mathrm{e}^{a_2 X}\right].$$

For an arbitrary $\lambda \in [0,1]$, we must also have

$$c = \Phi(X_\lambda c) = \frac{\beta}{a_1} \log \mathbb{E}\left[\lambda \mathrm{e}^{a_1 X} + (1-\lambda)\mathrm{e}^{a_1 c}\right] + \frac{1-\beta}{a_2} \log \mathbb{E}\left[\lambda \mathrm{e}^{a_2 X} + (1-\lambda)\mathrm{e}^{a_2 c}\right]. \tag{18}$$

Since (18) holds for every λ, we can differentiate it with respect to λ to obtain

$$0 = \frac{\beta(\mathbb{E}\left[\mathrm{e}^{a_1 X}\right] - \mathrm{e}^{a_1 c})}{a_1 \mathbb{E}\left[\lambda \mathrm{e}^{a_1 X} + (1-\lambda)\mathrm{e}^{a_1 c}\right]} + \frac{(1-\beta)(\mathbb{E}\left[\mathrm{e}^{a_2 X}\right] - \mathrm{e}^{a_2 c})}{a_2 \mathbb{E}\left[\lambda \mathrm{e}^{a_2 X} + (1-\lambda)\mathrm{e}^{a_2 c}\right]}.$$

Plugging in $\lambda = 0$ and $\lambda = 1$ gives, respectively,

$$\frac{\beta(\mathbb{E}\left[\mathrm{e}^{a_1 X}\right] - \mathrm{e}^{a_1 c})}{a_1 \mathrm{e}^{a_1 c}} = -\frac{(1-\beta)(\mathbb{E}\left[\mathrm{e}^{a_2 X}\right] - \mathrm{e}^{a_2 c})}{a_2 \mathrm{e}^{a_2 c}}. \tag{19}$$

$$\frac{\beta(\mathbb{E}\left[\mathrm{e}^{a_1 X}\right] - \mathrm{e}^{a_1 c})}{a_1 \mathbb{E}\left[\mathrm{e}^{a_1 X}\right]} = -\frac{(1-\beta)(\mathbb{E}\left[\mathrm{e}^{a_2 X}\right] - \mathrm{e}^{a_2 c})}{a_2 \mathbb{E}\left[\mathrm{e}^{a_2 X}\right]}. \tag{20}$$

Since $c = \beta K_{a_1}(X) + (1-\beta) K_{a_2}(X)$, the fact that $K_{a_2}(X) > K_{a_1}(X)$ implies c is strictly between $K_{a_1}(X)$ and $K_{a_2}(X)$. Thus, using $a_1 < 0 < a_2$ we deduce $\mathrm{e}^{a_1 c} < \mathbb{E}\left[\mathrm{e}^{a_1 X}\right]$ and $\mathrm{e}^{a_2 c} < \mathbb{E}\left[\mathrm{e}^{a_2 X}\right]$.

We can therefore divide (19) by (20) to obtain

$$\frac{\mathbb{E}\left[\mathrm{e}^{a_1 X}\right]}{\mathrm{e}^{a_1 c}} = \frac{\mathbb{E}\left[\mathrm{e}^{a_2 X}\right]}{\mathrm{e}^{a_2 c}}.$$

Plugging this back to (19), we conclude $\frac{\beta}{a_1} = -\frac{1-\beta}{a_2}$, so $\beta = \frac{-a_1}{a_2 - a_1}$ as we desire to show. $\square$

F Monotone Additive Statistics and the Independence Axiom

In this appendix we discuss the classic independence axiom and what it implies for preferences represented by monotone additive statistics.

Axiom F.1 (Independence)**.** *For all X, Y, Z and all $\lambda \in (0, 1)$, $X \succeq Y$ implies $X_\lambda Z \succeq Y_\lambda Z$.*

Proposition 7. *Suppose a preference $\succeq$ is represented by a monotone additive statistic $\Phi(X) = \int_{\overline{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu(a)$. Then $\succeq$ satisfies the independence axiom if and only if μ is a point mass at some $a \in \overline{\mathbb{R}}$.*

Proof. The "if" direction is relatively straightforward. If $a = 0$ then $\Phi(X) = \mathbb{E}[X]$. In this case $\mathbb{E}[X] \geq \mathbb{E}[Y]$ does imply

$$\mathbb{E}[X_\lambda Z] = \lambda\mathbb{E}[X] + (1-\lambda)\mathbb{E}[Z] \geq \lambda\mathbb{E}[Y] + (1-\lambda)\mathbb{E}[Z] = \mathbb{E}[Y_\lambda Z].$$

If $a > 0$ then $\Phi(X) \geq \Phi(Y)$ implies $\mathbb{E}\left[\mathrm{e}^{aX}\right] \geq \mathbb{E}\left[\mathrm{e}^{aY}\right]$ and thus

$$\lambda\mathbb{E}\left[\mathrm{e}^{aX}\right] + (1-\lambda)\mathbb{E}\left[\mathrm{e}^{aZ}\right] \geq \lambda\mathbb{E}\left[\mathrm{e}^{aY}\right] + (1-\lambda)\mathbb{E}\left[\mathrm{e}^{aZ}\right],$$

so that $\Phi(X_\lambda Z) \geq \Phi(Y_\lambda Z)$. A similar argument applies to the case of $a < 0$. Finally it is easy to see that $\max[X] \geq \max[Y]$ implies $\max[X_\lambda Z] \geq \max[Y_\lambda Z]$ and the same holds for the minimum. So the above independence axiom holds for $a = \pm\infty$ as well.[21]

We turn to the "only if" direction of the result. By the independence axiom, whenever c is a constant we have $X \succeq c$ implies $X_\lambda c \succeq c$ and $c \succeq X$ implies $c \succeq X_\lambda c$. Therefore $X \sim c$ implies $X_\lambda c \sim c$, which allows us to directly apply Lemma 16 from before. It remains to show that independence rules out $\Phi(X) = \beta K_{-a\beta}(X) + (1-\beta)K_{a(1-\beta)}(X)$ for some $\beta \in (0, 1)$ and $a \in (0, \infty]$.

Suppose Φ takes the above form. If $a = \infty$ then $\Phi(X) = \beta\min[X] + (1-\beta)\max[X]$ for some $\beta \in (0, 1)$. To see that it violates independence, choose X supported on 0 and $\frac{1}{1-\beta}$, and $Y = 1$ so that $\Phi(X) = \Phi(Y)$. But with Z being a sufficiently large constant we see that $X_\lambda Z$ has the same maximum as $Y_\lambda Z$, but a strictly smaller minimum. Hence $\Phi(X_\lambda Z) < \Phi(Y_\lambda Z)$, contradicting independence.

If instead $a \in (0, \infty)$, then we can do a similar construction by choosing X and Y such that $\Phi(X) > \Phi(Y)$ but $K_{-a\beta}(X) < K_{-a\beta}(Y)$. For example, let $Y = 1$, and let X be supported on $\{0, k\}$, with $\mathbb{P}[X = k] = \frac{1}{k}$. Then

$$K_b(X) = \frac{1}{b}\log\mathbb{E}\left[1 - \frac{1}{k} + \frac{\mathrm{e}^{bk}}{k}\right].$$

[21] Note however that $\Phi(X) = \max[X]$ or $\min[X]$ would violate a stronger form of independence that additionally requires $X \succ Y$ to imply $X_\lambda Z \succ Y_\lambda Z$ with strict preferences. This is related to the fact that these extreme monotone additive statistics do not satisfy mixture continuity.

For k tending to infinity, $K_b(X)$ tends to zero if $b < 0$, and to infinity if $b > 0$. Hence, for k large enough, X and Y will have the desired property.

Now let $Z = n$ where n is a large positive integer. Then

$$K_b(Y_\lambda n) = \frac{1}{b}\log \mathbb{E}\left[\lambda \mathbb{E}\left[\mathrm{e}^{bY}\right] + (1-\lambda)\mathrm{e}^{bn}\right]$$
$$K_b(X_\lambda n) = \frac{1}{b}\log \mathbb{E}\left[\lambda \mathbb{E}\left[\mathrm{e}^{bX}\right] + (1-\lambda)\mathrm{e}^{bn}\right]$$

and so

$$K_b(Y_\lambda n) - K_b(X_\lambda n) = \frac{1}{b}\log\left(\frac{\lambda \mathbb{E}\left[\mathrm{e}^{bY}\right] + (1-\lambda)\mathrm{e}^{bn}}{\lambda \mathbb{E}\left[\mathrm{e}^{bX}\right] + (1-\lambda)\mathrm{e}^{bn}}\right).$$

It easily follows that for fixed $\lambda \in (0,1)$ and b,

$$\lim_{n\to\infty} K_b(Y_\lambda n) - K_b(X_\lambda n) = 0 \;\text{ if } b > 0;$$
$$\lim_{n\to\infty} K_b(Y_\lambda n) - K_b(X_\lambda n) = K_b(Y) - K_b(X) \;\text{ if } b < 0.$$

Thus, as n tends to infinity,

$$\begin{aligned}
&\lim_n \Phi(Y_\lambda n) - \Phi(X_\lambda n) \\
&= \lim_n \beta\left[K_{-a\beta}(Y_\lambda n) - K_{-a\beta}(X_\lambda n)\right] + (1-\beta)\left[K_{a(1-\beta)}(Y_\lambda n) - K_{a(1-\beta)}(X_\lambda n)\right] \\
&= \beta\left[K_{-a\beta}(Y_\lambda n) - K_{-a\beta}(X_\lambda n)\right] > 0.
\end{aligned}$$

Therefore, for n large enough, we have found X and Y such that $\Phi(X) > \Phi(Y)$ but $\Phi(X_\lambda n) < \Phi(Y_\lambda n)$. This implies $X \succ Y$ but $X_\lambda n \prec Y_\lambda n$, which contradicts the independence axiom and completes the proof of Proposition 7. □

F.1 Proof of Proposition 1

We now prove Proposition 1 as a corollary of Proposition 7. Suppose that $\succeq$ is an MSTP that satisfies strong stationarity. Let $\succeq_*$ denote the preference over random times induced by $\succeq$ when fixing the payoff. That is, $T \succeq_* S$ if and only if $(x,T) \succeq (x,S)$ for any (and every) $x > 0$.

Fix any $X \succeq_* Y$ and any $Z \in L^\infty_+$, which can be considered as random times. For a given $\lambda \in (0,1)$, choose D to be a random variable that is equal to either 0 or 1, with probability λ and $1-\lambda$, respectively. Let $\tilde{X}$ be a random variable that conditioned on $D = 0$ has the same distribution as $X + 1$, and conditioned on $D = 1$ has the same distribution as Z. Likewise, let $\tilde{Y}$ be a random variable that conditioned on $D = 0$ has the same distribution as $Y + 1$, and conditioned on $D = 1$ has the same distribution as Z.

By construction $\tilde{X}_D \succeq_* \tilde{Y}_D$ for every possible value of D, so by strong stationarity $\tilde{X} + D \succeq_* \tilde{Y} + D$ must hold. But $\tilde{X} + D$ has the same distribution as $(X_\lambda Z) + 1$ while $\tilde{Y} + D$ has the same distribution as $(Y_\lambda Z) + 1$, which implies $(X_\lambda Z) + 1 \succeq_* (Y_\lambda Z) + 1$. Since this is an MSTP, we deduce $X_\lambda Z \succeq_* Y_\lambda Z$ as the independence axiom requires.

Note that even though $\succeq_*$ and the associated monotone additive statistic Φ are defined only for non-negative bounded random variables, it can be extended to all of L^∞ as shown in the proof of Proposition 6. Given additivity, it is easy to see that the extension preserves independence. So we can assume $\succeq_*$ and Φ satisfy independence on L^∞. This allows us to apply Proposition 7 and deduce that Φ must have a point-mass mixing measure μ, which proves the "only if" direction of Proposition 1.

As for the "if" direction, we need to verify that an MSTP represented by $V(x,T) = u(x) \cdot \mathrm{e}^{-rK_a(T)}$ does satisfy strong stationarity. First consider $a = 0$, in which case the representation simplifies to $u(x) \cdot \mathrm{e}^{-\mathbb{E}[T]}$ with the normalization $r = 1$. From the assumption $(x, T_D) \succeq (y, S_D)$ a.s. we obtain $u(x) \cdot \mathrm{e}^{-\mathbb{E}[T_D]} \geq u(y) \cdot \mathrm{e}^{-\mathbb{E}[S_D]}$, which can be rewritten as $\mathbb{E}[S_D] - \mathbb{E}[T_D] \geq \log(u(y)/u(x))$. Thus we can also add any fixed value of D to obtain

$$\mathbb{E}[S_D + D] - \mathbb{E}[T_D + D] \geq \log(u(y)/u(x)) \;\; a.s.$$

Averaging across different values of D , this implies $\mathbb{E}[S+D] - \mathbb{E}[T+D] \geq \log(u(y)/u(x))$, which after rearranging yields $u(x) \cdot \mathrm{e}^{-\mathbb{E}[T+D]} \geq u(y) \cdot \mathrm{e}^{-\mathbb{E}[S+D]}$. So $(x, T+D) \succeq (y, S+D)$ as demanded by strong stationarity.

Next consider $a > 0$. In this case we normalize $r = a$ and adjust u accordingly, to arrive at an equivalent representation $V(x,T) = u(x)/\mathbb{E}\left[\mathrm{e}^{aT}\right]$. From $(x, T_D) \succeq (y, S_D)$ we obtain $u(x) \cdot \mathbb{E}\left[\mathrm{e}^{aS_D}\right] \geq u(y) \cdot \mathbb{E}\left[\mathrm{e}^{aT_D}\right]$ and thus

$$u(x) \cdot \mathbb{E}\left[\mathrm{e}^{a(S_D+D)}\right] \geq u(y) \cdot \mathbb{E}\left[\mathrm{e}^{a(T_D+D)}\right] \;a.s.$$

Averaging across different values of D then yields $u(x) \cdot \mathbb{E}\left[\mathrm{e}^{a(S+D)}\right] \geq u(y) \cdot \mathbb{E}\left[\mathrm{e}^{a(T+D)}\right]$, which after rearranging gives the desired conclusion $V(x, T+D) \geq V(y, S+D)$.

If instead $a < 0$, then we normalize $r = -a$ and recover the usual EDU representation $V(x,T) = u(x) \cdot \mathbb{E}\left[\mathrm{e}^{aT}\right]$. Essentially the same argument as above applies to this case.

Finally consider $a = \infty$, so that $V(x,T) = u(x) \cdot \mathrm{e}^{-\max[T]}$ after normalizing $r = 1$. Here $(x, T_D) \succeq (y, S_D)$ implies $\max[S_D] - \max[T_D] \geq \log(u(y)/u(x))$, and thus

$$\max[S_D + D] - \max[T_D + D] \geq \log(u(y)/u(x)) \;\; a.s.$$

Let $\alpha = \max[S+D]$ and $c = \log(u(y)/u(x))$ be constants. Then the above implies that for almost every value of D, $T_D + D \leq \alpha - c$. Thus $T + D \leq \alpha - c$ almost surely, which gives $\max[S+D] - \max[T+D] \geq c$. This implies $V(x, T+D) \geq V(y, S+D)$ as desired.

A similar argument applies to the case of $a = -\infty$, completing the proof.