Appendix

The appendix contains the omitted proofs for most of the results in the main text, in the order in which they appeared. The only exceptions are Theorem 2 regarding the larger domain L_M, Proposition 1 regarding strong stochastic dynamic consistency and a few results in Section 4, whose proofs are relegated to the online appendix.

Throughout the proofs we will 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. *$K_X \colon \overline{\mathbb{R}} \to \mathbb{R}$ is well defined, non-decreasing and continuous.*

2. *If $K_X = K_Y$ then X and Y have the same distribution.*

Proof. Over $\mathbb{R}$ the map K_X is continuous and non-decreasing. This follows directly from the fact that $K_X(a)$ is the certainty equivalent of a CARA expected utility preference with coefficient of risk aversion equal to $-a$. That $\lim_{a\to\infty} K_X(a) = \max[X]$ and $\lim_{a\to-\infty} K_X(a) = \min[X]$ follow from a simple application of Laplace's method. It is a standard fact that $K_X = K_Y$ implies that X and Y have the same distribution (see for instance Curtiss, 1942). □

A Proof of Theorem 1

We follow the proof outlined in §5 of the main text and first establish Theorem 6.

A.1 Proof of Theorem 6

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.[20] 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:

$$\int \sigma(x) h(y-x) \, \mathrm{d}x = \mathrm{e}^{-\frac{y^2}{2V}} \cdot \int_{-N}^{N} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V} \cdot x} \cdot \mathrm{e}^{-\frac{x^2}{2V}} \, \mathrm{d}x. \tag{9}$$

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,

$$\begin{aligned} \int_{-N}^{N} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V} \cdot x} \cdot \mathrm{e}^{-\frac{x^2}{2V}} \, \mathrm{d}x &= M_\sigma\left(\frac{y}{V}\right) - \int_{-N}^{N} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V} \cdot x} \cdot (1 - \mathrm{e}^{-\frac{x^2}{2V}}) \, \mathrm{d}x \\ &\geq M_\sigma\left(\frac{y}{V}\right) - 2N \cdot \mathrm{e}^{AN} \cdot (1 - \mathrm{e}^{\frac{-N^2}{2V}}), \end{aligned} \tag{10}$$

which is positive when V is sufficiently large. So the right-hand side of (9) is positive.

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 $AV \geq N - \varepsilon$,

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

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: $F(K_X + K_Y) = F(K_X) + F(K_Y)$.[21] 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 6 (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 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:*

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

[21] 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 $K_X + K_Y = K_{X'} + K_{Y'} = K_{X'+Y'} \in \mathcal{L}$.

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. □

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. *Let $\mathcal{V}$ be 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 $X \in L^\infty$.[22] *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 and value 0 with probability $1-\varepsilon$. 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 $1 - \varepsilon$ and value 0 with probability ε. 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

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

[22] 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

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

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,

$$\lim_{n\to\infty}\int_{\overline{\mathbb{R}}}\frac{1}{n}K_a(X_{n,b})\,\mathrm{d}\mu(a) = \int_{[b,\infty]}\frac{a-b}{a}\,\mathrm{d}\mu(a), \tag{11}$$

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 7. *$\Phi\colon L^\infty_+ \to \mathbb{R}$ 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 $X \in L^\infty$*

$$\Phi(X) = \int_{\overline{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu(a). \tag{12}$$

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 stochastic 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, stochastic 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 $t \geq 0$, $v_i(t) = \mathrm{e}^{-a_i t}$ and $v(t) = \mathrm{e}^{-at}$. 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 $c \in \mathbb{R}$ such that for every t, $v(t) = \sum_i \lambda_i v_i(t) + c$. By letting $t \to \infty$ we obtain $0 = c$ and by setting $t = 0$ it follows that $1 = \sum_i \lambda_i$. Further plugging in $t = 1$ and $t = 2$, we obtain

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

But the Cauchy-Schwarz inequality gives

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

Thus equality holds. Since the individual discount rates $\{a_i\}$ are assumed to be distinct, the equality condition of the Cauchy-Schwarz inequality implies that exactly one λ_i is nonzero (in fact equal to 1), and hence $a = a_i$ for some agent i.

Without loss of generality suppose $a = a_1$. It remains to show that $u(x)$ is a constant multiple of $u_1(x)$ so that the social preference coincides with agent 1. Note that by the same argument as above, $v_1(t) = \mathrm{e}^{-a_1 t}$ cannot be expressed as a linear combination of $1, v_2(t), v_3(t), \cdots, v_n(t)$ whenever a_1 is distinct from $a_2, \cdots, a_n$. So the contrapositive of Harsanyi's Theorem implies the existence of random times S and T such that $\mathbb{E}\left[v_i(S)\right] \geq \mathbb{E}\left[v_i(T)\right]$ for all $i > 1$ but $\mathbb{E}\left[v_1(T)\right] > \mathbb{E}\left[v_1(S)\right]$. In what follows we fix these particular S and T, and also fix $\varepsilon > 0$ sufficiently small so that $\mathbb{E}\left[v_1(T)\right] \geq (1+\varepsilon)\mathbb{E}\left[v_1(S)\right]$.

For any pair of rewards $x, y \in \mathbb{R}_{++}$, we now show that the Pareto property implies $\frac{u(y)}{u_1(y)} = \frac{u(x)}{u_1(x)}$ which will complete the proof. To do this, let k be a sufficiently large positive integer, and define $T^{\oplus k}$, $S^{\oplus k}$ to be the random variables obtained by adding k independent copies of T and S. Since the moment generating function $\mathbb{E}\left[\mathrm{e}^{-\alpha Z}\right]$ is multiplicative when we add two independent random variables Z_1 and Z_2, our previous assumptions about S and T imply that $\mathbb{E}\left[\mathrm{e}^{-a_i S^{\oplus k}}\right] \geq \mathbb{E}\left[\mathrm{e}^{-a_i T^{\oplus k}}\right]$ for all $i > 1$ but $\mathbb{E}\left[\mathrm{e}^{-a_1 T^{\oplus k}}\right] \geq (1+\varepsilon)^k \mathbb{E}\left[\mathrm{e}^{-a_1 S^{\oplus k}}\right]$.

Next, let $t_k \in \mathbb{R}$ be the number that satisfies

$$\mathrm{e}^{-a_1 t_k} \cdot u_1(x)\mathbb{E}\left[\mathrm{e}^{-a_1 T^{\oplus k}}\right] = u_1(y)\mathbb{E}\left[\mathrm{e}^{-a_1 S^{\oplus k}}\right].$$

Thus, the time lottery $(x, T^{\oplus k} + t_k)$ is indifferent to $(y, S^{\oplus k})$ for agent 1. At the same time, the above equality implies $e^{a_1 t_k} \geq (1+\varepsilon)^k \cdot \frac{u_1(x)}{u_1(y)}$, so that $\lim_{k\to\infty} t_k = \infty$. In particular, we deduce that for k large, $\mathrm{e}^{a_i t_k} \geq \frac{u_i(x)}{u_i(y)}$ for every $i > 1$ and thus

$$\mathrm{e}^{-a_i t_k} \cdot u_i(x)\mathbb{E}\left[\mathrm{e}^{-a_i T^{\oplus k}}\right] \leq u_i(y)\mathbb{E}\left[\mathrm{e}^{-a_i S^{\oplus k}}\right].$$

Therefore $(x, T^{\oplus k} + t_k)$ is less preferred than $(y, S^{\oplus k})$ for every agent $i > 1$.

Putting together the above analysis, we can find k and t_k such that $(x, T^{\oplus k} + t_k)$ is weakly less preferred than $(y, S^{\oplus k})$ for every agent, with indifference for agent 1. By the Pareto property, $(x, T^{\oplus k} + t_k)$ must be weakly less preferred than $(y, S^{\oplus k})$ under the social preference. That is, we must have

$$\mathrm{e}^{-a t_k} \cdot u(x)\mathbb{E}\left[\mathrm{e}^{-a T^{\oplus k}}\right] \leq u(y)\mathbb{E}\left[\mathrm{e}^{-a S^{\oplus k}}\right].$$

But we already know $\mathrm{e}^{-a_1 t_k} \cdot u_1(x)\mathbb{E}\left[\mathrm{e}^{-a_1 T^{\oplus k}}\right] = u_1(y)\mathbb{E}\left[\mathrm{e}^{-a_1 S^{\oplus k}}\right]$ and $a = a_1$, so after dividing out $e^{-a t_k}$, $\mathbb{E}\left[\mathrm{e}^{-a T^{\oplus k}}\right]$ and $\mathbb{E}\left[\mathrm{e}^{-a S^{\oplus k}}\right]$ we obtain $\frac{u(y)}{u_1(y)} \geq \frac{u(x)}{u_1(x)}$.

Finally, since x, y are arbitrary, we can switch them and use the same argument to deduce the opposite inequality $\frac{u(x)}{u_1(x)} \geq \frac{u(y)}{u_1(y)}$. This proves that $\frac{u(y)}{u_1(y)} = \frac{u(x)}{u_1(x)}$ for any pair of rewards x, y. Hence the social utility representation is a constant multiple of agent 1's.

B.4 Proof of Proposition 3

We 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 $u_i(x)\mathrm{e}^{-r_i\Phi_i(T)} \geq u_i(y)\mathrm{e}^{-r_i\Phi_i(S)}$, which can be rewritten as

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

Summing across i using the weights λ_i we obtain

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

where the last equality uses $u = \Pi_{i=1}^{n} u_i^{\lambda_i}$. Since $r\Phi = \sum_{i=1}^{n} \lambda_i r_i \Phi_i$, it follows that $r(\Phi(S) - \Phi(T)) \geq \log \frac{u(y)}{u(x)}$, 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.

B.5 Proof of Proposition 4

We assume the Pareto axiom holds and deduce its implications. 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 $(\Phi_1, \dots, \Phi_n, \Phi)$ have the *Pareto property* if $\Phi_i(T) \leq \Phi_i(S)$ for every i implies $\Phi(T) \leq \Phi(S)$. We have the following result:

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

Proof. Let $(\mu_1, \dots, \mu_n, \mu)$ be the mixing measures on $\overline{\mathbb{R}}$ that correspond to the monotone additive statistics $(\Phi_1, \dots, \Phi_n, \Phi)$. Define the linear functionals $(I_1, \dots, I_n, I)$ 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 $f, g \in \mathcal{D}$,

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

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. □

Thus, the Pareto axiom implies that the social certainty equivalent Φ must be a convex combination of the individual Φ_i. To complete the proof, we restrict to the case of identical utility functions $u_i = u$ which additionally satisfies $\lim_{x \to 0} u(x) = 0$ or $\lim_{x \to \infty} u(x) = \infty$. In this case, in order for $u = \Pi_{i=1}^n u_i^{\lambda_i}$ to hold, the weights $\lambda_1, \ldots, \lambda_n$ must sum to 1. Therefore the desired identity $r\Phi = \sum_{i=1}^n \lambda_i r_i \Phi_i$ requires us to show that not only Φ is a convex combination of (Φ_i), but $r\Phi$ is also a convex combination of $(r_i \Phi_i)$.

To prove this, 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 $S, T \in L_+^\infty$,

$$r\Phi(S) - r\Phi(T) \geq \min_{1 \leq i \leq n} \{r_i \Phi_i(S) - r_i \Phi_i(T)\}. \tag{13}$$

The conclusion that $r\Phi$ is a convex combination of $(r_i\Phi_i)$ will follow from the condition (13) 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 $\Phi_1, \dots, \Phi_m$ are linearly independent, and every Φ_i is a (not necessarily positive) linear combination of those m. Thus we can find vectors $\gamma^1, \dots, \gamma^n \in \mathbb{R}^m$ such that every $r_i\Phi_i$ can be rewritten as the following inner product (i.e., linear combination):

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

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

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 $(\Phi_1(S) - \Phi_1(T), \dots, \Phi_m(S) - \Phi_m(T))$ for some $S, T \in L_+^\infty$. Condition (13) 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 $q_1w_1 + \cdots + q_kw_k$ 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 $(\Phi_1(S), \dots, \Phi_m(S))$, and such vectors cannot all belong to a lower-dimensional subspace by the assumption that $\Phi_1, \dots, \Phi_m$ are linearly independent.

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

$$\gamma \cdot w \geq \min_{1 \leq i \leq n} \gamma^i \cdot w \ \text{ for all } w \in \mathbb{R}^m. \tag{14}$$

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 $-\gamma^1, \dots, -\gamma^n$, and whose last row is γ. Let b be the $n+1$-dimensional vector $(-1-\varepsilon, \dots, -1-\varepsilon, 1)$. Then $Aw \leq b$ has no solution $w \in \mathbb{R}^m$.

By Farkas' Lemma, there exists a non-negative $n+1$-dimensional vector $z = (z_1, \dots, z_{n+1})$ 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 $\gamma^1, \dots, \gamma^n$ that gives the zero vector, leading to the impossible implication that a positive linear combination of $\Phi_1, \dots, \Phi_n$ 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 (14) 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 Proof of Theorem 4

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 7. *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 Theorem 4, or $\Phi(X) = \beta \min[X] + (1-\beta)\max[X]$ for some $\beta \in [0,1]$.*

This result implies Theorem 4 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 7 is in turn based on the following lemma which further relaxes betweenness:

Lemma 8. *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

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

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 $\Phi(Y_{n,b}) \le c_n$ 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

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

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 7. From Lemma 8, 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

$$c = \Phi(X_\lambda c) = \frac{\beta}{a_1} \log \mathbb{E}\left[\lambda \mathrm{e}^{a_1 X} + (1-\lambda)\mathrm{e}^{a_1 c}\right] + \frac{1-\beta}{a_2} \log \mathbb{E}\left[\lambda \mathrm{e}^{a_2 X} + (1-\lambda)\mathrm{e}^{a_2 c}\right]. \tag{15}$$

Since (15) 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,

$$\frac{\beta(\mathbb{E}\left[\mathrm{e}^{a_1 X}\right] - \mathrm{e}^{a_1 c})}{a_1 \mathrm{e}^{a_1 c}} = -\frac{(1-\beta)(\mathbb{E}\left[\mathrm{e}^{a_2 X}\right] - \mathrm{e}^{a_2 c})}{a_2 \mathrm{e}^{a_2 c}}. \tag{16}$$

$$\frac{\beta(\mathbb{E}\left[\mathrm{e}^{a_1 X}\right] - \mathrm{e}^{a_1 c})}{a_1 \mathbb{E}\left[\mathrm{e}^{a_1 X}\right]} = -\frac{(1-\beta)(\mathbb{E}\left[\mathrm{e}^{a_2 X}\right] - \mathrm{e}^{a_2 c})}{a_2 \mathbb{E}\left[\mathrm{e}^{a_2 X}\right]}. \tag{17}$$

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 (16) by (17) 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 (16), 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. $\square$

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 6 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 9. *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 6, 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 (9), we have for $h(x) = \mathrm{e}^{-\frac{x^2}{2V}}$ that

$$\mathrm{e}^{\frac{y^2}{2V}} \int \sigma(x) h(y-x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V} \cdot x} \cdot \mathrm{e}^{-\frac{x^2}{2V}}\,\mathrm{d}x. \tag{18}$$

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 (18) 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 (18) 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,

$$\int_{-\infty}^{\infty} \sigma(x) \cdot \mathrm{e}^{ax} \cdot (1 - \mathrm{e}^{-\frac{x^2}{2V}}) \,\mathrm{d}x < \varepsilon \quad \text{for all } |a| \leq A. \tag{19}$$

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 (19) holds for sufficiently large V. This completes the proof. □

D.2 Step 2: A Perturbation Argument

With Lemma 9, 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 10. *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 11. *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 12. *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 13. *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 9. 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. □

Lemma 14. *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. $\square$

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 15. *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 13, 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 $K_{X|Z}(a) \le K_{Y|Z}(a) + \frac{m}{n}$ is equivalent to $K_a(X) \le K_a(Y) + \frac{m}{n}K_a(Z)$ and equivalent to $K_a(X^{*n}) \le K_a(Y^{*n} + Z^{*m})$, 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 13 we also have $\Phi(X^{*n}) \le \Phi(Y^{*n} + Z^{*m})$, and thus $\Phi(X) \le \Phi(Y) + \frac{m}{n}\Phi(Z)$ by additivity. An approximation argument shows that for any *real* number $\varepsilon > 0$, $K_{X|Z}(a) \le K_{Y|Z}(a) + \varepsilon$ for all a implies $\Phi(X) \le \Phi(Y) + \varepsilon\Phi(Z)$. 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 $\frac{\mathrm{d}\hat{\mu}_Z}{\mathrm{d}\mu_Z}(a) = \frac{\Phi(Z)}{K_a(Z)} \le \Phi(Z)$, 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 16. *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 $X \ge 0$,, since we can add any constant to X. Given the previous lemma, we just need to construct $Z \ge 1$ 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 $a \ge 0$. 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

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

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 17. *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 Omitted Proofs for Section 4

E.1 Proof of Proposition 5

The result can be derived as a corollary of Proposition 6 which we prove below, 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 $K_a(X) \ge \min[X] = 0$ for every $a \in \overline{\mathbb{R}}$. Moreover, for $a \ge \varepsilon$ we have

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

whenever n is sufficient large. Thus

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

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.

E.2 Proof of Proposition 6

We first show that conditions (i) and (ii) are necessary for $\int_{\overline{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu_1(a) \le \int_{\overline{\mathbb{R}}} K_a(Y)\,\mathrm{d}\mu_2(a)$ 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 (11). 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

$$\int_{\overline{\mathbb{R}}_{\neq 0}} (K_a(X) - \mathbb{E}[X])\,\mathrm{d}\mu_1(a) \le \int_{\overline{\mathbb{R}}_{\neq 0}} (K_a(X) - \mathbb{E}[X])\,\mathrm{d}\mu_2(a). \tag{20}$$

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

$$\int_{(0,\infty]} (K_a(X) - \mathbb{E}[X])\,\mathrm{d}\mu_1(a) \le \int_{(0,\infty]} (K_a(X) - \mathbb{E}[X])\,\mathrm{d}\mu_2(a). \tag{21}$$

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 (20).

To prove (21), 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 (21) 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 (21) follows from condition (i) and $L_X''(b) \geq 0$ for all b. This completes the proof.

E.3 Proof of Theorem 5

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.

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 8. *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.[23]

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

[23]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.

$X \sim c$ implies $X_\lambda c \sim c$, which allows us to directly apply Lemma 7 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

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

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

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

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,

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

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 8. $\square$

F.1 Proof of Proposition 1

We now prove Proposition 1 as a corollary of Proposition 8. The first observation is that under time invariance, strong stochastic dynamic consistency is equivalent to the following property of the preference $\succeq$:

Axiom F.2 (Strong Stochastic Stationarity)**.** *For every pair of time lotteries (x, T), (y, S) and every $D \in L^\infty_+$ not necessarily independent, if $(x, T_d) \succeq (y, S_d)$ for almost every realization d of D, then $(x, T + D) \succeq (y, S + D)$.*

Indeed, suppose strong stochastic dynamic consistency is satisfied, and $(x, T_d) \succeq (y, S_d)$ holds for almost every realization d of D. Then by time invariance $(x, T_d) \succeq_{t+d} (y, S_d)$ also holds for almost every d. Strong stochastic dynamic consistency thus implies $(x, T + D) \succeq_t (y, S + D)$ and therefore strong stochastic stationarity. A similar argument shows that conversely, strong stochastic stationarity also implies strong stochastic dynamic consistency.

For the "only if" direction of Proposition 1, suppose that $\succeq$ is an MSTP that satisfies strong stochastic 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 stochastic 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 $(Y_\lambda Z) + 1$, so $(X_\lambda Z) + 1 \succeq_* (Y_\lambda Z) + 1$. 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 7. 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 8 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 stochastic stationarity. First consider $a = 0$, in which case the representation simplifies to $u(x) \cdot \mathrm{e}^{-\mathbb{E}[T]}$ with the normalization $r = 1$. If $(x, T_d) \succeq (y, S_d)$ for almost every d, then $u(x) \cdot \mathrm{e}^{-\mathbb{E}[T_d]} \geq u(y) \cdot \mathrm{e}^{-\mathbb{E}[S_d]}$, which can be rewritten as

$\mathbb{E}[S_d] - \mathbb{E}[T_d] \geq \log(u(y)/u(x))$. Averaging across different realizations d, this implies $\mathbb{E}[S] - \mathbb{E}[T] \geq \log(u(y)/u(x))$, and thus $\mathbb{E}[S+D] - \mathbb{E}[T+D] \geq \log(u(y)/u(x))$. After rearranging, this 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 stochastic 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

$$u(x) \cdot \mathbb{E}\left[\mathrm{e}^{a(S_d+d)}\right] \geq u(y) \cdot \mathbb{E}\left[\mathrm{e}^{a(T_d+d)}\right].$$

Averaging across different realizations 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 $r = 1$. In this case $(x, T_d) \succeq (y, S_d)$ implies $\max[S_d] - \max[T_d] \geq \log(u(y)/u(x))$, and thus

$$\max[S_d + d] - \max[T_d + d] \geq \log(u(y)/u(x)).$$

Let $\alpha = \max[S+D]$ and $c = \log(u(y)/u(x))$ be constants. Then the above implies that for almost every realization d 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.