Technical Appendix: Hierarchical Evolutionary Preferences Explain Discrepancies in Expected Utility Theory

Michael Holton Price*[1] and James Holland Jones[2,3,4]

[1]Department of Anthropology, Pennsylvania State University,
[2]Department of Earth Systems Science, Stanford University,
[3]Woods Institute for the Environment, Stanford University
[4]Department of Life Sciences, Imperial College London

October 17, 2016

1 Example 1: Generalizing the Arrow-Pratt Risk Premium

We first introduce age-structured life history theory in the absence of stochastic environmental uncertainty in life history traits. We then provide a basic overview of expected utility theory and derive the Arrow-Pratt risk premium. Having introduced the key evolutionary and economic concepts, we link them and derive a generalization of the Arrow-Pratt risk premium. Example 1 requires only non-stochastic life history theory, but Example 2 utilizes stochastic life history theory. Hence, the theoretical material in this section provides an accessible introduction to the more sophisticated material for Example 2.

1.1 Evolutionary theory

Consider an age structured population where each age class j accounts for individuals between $y_j = j\,\Delta y$ and $y_{j+1} = (j+1)\,\Delta y$ years of age. The key demographic traits that govern the population's dynamics are the age specific fertility, F_j, and the age specific survival, P_j. F_j is the mean number of offspring that an individual in age class j contributes to the juvenile age class ($j = 1$) going from time t_n to time $t_{n+1} = t_n + \Delta y$, where n indexes time steps. P_j is the probability that an individual will survive from age class j to age class $j+1$. The Leslie matrix, $\mathbf{A}$, has the age specific fertilities on the first row and the age specific survivals on the

*Correspondence Address: Department of Anthropology, 409 Carpenter Building, University Park, PA 16802; phone: 814-865-2509, fax: 814-863-1474; email: mhp12@psu.edu

sub-diagonal,

$$\mathbf{A} = \begin{pmatrix} F_1 & F_2 & F_3 & F_4 & \cdots & F_M \\ P_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & P_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & P_3 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & \cdots & 0 \\ 0 & 0 & 0 & \ddots & P_{M-1} & 0 \end{pmatrix}, \tag{1}$$

where M is the number of age classes. The Leslie matrix projects the population vector $\mathbf{z}_n$ one time step into the future, $\mathbf{z}_{n+1} = \mathbf{A}\mathbf{z}_n$. We assume that the Leslie matrix is irreducible and primitive, so that $\mathbf{A}^n$ converges to a stable age distribution for large enough n. Given these assumptions, there exists a unique dominant eigenvalue λ that is the per-period growth factor of the population. In the absence of environmental stochasticity in the Leslie matrix elements, λ is the measure of fitness that natural selection, the principal, maximizes (i.e., the top level of Figure 1). Associated with λ are the dominant left and right eigenvectors of $\mathbf{A}$, which we represent by $\boldsymbol{\nu}$ and $\boldsymbol{\omega}$, respectively. $\boldsymbol{\nu}$ is the vector of age specific reproductive values and $\boldsymbol{\omega}$ is the stable age distribution vector. For reference, Figures 1 and 2 plot the reproductive value and stable age distribution using demographic data for Venezuela in 1965 (low mortality, high fertility) and for Madagascar in 1966 (high mortality, high fertility) from (1). As is conventional for human demographic data, age classes are broken into 5 year increments. The peak in reproductive value in the fourth age class (15 to 20 years) is typical of human populations.

The elements of the Leslie matrix, A_{ij}, are the determinants of fitness at the intermediate level of Figure 1. We assume that they depend on a valuable but limited resource that can be traded off between fertility and survival across an individual's life cycle. Stated formally, each Leslie matrix element A_{ij} is a function of the consumption x_{ij} allocated to it. x_{ij} is at the bottom level of Figure 1. We will use a first-order Taylor expansion to assess perturbative changes to the distribution of resources around the baseline optimal distribution $\mathbf{x}_0$. In expressing the Taylor expansion, it is convenient to adopt an indexing notation directly in terms of F_j and P_j rather than the 2D indexing in terms of A_{ij} (both indexing notations will prove useful below). The fertility in age class j is $F_j = A_{1j}$ whereas the survival in age class j is $P_j = A_{j+1,j}$. The corresponding consumption variables are $x_j^{(F)} = x_{1j}$ and $x_j^{(P)} = x_{j+1,j}$. Given this notation, the Taylor expansion for λ is

$$\lambda = \lambda_0 + \sum_{j=1}^{M} \frac{\partial \lambda}{\partial F_j} \frac{\partial F_j}{\partial x_j^{(F)}} dx_j^{(F)} + \sum_{j=1}^{M-1} \frac{\partial \lambda}{\partial P_j} \frac{\partial P_j}{\partial x_j^{(P)}} dx_j^{(P)}, \tag{2}$$

where it is implicit that the partial derivatives are calculated at the baseline distribution of resources $\mathbf{x}_0$ and corresponding baseline Leslie matrix $\mathbf{A}_0 = \mathbf{A}(\mathbf{x}_0)$. Caswell (2; 3) derives formulas for the first and second partial derivatives with respect to an element of the Leslie matrix, ∂A_{ij}. The first derivative is

$$\frac{\partial \lambda}{\partial A_{ij}} = \nu_i \omega_j. \tag{3}$$

Equation 3 assumes that $\boldsymbol{\nu}$ and $\boldsymbol{\omega}$ have been scaled so that their dot product is 1, $\boldsymbol{\nu} \cdot \boldsymbol{\omega} = 1$. The second derivative is given by a sum over the non-dominant eigenvalues,

$$\frac{\partial^2 \lambda}{\partial A_{ij}^2} = 2\, \nu_i\, \omega_j \sum_{k \neq 1} \frac{\nu_i^{(k)}\, \omega_j^{(k)}}{\left(\lambda - \lambda^{(k)}\right)}, \tag{4}$$

where $\lambda^{(k)}$ is the k-th eigenvalue, $\boldsymbol{\nu}^{(k)}$ is the k-th left eigenvector, $\boldsymbol{\omega}^{(k)}$ is the k-th right eigenvector, and we assume that all vectors have been scaled so that $\boldsymbol{\nu}^{(k)} \cdot \boldsymbol{\omega}^{(k)} = 1$. If the superscript is omitted, it is assumed to be (1), the dominant eigenvalue/eigenvector (e.g., $\lambda^{(1)} = \lambda$).

1.2 Economic theory

Expected Utility Theory (EUT) can be derived from four axioms concerning preferences over risky prospects: completeness, transitivity, continuity, and independence. Let

$$\mathbf{q} = (p_1, y_1; p_2, y_2; \cdots; p_N, y_N) \tag{5}$$

be a pure (non-compound) prospect defined over N mutually exclusive and exhaustive outcomes y_n, each occurring with probability p_n (notation and presentation largely follow (4)). Given this definition, the four axioms can be succinctly stated as follows.

Axiom 1.1. Completeness *For all prospects* $\mathbf{q}$ *and* $\mathbf{q}'$*, either* $\mathbf{q} \succeq \mathbf{q}'$ *or* $\mathbf{q}' \succeq \mathbf{q}$ *or both, where* $\succeq$ *indicates weak preference.*

Axiom 1.2. Transitivity *For all prospects* $\mathbf{q}$, $\mathbf{q}'$, *and* $\mathbf{q}''$*, if* $\mathbf{q} \succeq \mathbf{q}'$ *and* $\mathbf{q}' \succeq \mathbf{q}''$ *then it must be that* $\mathbf{q} \succeq \mathbf{q}''$.

Axiom 1.3. Continuity *For all prospects* $\mathbf{q}$, $\mathbf{q}'$, *and* $\mathbf{q}''$*, if* $\mathbf{q} \succeq \mathbf{q}'$ *and* $\mathbf{q}' \succeq \mathbf{q}''$ *then there exists some* p *such that* $(\mathbf{q}, p; \mathbf{q}'', 1-p) \approx \mathbf{q}'$*, where* $\approx$ *indicates indifference and* $(\mathbf{q}, p; \mathbf{q}'', 1-p)$ *is a compound lottery in which* $\mathbf{q}$ *occurs with probability* p *and* $\mathbf{q}''$ *with probability* $1-p$.

Axiom 1.4. Independence *For all prospects* $\mathbf{q}$, $\mathbf{q}'$, *and* $\mathbf{q}''$ *and probabilities* p*, if* $\mathbf{q} \succeq \mathbf{q}'$ *then* $(\mathbf{q}, p; \mathbf{q}'', 1-p) \succeq (\mathbf{q}', p; \mathbf{q}'', 1-p)$.

If the completeness, transitivity, and continuity axioms are satisfied, preferences over prospects can be represented by a real valued function $\mathcal{U}$ such that if $\mathcal{U}(\mathbf{q}) \geq \mathcal{U}(\mathbf{q}')$ then $\mathbf{q} \succeq \mathbf{q}'$. If, furthermore, the independence axiom is also satisfied, preferences can be represented by a weighted sum over outcomes (i.e., by their expected value),

$$U = \mathbb{E}\big[u(\mathbf{y}_n)\big] = \sum_n p_n\, u(\mathbf{y}_n), \tag{6}$$

where $u(\cdot)$ is a utility function defined only over outcomes, and not over probabilities.

An influential treatment of risk within the standard economic theory of utility is that due to Arrow and Pratt (5; 6). Consider two lotteries (or prospects) that an individual can choose between: a risk free one and a risky one. The former offers a payoff $\mu_{rf} = \langle x_{rf} \rangle$ with certainty, whereas the latter has a mean payoff $\mu = \langle x \rangle$, variance $\sigma^2 = \big\langle (x-\mu)^2 \big\rangle$, and third moment about the mean $\gamma^3 = \left\langle (x-\mu)^3 \right\rangle$. To simplify notation a subscript is used to mark the risk free gamble, but not the risky gamble. The risk premium is the difference in means of the two gambles when the mean utilities of the lotteries are set equal, $\pi = \mu - \mu_{rf}$. It is the amount of extra compensation a risk averse individual must receive to accept the risky gamble. We will derive an equation for π by making a third order Taylor expansion of the expectation utility for the two gambles about the center point x_0,

$$u(x) = u_0 + \frac{\partial u}{\partial x}\left[x - x_0\right] + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}\left[x - x_0\right]^2 + \frac{1}{6}\frac{\partial^3 u}{\partial x^3}\left[x - x_0\right]^3, \tag{7}$$

where the derivatives are all evaluated at x_0. The expectation of u in Equation 7 is

$$\langle u \rangle = u_0 + \frac{\partial u}{\partial x}\left[\langle x \rangle - x_0\right] + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}\left[\left\langle x^2 \right\rangle - 2\langle x \rangle x_0 + x_0^2\right] \\ + \frac{1}{6}\frac{\partial^3 u}{\partial x^3}\left[\left\langle x^3 \right\rangle - 3\left\langle x^2 \right\rangle x_0 + 3\langle x \rangle x_0^2 - x_0^3\right]. \tag{8}$$

Utilizing Equations 34 and 35 derived below, Equation 8 becomes

$$\langle u \rangle = u_0 + \frac{\partial u}{\partial x}\left[\mu - x_0\right] + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}\left[\sigma^2 + \mu^2 - 2\mu x_0 + x_0^2\right] \\ + \frac{1}{6}\frac{\partial^3 u}{\partial x^3}\left[\gamma^3 + \mu^3 - x_0^3 + 3(\sigma^2 - \mu x_0)(\mu - x_0)\right]. \tag{9}$$

Utilizing Equation 9 for both the risk free and risky gamble and setting the expectation of the two lotteries equal to each other, $\langle u \rangle = \langle u_{rf} \rangle$, yields

$$0 = \frac{\partial u}{\partial x}\pi + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}\left[\sigma^2 - \sigma_{rf}^2 + \mu^2 - \mu_{rf}^2 - 2x_0\pi\right] \\ + \frac{1}{6}\frac{\partial^3 u}{\partial x^3}\left[\gamma^3 - \gamma_{rf}^3 + \mu^3 - \mu_{rf}^3 + 3\mu\sigma^2 - 3\mu_{rf}\sigma_{rf}^2 - 3x_0(\sigma^2 - \sigma_{rf}^2 + \mu^2 - \mu_{rf}^2) + 3x_0^2\pi\right]. \tag{10}$$

Since $\mu_{rf} = \mu - \pi$, the following identities hold:

$$\mu^2 - \mu_{rf}^2 = 2\mu\pi - \pi^2 \tag{11}$$

and

$$\mu^3 - \mu_{rf}^3 = 3\mu^2\pi - 3\mu\pi^2 + \pi^3. \tag{12}$$

Substituting Equation 11 and 12 in Equation 10 and utilizing the fact that $\sigma_{rf} = 0$ and $\gamma_{rf} = 0$ yields

$$0 = \frac{\partial u}{\partial x}\pi + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}\left[\sigma^2 - 2(\mu - x_0)\pi - \pi^2\right] \\ + \frac{1}{6}\frac{\partial^3 u}{\partial x^3}\left[\gamma^3 + 3(\mu - x_0)\sigma^2 + 3(\mu - x_0)^2\pi + 3(\mu - x_0)\pi^2 + \pi^3\right]. \tag{13}$$

Since it is the risk sensitivity to the gamble that is under consideration, a sensible value for the center of expansion is the mean of the risky lottery. Setting $x_0 = \mu$ in Equation 13 yields

$$0 = \frac{\partial u}{\partial x}\pi + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}\left[\sigma^2 - \pi^2\right] + \frac{1}{6}\frac{\partial^3 u}{\partial x^3}\left[\gamma^3 + \pi^3\right]. \tag{14}$$

Keeping only first order terms in π and setting the third moment to zero gives the desired formula for the risk premium,

$$\pi = -\frac{1}{2}\frac{\frac{\partial^2 u}{\partial x^2}}{\frac{\partial u}{\partial x}}\sigma^2 = -\frac{1}{2}\frac{u''}{u'}\sigma^2 = \frac{1}{2}\alpha_u\sigma^2, \tag{15}$$

where $\alpha_u = -u''(x)/u'(x)$ is the classical Arrow-Pratt Index of Absolute Risk Aversion (5; 6).

1.3 Implementing the hierarchical principal-agent framework

In the previous two sections, we established the evolutionary and economic theory needed to derive a generalization of the Arrow-Pratt risk premium using the hierarchical principal-agent framework described in the main text. We assume that all life histories traits in the Leslie matrix are fixed aside from juvenile survival, P_1. We allow juvenile survival to vary because it is one of the most evolutionary salient components of the Leslie matrix (7), but qualitatively identical results are reached if any other element of the Leslie matrix is the one chosen to be varied. To link the evolutionary and economic theory, we assume that juvenile survival is the determinant of fitness to be associated with utility, $P_1(x_1^{(P)}) = u(x)$. For further justification of this crucial assumption we refer the reader to recent work on the evolution of preferences in which some determinant of fitness (usually survival, fertility, or number of offspring) is associated with utility (8; 9; 10; 11)[1]. The measure of fitness is λ, the dominant eigenvalue of the Leslie matrix and growth parameter of the stable age distribution (see Section 1.1), and the full hierarchical link (or, mathematically, compositional link) between the evolutionary and economic theory is the relation

$$f(u(x)) = \lambda(P_1(x_1^{(P)})). \tag{16}$$

To generalize the Arrow-Pratt risk premium, we consider gambles that leave fitness, f, unchanged, as opposed to leaving utility unchanged in the conventional result. Equation 15 remains the correct formula to use, but with u replaced by f. Furthermore, the chain rule must be used to account for the hierarchical dependence of fitness on consumption. In particular, $\partial f/\partial x = f'\,u'$ and $\partial^2 f/\partial x^2 = f''\,u'^2 + f'\,u''$ and the formula for the evolutionary risk premium is

$$\pi_f = -\frac{1}{2}\frac{\frac{\partial^2 f}{\partial x^2}}{\frac{\partial f}{\partial x}}\sigma^2 = -\frac{1}{2}\left[\frac{u''}{u'} + \frac{f''}{f'}u'\right]\sigma^2 = \frac{1}{2}\left[\alpha_u + \alpha_f\,u'\right]\sigma^2, \tag{17}$$

[1] Other assumptions concerning the relation between utility and different evolutionary quantities exist, but they have become increasingly rare since the publication of Alan Rogers' seminal 1994 article (8). A partial list of the relations assumed or derived include: utility is a function of fitness (12; 13); utility equals fitness (14; 15; 16; 17; 18); and utility equals some determinant of fitness (8; 9; 19; 20; 10; 11).

where $\alpha_f = -f''(u)/f'(u)$. α_f can be calculated using Equations 3 and 4,

$$\alpha_f = -\frac{1}{2\sum_{k\neq 1} \frac{\nu_p^{(k)}\,\omega_q^{(k)}}{\left(\lambda - \lambda^{(k)}\right)}}. \tag{18}$$

For juvenile survival, $i = 2$ and $j = 1$ ($P_1 = A_{21}$). To illustrate our arguments in the main text and create Figure 2, we utilize demographic data from Madagascar in 1966 from (1) and adopt a utility function with constant absolute risk aversion c, $u(x) = 1 - \exp(-c\,x)$. This is a common functional form to assume for the utility function in econometric analysis and, like survival, lies between 0 and 1. We set c so that at a consumption level of $x = 1$ the utility function equals the actual juvenile survival of the Madagascar data, $c = -\log(1 - P_1^{(MAD)})$. We then allow x to vary, which changes juvenile survival in the Leslie matrix (see Equation 16) and plot α_f as a function of x to create Figure 2. In Figure 2, the consumption level x is called the Resource Level. The stochastic evolutionary curve in Figure 2 utilizes Equation 23 below for the fitness in order to calculate α_f.

2 Example 2: Pessimism due to Environmental Uncertainty

We first extend the non-stochastic age-structured life history theory used in Example 1 to allow stochasticity in the Leslie matrix elements, then describe Rank Dependent Expected Utility Theory (RDEUT), which replaces the Independence axiom of EUT with the requirement that preference orderings cannot violate first order stochastic dominance. We then link the evolutionary and economic theory by associating utility with fertility (the determinant of fitness), and demonstrate that stochastic uncertainty in the determinant of fitness induces natural selection, the principal, to instill pessimistic probability weighting *sensu* RDEUT in the agents.

2.1 Evolutionary theory

We now generalize the preceding material to allow for stochastic uncertainty in the Leslie matrix elements. The key idea is that the Leslie matrix is not fixed from one time step to another, but instead depends on the changing state of the world. Following (21), consider an environmental sequence $\mathcal{E}$ with an associated sequence of Leslie matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$. Let

$${}^{n}\mathbf{Y}^{1} = \mathbf{A}_n\,\mathbf{A}_{n-1}\,\cdots\,\mathbf{A}_1 \tag{19}$$

represent the product matrix which governs population growth to time period n, and let λ_n, $\boldsymbol{\omega}_n$, and $\boldsymbol{\nu}_n$ represent, respectively, the dominant eigenvalue, corresponding right eigenvector, and corresponding left eigenvector of ${}^{n}\mathbf{Y}^{1}$. Let λ, $\boldsymbol{\omega}$, and $\boldsymbol{\nu}$ represent the corresponding quantities for the mean Leslie matrix. In mathematical jargon, this is abuse of notation since ${}^{1}\mathbf{Y}^{1}$ is not typically the same as the mean Leslie matrix and $\boldsymbol{\omega}$, etc., have already been defined with respect to the non-stochastic Leslie matrix. However, the meaning of symbols is always clear from context so the conceptual clarity gained by using the same symbol for similar objects justifies the abuse of notation. Without loss of generality, we assume that $\boldsymbol{\omega}_n$ and $\boldsymbol{\nu}_n$ are normalized to 1. Weak ergodicity guarantees a few key results. First, it guarantees that

$$\lim_{n\to\infty} {}^n\mathbf{Y}^1 \to \lambda_n \frac{\boldsymbol{\omega}_n \boldsymbol{\nu}'_n}{\boldsymbol{\nu}'_n \boldsymbol{\omega}_n}. \tag{20}$$

Second, it guarantees the convergence of age structure for any non-negative and non-zero starting vector $\mathbf{z}_0$. To formalize this, let

$$\mathbf{z}_{i+1} = \frac{\mathbf{A}_{i+1}\mathbf{z}_i}{|\mathbf{A}_{i+1}\mathbf{z}_i|} \tag{21}$$

be the normalized age structure at time period n. Then

$$\lim_{n\to\infty} \mathbf{z}_n \to \boldsymbol{\omega}_n \tag{22}$$

An analogous result holds for $\boldsymbol{\nu}'_n$ with left multiplication. The logarithmic growth rate is

$$a = \lim_{n\to\infty} \frac{1}{n} \log \left| {}^n\mathbf{Y}^1 \boldsymbol{\zeta} \right|, \tag{23}$$

where $\boldsymbol{\zeta}$ is an arbitrary initial age structure. (21) shows that a phenotype that maximizes this logarithmic growth rate cannot be invaded by a phenotype with a lower logarithmic a, and hence is following an evolutionarily stable strategy (ESS). Therefore, a is the relevant fitness measure that natural selection, the principal in the evolutionary principal-agent framework, should maximize given stochastic uncertainty in the Leslie matrix elements. Stochasticity lowers the long term growth rate compared to a non-stochastic reference Leslie matrix with the same mean life history traits. Effectively, this is because growth is a multiplicative process and the geometric mean is always less then the arithmetic mean (22). In addition to proving that a phenotype that maximizes a is following an ESS, (21) derives a useful "small noise approximation" for the logarithmic growth rate assuming relatively small fluctuations of the Leslie matrix elements relative to the mean values,

$$a = \log \lambda - \frac{1}{2\lambda^2 \left(\boldsymbol{\nu} \cdot \boldsymbol{\omega}\right)^2} \left(\boldsymbol{\nu} \otimes \boldsymbol{\nu}\right)^\dagger \mathbf{C} \left(\boldsymbol{\omega} \otimes \boldsymbol{\omega}\right), \tag{24}$$

where $\mathbf{C}$ is the covariance matrix of $\mathbf{A}$, $\otimes$ is the Kronecker product, and the serial auto-correlation term in Tuljapurkar's approximation is not included. If only one Leslie matrix element is considered, Equation 24 simplifies to

$$a = \log \lambda - \frac{1}{2\lambda^2} \left(\frac{\partial \lambda}{\partial A_{ij}}\right)^2 \phi_{ij}^2, \tag{25}$$

where ϕ_{ij}^2 is the environmental variance associated with matrix element A_{ij} and Equation 3 provides a formula for $\frac{\partial \lambda}{\partial A_{ij}}$ in terms of the mean Leslie matrix. This is the key result for the evolutionary component of Example 2. It will prove useful to explicitly express the functional dependence of the log growth rate as $a(\langle \mathbf{A} \rangle, \phi_{ij})$, where we explicitly use $\langle \mathbf{A} \rangle$ for the mean Leslie matrix.

2.2 Economic theory

In this section we motivate RDEUT as a generalization of EUT with a discussion of the Allais paradox, then discuss RDEUT's axiomatic foundation and mathematical formulation. The Allais paradox, or common consequence effect, is perhaps the most notable empirical violation of EUT. For example, Quiggin (23) writes: "The Allais problem is the *pons asinorum* of theories of choice under uncertainty. Almost all of the many authors who have introduced new models of choice under uncertainty in the last ten years [i.e., alternatives to expected utility theory] have included a demonstration that the model is consistent with the behavior revealed in this problem." The Allais paradox calls into question one of the four foundational axioms of EUT, the independence axiom, which asserts that adding a common outcome to two prospects should not change the preference ordering of the prospects.[2] To illustrate this, consider the following two choices or problems, which are also summarized in Table 1.

Problem 1 involves a choice between between prospect 1A, which yields \$1,000,000 with probability 11% and nothing otherwise, and prospect 1B, which yields \$5,000,000 with probability 10% and nothing otherwise. Problem 2 involves a choice between between prospect 2A, which yields \$1,000,000 with certainty, and prospect 2B, which yields \$1,000,000 with probability 89%, \$5,000,000 with probability 10%, and nothing otherwise (1% of the time). Table 2 shows that Problem 1 and Problem 2 just differ by a common outcome since Problem 1 can be transformed into Problem 2 by moving an 89% probability of receiving nothing to an 89% probability of receiving \$1,000,000 for each prospect (A and B). Because EUT assumes the independence axiom, it predicts that individuals who prefer 1A to 1B should also prefer 2A to 2B (and vice versa). However, when choice problems such as that encapsulated by Table 1 are presented to individuals in economic choice experiments subjects often exhibit a preference reversal. For example, (24) found, using different numbers than those in Table 1 but a similar set-up, that 82% of the subjects preferred the certain outcome (corresponding to 2A in my notation) to its alternate, but 83% preferred the equivalent of 1B to 1A, which implies that the majority of individuals violated the independence axiom.

2.2.1 Subjective probability weighting and generalizations of EUT

Subjective probability weighting has been suggested as one way to accommodate the Allais paradox. In the simplest formulation of probability weighting[3], the weights which enter the cumulative utility sum for probabilistic outcomes are a direct function of the probability of each individual outcome,

$$\mathcal{U} = \sum \mathcal{W}(p_i) u(x_i), \tag{26}$$

where $\mathcal{W}$ is the weighting function and p_i is the probability of outcome i. Note that $\mathcal{U}$ is used instead of U in Equation 26 so U can be reserved for expected utility sums (which Equation 26 is not, despite its nearly identical form). This straightforward method of probability weighting was adopted in the original formulation of prospect theory (PT (24)), one of the first generalizations

[2] If expected utility is assumed, these outcomes literally add in the expected utility calculation.

[3] (25) may have been the first to suggest the use of subjective probability weights. They showed that when individuals could competitively bid on lotteries with uncertain outcomes they systematically overweighted low probabilities (up to about $p = .1$) and underweighted high probabilities.

of EUT to gain widespread attention. However, it was recognized that PT violated first order stochastic dominance, which occurs if one gamble is preferred to another, even though the second gamble is strictly better than the first in the sense that the second gamble's cumulative density function is at least equal to that of the first over its entire domain, and is strictly greater over at least part of the domain. A violation of stochastic dominance was therefore avoided in PT through an editing process that occurred before the probability weighting. The appeal of PT was its selective incorporation of the more desirable features of EUT, while nevertheless rejecting the independence axiom, so that the Allais paradox and other behavioral violations could be accommodated.

Following the formulation of PT, a number of scholars began to formulate theories that adopted the three less controversial axioms of EUT (completeness, transitivity, and continuity) while simultaneously avoiding the violation of stochastic dominance exhibited by PT in its unedited form. The culmination of this work was the simultaneous and independent discovery of RDEUT by at least three scholars (26; 27; 28). RDEUT is the only axiomatic formulation of utility theory that simultaneously satisfies completeness, transitivity, and continuity while avoiding the violation of stochastic dominance. Violations of stochastic dominance can only be avoided if the subjective probability weighting depends in a quite specific manner on the entire vector of probabilities for a prospect. Assume that the probabilities of some prospect are sorted from most desirable to least desirable according to the values of their corresponding outcomes ($x_1 \geq x_2 \geq x_3 \geq \cdots$). Next, define the cumulative probability, or rank (29), as

$$\theta_i = \sum_{j=1}^{i} p_j, \tag{27}$$

where $\theta_0 = 0$. θ_i is the probability of receiving an outcome that is as good or better than i. The probability weighting on outcome i is given by

$$\rho_i = w(\theta_i) - w(\theta_{i-1}), \tag{28}$$

where w is the weighting function defined on ranks. Note that w is used rather than $\mathcal{W}$ to emphasize that w is defined on probability ranks, whereas $\mathcal{W}$ is defined on probabilities. w maps the interval $[0, 1]$ onto itself and must be an increasing function of θ. The rank dependent expected utility is

$$\mathcal{U} = \sum_i \rho_i u(x_i). \tag{29}$$

Figure 3 plots three different probability weighting functions. The middle curve is a standard form that exhibits two suggested biases in the psychophysics of probability distortions: likelihood sensitivity and pessimism (30; 29). Likelihood sensitivity occurs because differences near the endpoints of the probability scale (0 and 1) loom larger than differences in the interior of the interval. Conversely, differences in the middle of the scale loom smaller, so that the difference between 0% and 1% is psychologically much more salient than the difference between 60% and 61%. To account for diminished likelihood sensitivity for intermediate probabilities, the probability weighting function must have a sideways-Z shape so that the center is more flat than the endpoints.

Pessimism involves the under-weighting of desirable outcomes relative to their probability of occurrence. Two types of pessimism can be defined relative to the probability weighting function, $w(\theta)$: regular pessimism and strong pessimism (23). Regular pessimism (or simply pessimism) occurs where w is below the identity line $w = \theta$, whereas optimism occurs where w is above the identity line. Strong pessimism occurs where w is convex, whereas strong optimism occurs where w is concave. For many functional forms of w, regular pessimism and strong pessimism occur on the same or nearly the same intervals (and similarly for optimism).

To connect this to the Allais paradox, the reason RDEUT utility theory can accommodate the Allais paradox is pessimistic probability weighting. (31), for example, has formulated a general statement of the Allais paradox using arbitrary payoffs and non-discrete probability distributions, and shown that RDEUT with a probability weighting function w exhibiting strong pessimism on its entire domain (0 to 1) can accommodate any and all formulations of the generalized Allais paradox. However, only a subset of the generalized formulations are encountered in daily life or economic choice experiments, so less restrictive probability weighting functions can accommodate the empirically relevant cases, although some pessimism is still a necessary element. For example, as already discussed pessimism is one of the two primary components of the standard probability weighting curve (the other being likelihood sensitivity), which exhibits pessimism over only part of its domain. Consequently, we utilize regular pessimism as opposed to strong pessimism as the definition of pessimism used in the main text.

When it was realized that RDEUT avoided violations of stochastic dominance and could accommodate the Allais paradox, PT was reformulated to include RDEUT, becoming Cumulative Prospect Theory (CPT (32)). Consequently, RDEUT, either in its own right or as a component of CPT, is one of the most important components of contemporary work to generalize EUT (4; 29).

2.3 Implementing the hierarchical principal-agent framework

In this section, we link the evolutionary and economic theory described above to demonstrate how pessimistic probability weighting emerges as a solution to the evolutionary principal-agent problem due to stochastic uncertainty in the Leslie matrix elements. As with Example 1, the linking assumption is that one element of the Leslie matrix (a determinant of fitness) can be associated with utility and that fitness depends hierarchically on consumption towards that Leslie matrix element. In addition, variance in consumption due to environmental uncertainty influences fitness as it is a component of Equation 25. However, it is not necessary to explicitly model the link between consumption and utility (the bottom levels in Figure 2) since pessimism arises from the link between utility and fitness (the top levels in Figure 2). Rather, we assume that uncertainty in consumption leads to uncertainty in the Leslie matrix elements, and explicitly model the latter.

Given the framing in the preceding paragraph, it is straightforward to demonstrate that environmental uncertainty induces a new preference ordering equivalent to pessimism in RDEUT. Let $\mathbf{A}$ be the mean Leslie matrix and A_{ij} an element that is perturbed, with a variance due to environmental uncertainty of ϕ_{ij}^2. The measure of fitness that the principal, natural selection, maximizes is a in Equation 25. The effect of environmental uncertainty is to reduce a if the mean Leslie matrix is held constant. That is, $\phi_{ij} < \phi'_{ij}$ implies $a(\langle \mathbf{A} \rangle, \phi_{ij}) < a(\langle \mathbf{A} \rangle, \phi'_{ij})$, and vice versa.

To introduce the economic formalism, let s index states of the world (e.g., good years and bad years), which occur with probability p_s, and let k index strategies that the agent can choose, which determine the value of the (ij)-th Leslie matrix element given the state of the world, $A_{ij,s}^{(k)}$. The mean and standard deviation of strategy k are

$$A_{ij}^{(k)} = \sum_s p_s \, A_{ij,s}^{(k)} \tag{30}$$

and

$$\phi_{ij}^{(k)} = \sqrt{\sum_s p_s \, (A_{ij,s}^{(k)} - A_{ij}^{(k)})^2}. \tag{31}$$

An expected utility maximizer assesses strategy k solely by the mean value of the Leslie matrix element being perturbed, $A_{ij}^{(k)}$. In terms of the measure of fitness, a, this can be written $a^{(k,EUT)} = a(A_{ij}^{(k)}, 0)$. Assuming that $\sigma_{ij}^{(k)} > 0$, an EUT evaluates a as less than its true value,

$$a^{(k,EUT)} < a^{(k)}. \tag{32}$$

As discussed in the preceding section, a pessimistic RDEUT decision maker evaluates the value of a lottery as less than its EUT value,

$$\mathcal{U}^{(EUT)} < \mathcal{U}^{(PES)}. \tag{33}$$

Comparing Equation 32 with Equation 33, one concludes that a principal wishing to correct the bias of the EUT decision maker in Equation 32 via RDEUT probability weighting will instill pessimistic weighting preferences in the agent.

3 Identities of the moments of random variables

In this section, we establish some identities that are used in the derivations above related to the first, second, and third moments of random variables. Let x be some random variable, $\mu = \langle x \rangle$ its mean, $\sigma^2 = \left\langle (x-\mu)^2 \right\rangle$ its variance, and $\gamma^3 = \left\langle (x-\mu)^3 \right\rangle$ its third moment about the mean. The angular braces, $\langle \cdot \rangle$, indicate an expectation. The formula for the variance can be expanded to yield

$$\sigma^2 = \left\langle (x - \langle x \rangle)^2 \right\rangle = \left\langle x^2 - 2x\langle x \rangle + \langle x \rangle^2 \right\rangle = \left\langle x^2 \right\rangle - 2\langle x \rangle \langle x \rangle + \langle x \rangle^2 = \left\langle x^2 \right\rangle - \mu^2.$$

Solving this for $\left\langle x^2 \right\rangle$ gives the identity

$$\left\langle x^2 \right\rangle = \sigma^2 + \mu^2. \tag{34}$$

Similarly, the third order moment can be expanded to yield

$$\begin{aligned}\gamma^3 = \left\langle (x - \langle x \rangle)^3 \right\rangle &= \left\langle x^3 - 3x^2\langle x \rangle + 3x\langle x \rangle^2 + \langle x \rangle^3 \right\rangle \\ &= \left\langle x^3 \right\rangle - 3\left\langle x^2 \right\rangle \langle x \rangle + 3\langle x \rangle \langle x \rangle^2 - \langle x \rangle^3 = \left\langle x^2 \right\rangle - \mu^2.\end{aligned}$$

Utilizing Identity 34 to substitute for $\langle x^2 \rangle$ in this expansion and solving for $\langle x^3 \rangle$ gives the identity

$$\left\langle x^3 \right\rangle = \gamma^3 + 3\mu\sigma^2 + \mu^3. \tag{35}$$

Let $\Delta\mu = \mu_2 - \mu_1$ be the difference of two means. The difference of squares of the means is

$$\mu_2^2 - \mu_1^2 = \mu_2^2 - (\mu_2 - \Delta\mu)^2 = \mu_2^2 - (\mu_2^2 - 2\mu_2\Delta\mu + \Delta\mu^2) = 2\mu_2\Delta\mu - \Delta\mu^2,$$

which establishes the identity

$$\mu_2^2 - \mu_1^2 = 2\mu_2\Delta\mu - \Delta\mu^2. \tag{36}$$

Similarly, the difference of cubes of the means is

$$\begin{aligned}\mu_2^3 - \mu_1^3 = \mu_2^3 - (\mu_2 - \Delta\mu)^3 = \mu_2^3 - (\mu_2^3 - 3\mu_2^2\Delta\mu + 3\mu_2\Delta\mu^2 - \Delta\mu^3)\\ = 3\mu_2^2\Delta\mu - 3\mu_2\Delta\mu^2 + \Delta\mu^3,\end{aligned}$$

which establishes the identity

$$\mu_2^3 - \mu_1^3 = 3\mu_2^2\Delta\mu - 3\mu_2\Delta\mu^2 + \Delta\mu^3. \tag{37}$$

References

[1] Goodman LA, Keyfitz N, Pullum TW, 1974 Family Formation and the Frequency of Various Kinship Relationships. *Theoretical Population Biology* **5**, 1–27. doi: 10.1016/0040-5809(74)90049-5

[2] Caswell H, 1978 A general formula for the sensitivity of population growth rate to changes in life history parameters. *Theoretical Population Biology* **14**, 215–230. doi: 10.1016/0040-5809(78)90025-4

[3] Caswell H, 1996 Second derivatives of population growth rate: Calculation and applications. *Ecology* **77**, 870–879. doi: 10.2307/2265507

[4] Starmer C, 2000 Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature* **38**, 332–382. doi: 10.1257/jel.38.2.332

[5] Arrow KJ, 1965 *Aspects of the theory of risk-bearing.* Helsinki: Yrjö Jahnssonin Säätiö Lectures

[6] Pratt JW, 1964 Risk aversion in the small and in the large. *Econometrica* **32**, 122–136. doi: 10.2307/1912743

[7] Jones JH, 2009 The force of selection on the human life cycle. *Evolution and Human Behavior* **30**, 305–314. doi: 10.1016/j.evolhumbehav.2009.01.005

[8] Rogers AR, 1994 Evolution of time preference by natural selection. *American Economic Review* **84**, 460–481

[9] Robson AJ, 1996 A biological basis for expected and non-expected utility. *Journal of Economic Theory* **68**, 397–424. doi: 10.1006/jeth.1996.0023

[10] Samuelson L, Swinkels J, 2006 Information, evolution and utility. *Theoretical Economics* **1**, 119–142

[11] Robson AJ, Samuelson L, 2009 The evolution of time preference with aggregate uncertainty. *American Economic Review* **99**, 1925–1953. doi: 10.1257/aer.99.5.1925

[12] Becker GS, 1976 Altruism, egoism, and genetic fitness: Economics and sociobiology. *Journal of Economic Literature* **14**, 817–826

[13] Galor O, Moav O, 2002 Natural selection and the origin of economic growth. *The Quarterly Journal of Economics* **117**, 1133–1191. doi: 10.1162/003355302320935007

[14] Rubin PH, Ii CWP, 1979 An evolutionary model of taste for risk. *Economic Inquiry* **17**, 585–596. doi: 10.1111/j.1465-7295.1979.tb00549.x

[15] Rubin PH, Kau JB, Meeker EF, 1979 Forms of wealth and parent-offspring conflict. *Journal of Social and Biological Structures* **2**, 53–64. doi: 10.1016/0140-1750(79)90021-6

[16] Ursprung HW, 1988 Evolution and the economic approach to human behavior. *Journal of Social and Biological Structures* **11**, 257–279. doi: 10.1016/0140-1750(88)90017-6

[17] Hansson I, Stuart C, 1990 Malthusian selection of preferences. *The American Economic Review* **80**, 529–544

[18] Rayo L, Becker GS, 2007 Evolutionary Efficiency and Happiness. *Journal of Political Economy* **115**, 302–337. doi: 10.1086/516737

[19] Robson AJ, 2001 Why would Nature give individuals utility functions? *Journal of Political Economy* **109**, 900–914. doi: 10.1086/322083

[20] Robson AJ, 2001 The biological basis of economic behavior. *Journal of Economic Literature* **39**, 11–33. doi: 10.1257/jel.39.1.11

[21] Tuljapurkar S, 1982 Population dynamics in variable environments. III. evolutionary dynamics of r-selection. *Theoretical Population Biology* **21**, 141–165. doi: 10.1073/pnas.87.3.1139

[22] Lewontin RC, Cohen D, 1969 On population growth in a randomly varying environment. *Proceedings of the National Academy of Sciences of the United States of America* **62**, 1056–1060. doi: 10.1073/pnas.62.4.1056

[23] Quiggin J, 1993 *Generalized Expected Utility Theory: The Rank-Dependent Model*. Norwell, MS: Kluwer

[24] Kahneman D, Tversky A, 1979 Prospect Theory: An analysis of decision under risk. *Econometrica* **47**, 263–292. doi: 10.2307/1914185

[25] Preson MG, Baratta P, 1948 An experimental study of the auction-value of an uncertain outcome. *American Journal of Psychology* **61**, 183–193. doi: 10.2307/1416964

[26] Quiggin J, 1982 A theory of anticipated utility. *Journal of Economic Behavior and Organization* **3**, 323–343. doi: 10.1016/0167-2681(82)90008-7

[27] Schmeidler D, 1989 Subjective probability and expected utility without additivity. *Econometrica* **57**, 571–587. doi: 10.2307/1911053

[28] Yaari ME, 1987 The dual theory of choice under risk. *Econometrica* **55**, 95–115. doi: 10.2307/1911158

[29] Wakker PP, 2010 *Prospect Theory: For Risk and Ambiguity*. Cambridge: Cambridge University Press

[30] Gonzalez R, Wu G, 1999 On the shape of the probability weighting function. *Cognitive Psychology* **38**, 129–166. doi: 10.1006/cogp.1998.0710

[31] Segal U, 1987 Some remarks on Quiggin's anticipated utility. *Journal of Economic Behavior & Organization* **8**, 145–154. doi: 10.1016/0167-2681(87)90027-8

[32] Tversky A, Kahneman D, 1992 Advances in Prospect Theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* **5**, 297–323. doi: 10.1007/BF00122574

4 Figures and Tables for the Appendix

Prospect	$0	$1M	$5M
1A	89%	11%	
1B	90%		10%
2A		100%	
2B	1%	89%	10%

Table 1: The Allais choice problem. Problem 1 is the choice between 1A and 1B. Problem 2 is the choice between 2A and 2B..

Prospect	$0	$0	$1M	$1M	$5M	$5M
1A		89%	11%			
1B	1%	89%			10%	
2A			11%	89%		
2B	1%			89%	10%	

Table 2: Illustrating the violation of the independence axiom in the choice problem. The prospects are identical to those in Table 1, but some of the probabilities have been broken down to show the common components of Problem 1 and Problem 2. For example, for 1B the 90% probability of receiving nothing has been written as 1% and 89%, which sum to 90%. Problem 1 is identical to Problem 2 except that both prospects in Problem 2 have an additional 89% chance of receiving $1,000,000. Since Problem 1 and Problem 2 just differ by a common outcome, the independence axiom of EUT predicts that any individual who prefers 1A to 1B should also prefer 2A to 2B (and vice versa). Empirically, however, subjects often exhibit a preference reversal, which violates the independence axiom..

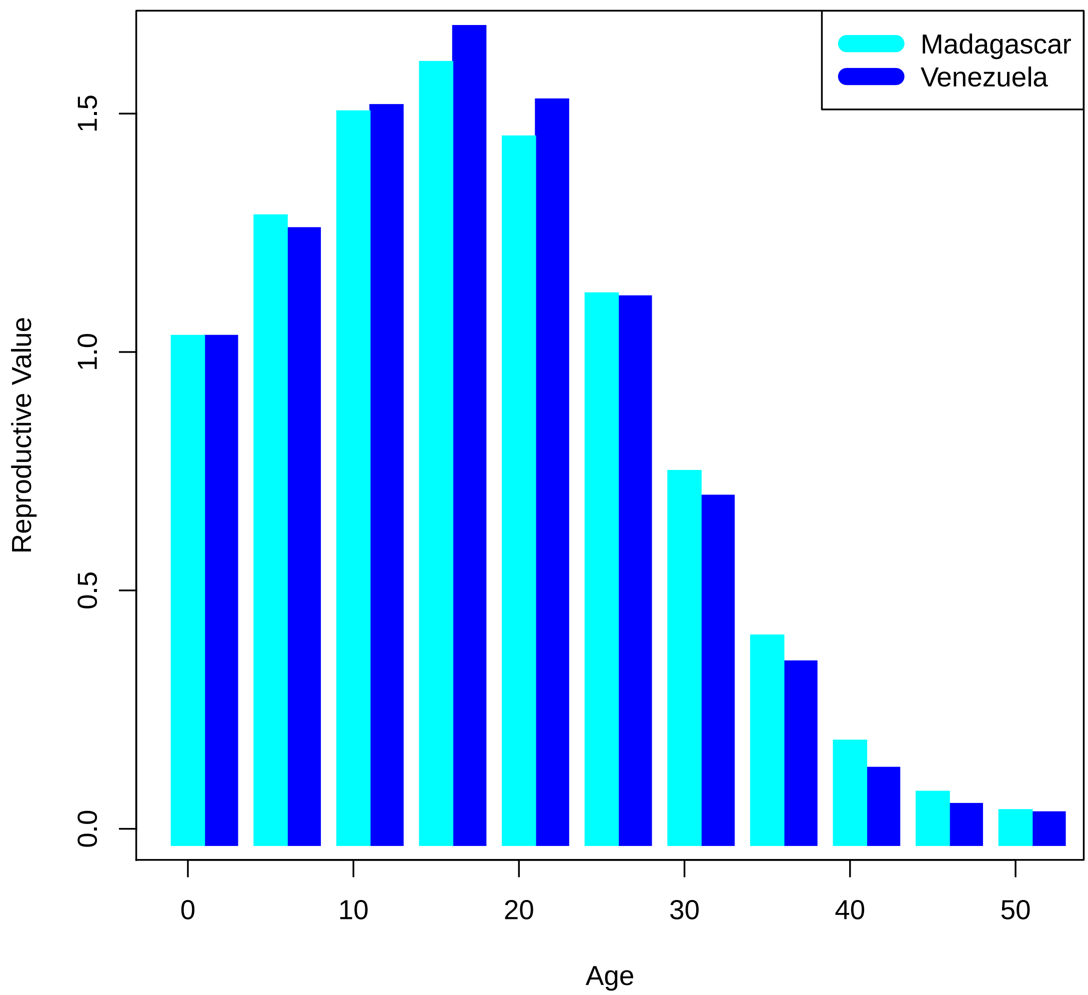

Figure 1: Reproductive value of the stable population for the Venezuela and Madagascar data sets, normalized so that the reproductive value of the first age class is 1.

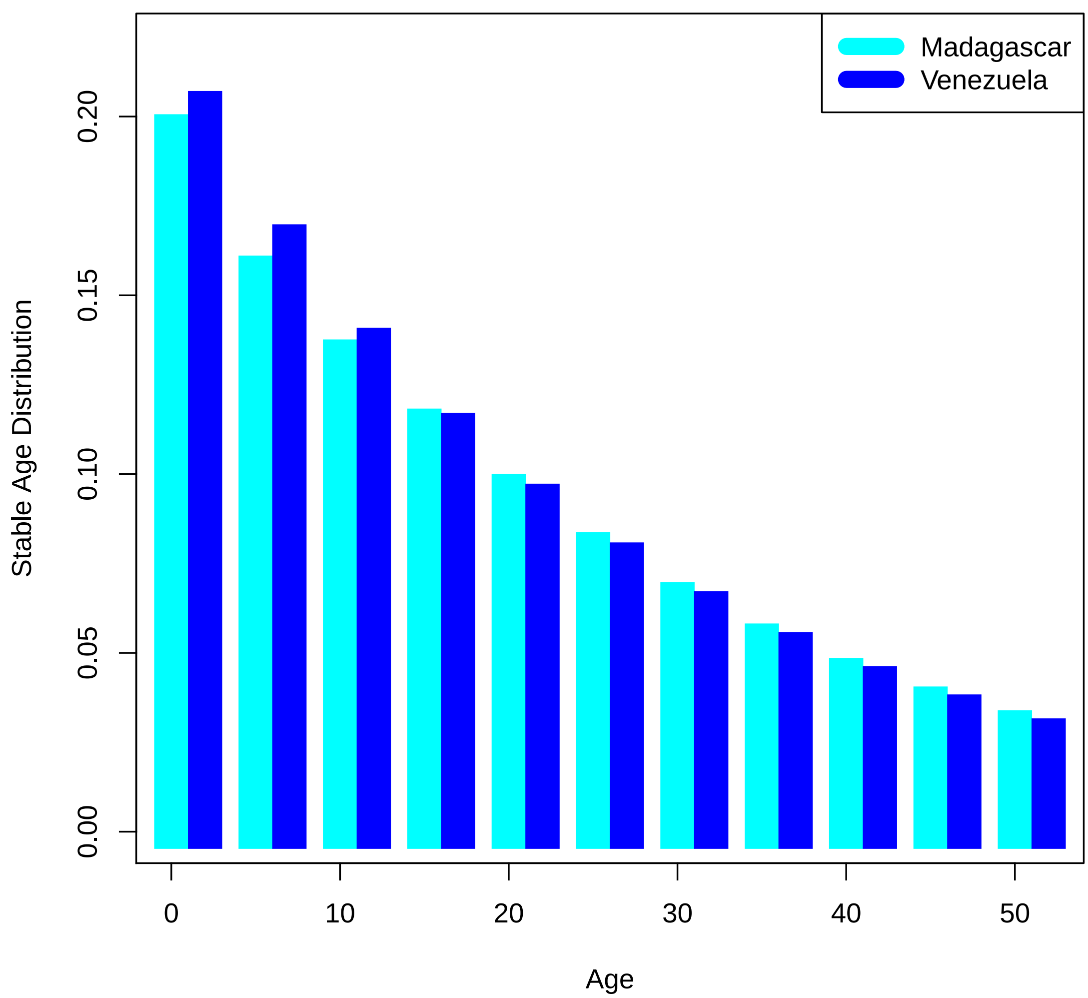

Figure 2: Stable age distribution of the stable population for the Venezuela and Madagascar data sets.

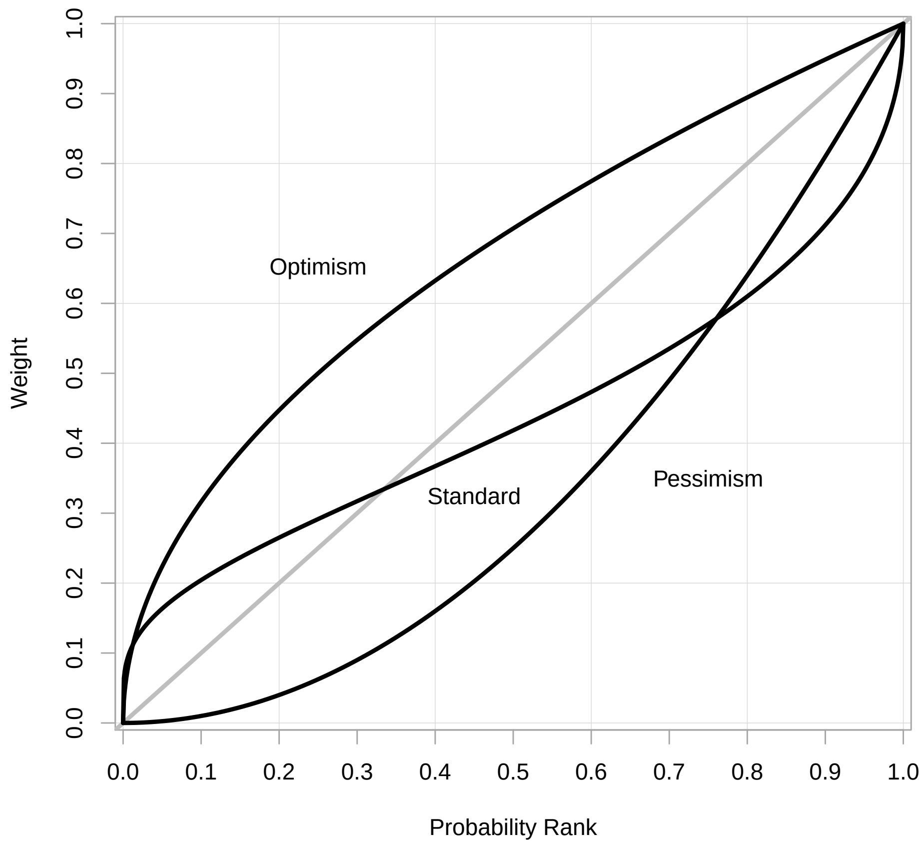

Figure 3: Probability weighting functions for the RDEUT model. The top curve exhibits optimism over its entire domain. The bottom curve exhibits pessimism. The middle curve is a standard curve adopted in much of the recent literature; it combines pessimism with likelihood sensitivity.