# Optimal behavior in a stochastic environment

Adam Narkiewicz[1]

July 13th, 2016

**Abstract:** This article proposes a general model of behavior in a stochastic environment in which an agent must sequentially choose among prospects in order to maximize her expected performance. A single utility function is optimal decision policy if the environment is stationary but is not optimal in general. The optimal utility function must be equal to a positive affine transformation of the expected performance of the agent. This theory challenges popular assumption that rational agents use a single utility function to make sequential choices. It implies that the value of utility function has the same units as performance and is, in principle, empirically measurable. It also shows how natural selection can create agents that use maximum expected utility principle. To illustrate the usefulness of the theory I build a model of foraging in a stochastic environment and identify conditions under which natural selection may lead to risk-averse preferences.

**Keywords:** stochastic environment, natural selection, Markov decision processes, risk aversion

**JEL Classification:** C61, D01

[1] Private Enterprise Research Center, Texas A&M University, College Station, TX 77840, U.S.A.; anarkiewicz@tamu.edu.

## 1. Introduction

The objective of this article is to propose a general model of behavior in a stochastic environment, to find sufficient conditions under which optimal decision policy is a single utility function, and to show how to calculate value of such utility function.

Theoretical models of decisions and corresponding empirical research often assume that agent uses a single utility function throughout a sequence of decisions. Examples include models concerned with sequential consumption or investment decisions or experiments trying to elicit risk preferences of an agent (human or animal) by presenting her with a sequence of gambling choices. In this paper I show that the assumption of a single utility function is not generally correct and the agent's utility function can change over time unless certain conditions are satisfied. For example, it is not optimal to use a single utility function of money for a sequence of gambling decisions if a person wishes to achieve a financial goal by a deadline.

To show that utility function can change over time, I construct a discrete-time model of behavior in a stochastic environment (MBSE). In every period, agent is provided with a set of prospects – probability distributions over environment's states. The agent selects one of these prospects and the environment changes its state accordingly. Environment's state determines agent's performance. Agent's fitness depends on how well she maximizes her expected performance. Natural selection ensures that only agents with the highest fitness survive. Selecting prospects in a best way possible is thus crucial for survival. The optimal decision policy follows a single utility function if the environment is stationary, that is, roughly speaking, current events depend only on the current state rather than the entire state history or the period number. For non-stationary environments it is easy to find examples where single utility function does not maximize fitness.

Fitness and utility are sometimes considered to be equivalent. Conversely, I consider them to be two distinct objects. My interpretation of the expected utility theory follows von Neumann and Morgenstern (1963) and Friedman and Savage (1948). Utility function is a function $u: X \to R$ where $X$ is a set of consumption bundles, a set of possible wealth or, in general, a set of states of the world. $u$ can be constructed from preferences over $X$ using representation theorems. When faced with two prospects (probability distributions over $X$) individual choses prospect $p$ over prospect $q$ only if $E_p u(x) \geq E_q u(x)$. Utility function does not depend on time and is superfluous if there is no risk involved, since in such case representation theorems cannot be applied.

Fitness is often defined as the ability of an organism to propagate its genes or, more precisely, the reproductive success of a genotype compared with other genotypes in a population (Pierce, 2012, p. 710). This general definition allows to explain many traits observed in animals, including social behaviors such as love between siblings or general altruism towards strangers (Buss, 2016, p. 227, 257). However, it is hard to construct an empirical measure based on this definition, and various other proxy measures of fitness are often used instead, for example expected number of offspring or probability of survival.

Natural selection ensures that individuals maximize fitness. Fitness in turn depends on the decision policy (e.g. utility function) used by an individual. This suggests a maximization problem in which objective function is fitness and the optimal utility function is a point in the space of all allowed decision policies for which the objective function is maximized. Rayo and Becker (2007) suggest that such setup can be interpreted as a principal-agent problem in which the principal (natural selection) designs the agent in such a way so that the agent using means available to her (utility function) maximizes principal’s objective (fitness). The optimization problem solved by natural selection is thus equivalent to the optimization problem an artificial intelligence engineer solves in order to design an optimal intelligent agent.

To help the reader understand the meaning of the terms used in MBSE, I start with an example application and show the general theory behind it afterwards. In Section 2, I present a series of stochastic models of foraging whose aim is to investigate what are the evolutionary reasons for risk preferences. In every period, an individual consumes certain amount of resources. She must also ensure that she has enough resources if she wants to successfully procreate. This forces her to accept at least some of foraging opportunities (gambles on her resources) the environment provides her with. Natural selection ensures that only those individuals who optimally select foraging opportunities survive. Hence, the problem is to find a utility function reflecting foraging preferences that maximize fitness.

I look for the optimal utility functions using numerical methods. A few of the environments I consider yield at least partially concave utility function and one of the environments yields an S-shaped utility function. This allows for identifying the possibility of death and the diminishing marginal returns to having more resources as potential evolutionary sources of traditional risk aversion. Moreover, the possibility of losing or gaining social status can potentially explain the risk preferences described by Prospect Theory (Kahneman and Tversky, 1979). Some of the utility functions are discontinuous and have flat areas, which makes the use of Arrow-Pratt measure of absolute risk aversion impractical. To analyze risk profiles of such utility functions, I propose fractional risk aversion, a measure equal to the percentage of potential gambles an individual would reject. Section 2 ends with sample foraging models in which a single utility function is not the optimal way to select favorable gambles.

In Section 3, I construct a mathematical model of behavior in a stochastic environment (MBSE) which generalizes concepts used in Section 2. It contains the main result of the paper: the Theorem proving that a single utility function can be used by an individual as the optimal way to make sequential decisions in a stationary environment. The Theorem also provides a way to calculate the value of utility function which must be equal to a positive affine transformation of the expected performance function. I use the Theorem in Proposition 1 and Proposition 2 to analytically verify the numerical results presented in Section 2. All proofs are in the Appendix.

Section 2 and Section 3 focus mostly on technical aspects. The discussion of related literature, terminology, assumptions, and implications of the results are combined in Section 4. I start that section by explaining how MBSE relates to the literature on Markov decision processes. I than further discuss the distinction between utility and fitness as well as the assumption that the environment is stochastic. The discussion of the main result follows: it is optimal to use a single utility function in a stationary environment but it may not be optimal to do so in an environment that violates stationarity. This challenges a common assumption that rational individuals use a single utility function over time. The Theorem also provides arguments that the theory that people (and other animals) behave as if they were maximizing expected utility is methodologically reducible. Finally, by linking value of utility function to the value of the performance function it suggests that utility functions may have particular units and can be in principle empirically measurable.

Section 4 also discusses the models presented in Section 2. These models are developed in the spirit of optimal foraging theory and are consistent with a number of experimental studies. They predict typical attitudes to risk (traditional risk aversion and S-shaped value function) as well as introduce new hypotheses. They explain why a rational individual would use a globally concave utility function and they suggest that individuals just below reference point may be more risk-loving than individuals far below it. In the summary, identified evolutionary reasons for risk aversion suggest that risk aversion may be an evolutionary mismatch.

## 2. Example applications

In this section I present a number of foraging models which are aimed at identifying evolutionary sources of risk preferences. The models are inspired by empirical research in humans and other animals. For a sample empirical analysis of the relationship between foraging, social status, and reproductive success in humans see Wiessner (2002). For a review of foraging experiments with animals see Real and Caraco (1986). The models allow me to derive various utility functions using numerical methods and to show that in some cases utility function used by an individual should change over time. Since some of the utility functions are discontinuous or flat, using Arrow-Pratt absolute risk aversion measure is impractical. To analyze risk profiles of such utility functions I use an ad-hoc measure I call fractional risk aversion.

Consider an individual inhabiting a stochastic environment with discrete time and periods numbered $t = 1,2,3, \ldots.$ In every period the individual consumes $d_t \sim \mathrm{Exp}\left(\frac{1}{\bar{c}}\right)$ resources where $\bar{c} = 0.3$. Also, in every period the individual is provided with a foraging opportunity. The opportunity yields $g_t$ net resources with probability $p_t$ if it is successful or costs $l_t$ resources with probability $1 - p_t$ if it is unsuccessful, where $p_t \sim \mathrm{U}[0,1]$, and

$g_t, l_t \sim \text{Exp}(1)$. The individual observes values of $d_t$, $p_t$, $g_t$, and $l_t$ and then chooses whether to accept or reject the foraging opportunity. The individual starts with $x_1 > 0$ resources. If she does not accept the foraging opportunity, then $x_{t+1} = x_t - d_t$. If she accepts the foraging opportunity, then $x_{t+1} = x_t - d_t + g_t$ with probability $p_t$ or $x_{t+1} = x_t - d_t - l_t$ with probability $1 - p_t$. If, after foraging decision has been made, the amount of resources available to the individual is below zero, then the individual dies without leaving progeny and her performance is measured at 0. Every period, before foraging choice is made, there is a chance of $p_0 = 0.2$ that the individual encounters a mate. If the individual encounters a mate before starving, she reproduces and her performance is measured at 1. All random variables and events are independently distributed.

An individual is equipped with a utility function that allows her to make decisions about which foraging opportunities to accept and which to reject, according to the maximum expected utility principle. Two individuals with different utility functions will presumably have different expected performance. Natural selection ensures that only individuals with the highest expected performance (probability of finding a mate) survive. The goal is thus to find a utility function that maximizes the expected performance. In other words, the goal is to maximize fitness, where fitness is defined as the expected performance and depends on the adopted utility function.

Let us call the above stochastic environment $E_1$. For the ease of exposition, let us explicitly characterize the performance function $f_1(x) = \mathbf{1}(x \geq 0)$, where $\mathbf{1}$ is the indicator function. Let us also explicitly characterize probability of process termination which is $\Phi_1(x) = \begin{cases} 1 & \Leftrightarrow & x < 0 \\ p_0 & \Leftrightarrow & x \geq 0 \end{cases}$ and embodies both possibilities of starving and finding a mate. Now, let us consider an alternative stochastic environment $E_2$ which is exactly the same as $E_1$ except that the individual can have a negative amount of resources without starving, that is $\Phi_2(x) = p_0$. An interpretation of $E_2$ is that the individual lives in a group and the state variable $x_t$ indicates how much resources she has relatively to the group average (or some other reference point). A potential mate choses our individual only if her social status is high and she has more resources than average. Otherwise, the opportunity to procreate is forgone.

| Env. | Performance function | Probability of termination |
|---|---|---|
| $E_1$ | $f_1(x) = \mathbf{1}(x \geq 0)$ | $\Phi_1(x) = \begin{cases} 1 & \Leftrightarrow \quad x < 0 \\ p_0 & \Leftrightarrow \quad x \geq 0 \end{cases}$ |
| $E_2$ | $f_2(x) = \mathbf{1}(x \geq 0)$ | $\Phi_2(x) = p_0$ |
| $E_3$ | $f_3(x) = x\mathbf{1}(x \geq 0)$ | $\Phi_3(x) = \begin{cases} 1 & \Leftrightarrow \quad x < 0 \\ p_0 & \Leftrightarrow \quad x \geq 0 \end{cases}$ |
| $E_4$ | $f_4(x) = x\mathbf{1}(x \geq 0)$ | $\Phi_4(x) = p_0$ |
| $E_5$ | $f_5(x) = \begin{cases} 0 & \Leftrightarrow \quad x < 0 \\ \ln(x+1) & \Leftrightarrow \quad x \geq 0 \end{cases}$ | $\Phi_5(x) = \begin{cases} 1 & \Leftrightarrow \quad x < 0 \\ p_0 & \Leftrightarrow \quad x \geq 0 \end{cases}$ |
| $E_6$ | $f_6(x) = \begin{cases} 0 & \Leftrightarrow \quad x < 0 \\ \ln(x+1) & \Leftrightarrow \quad x \geq 0 \end{cases}$ | $\Phi_6(x) = p_0$ |
| $E_7$ | $f_7(x) = x$ | $\Phi_7(x) = p_0$ |

**Table I.** Characteristics of stochastic environments under consideration.

Environment $E_3$ is the same as $E_1$ except that performance function is $f_3(x) = x\mathbf{1}(x \geq 0)$: the number of children depends linearly on the amount of resources individual was able to acquire. Environment $E_5$ is the same as $E_1$ except that performance function is $f_5(x) = \begin{cases} 0 & \Leftrightarrow \quad x < 0 \\ \ln(x+1) & \Leftrightarrow \quad x \geq 0 \end{cases}$, to reflect diminishing returns to having more resources in terms of ability to procreate. For $i = 2, 4, 6$ the environment $E_i$ is exactly the same as $E_{i-1}$ except that in $E_i$ we have $\Phi_i(x) = p_0$ (no possibility of starvation). Finally, the environment $E_7$ has $f_7(x) = x$ and $\Phi_7(x) = p_0$. The basic characteristics of all environments are summarized in Table I.

Furthermore, for each of the seven environments, consider four variants:

I. $d_t \sim \text{Exp}\left(\frac{1}{\bar{c}}\right)$ and $\bar{c} = 0.3$, $p_0 = 0.2$, $g_t, l_t \sim \text{Exp}(1)$,
II. $d_t \sim \text{Exp}\left(\frac{1}{\bar{c}}\right)$ and $\bar{c} = 0.4$, $p_0 = 0.2$, $g_t, l_t \sim \text{Exp}(1)$,
III. $d_t \sim \text{Exp}\left(\frac{1}{\bar{c}}\right)$ and $\bar{c} = 0.3$, $p_0 = 0.15$, $g_t, l_t \sim \text{Exp}(1)$,
IV. $d_t \sim \text{U}[0,2\bar{\text{c}}]$ and $\bar{c} = 0.3$, $p_0 = 0.2$, $g_t, l_t \sim \text{U}[0,2]$.

These four variants are supposed to provide a rudimentary robustness analysis. The only justification for the parameters and probability distributions I use is that they are easy to analyze, both numerically and analytically. It is thus important to see whether the results change much when the parameters and the distributions change.

For each of the 28 cases described above, I numerically look for an optimal utility function. Using Microsoft Visual C++ 2015, I implement the simultaneous perturbations

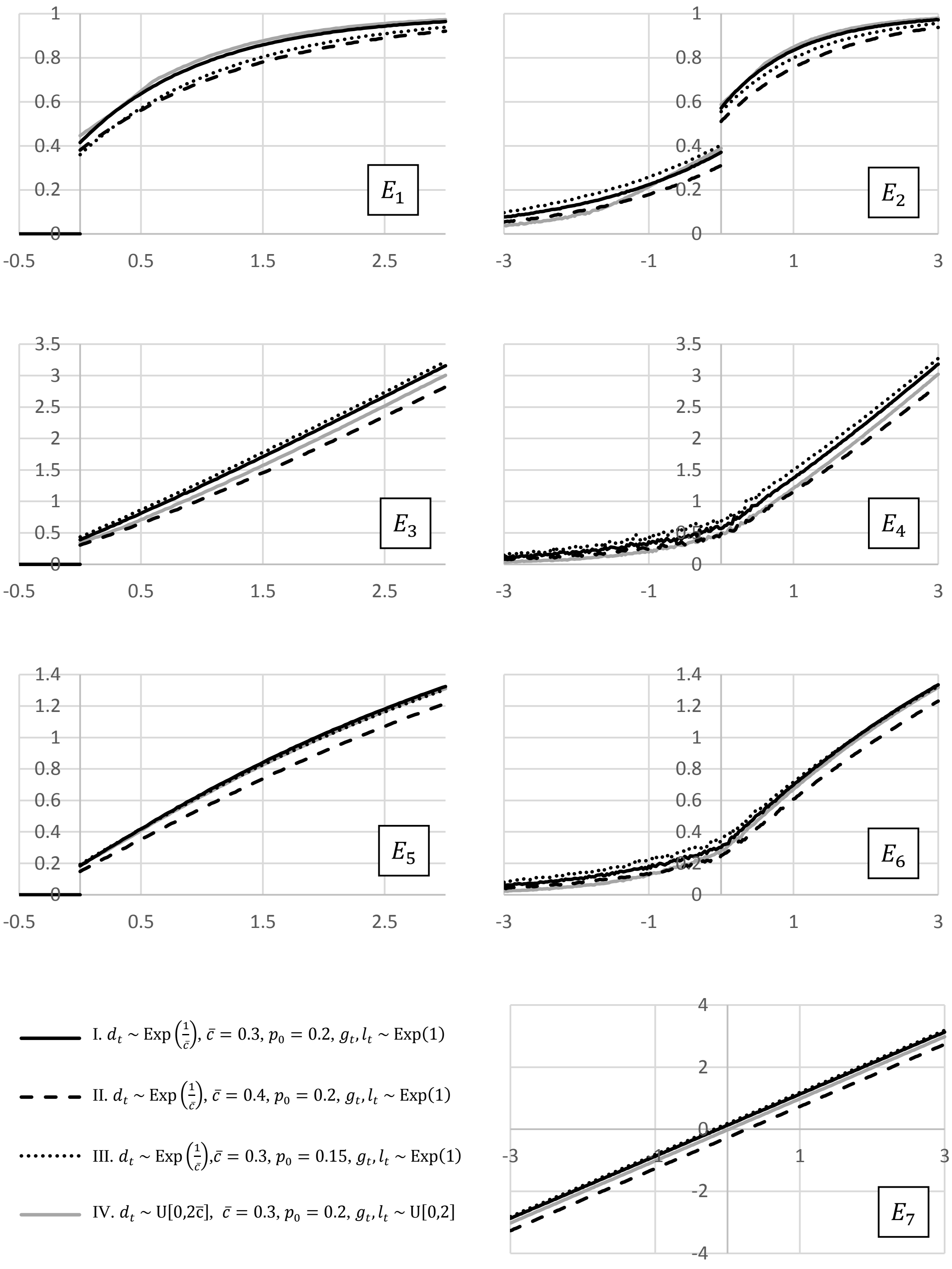

**Figure 1.** Numerical utility functions in the seven stochastic environments $E_1 - E_7$.

algorithm (Spall, 2003) and use a piecewise linear approximations of utility functions. In a nutshell, the algorithm takes a randomly generated utility function and simulates a life of an agent using this utility function in order to obtain her performance. Then, the algorithm tweaks the utility function in a random way and simulates individual's life again to see how the tweak affects the performance. If the tweak is improving, it is kept, otherwise, an opposite tweak is implemented. Then, the algorithm moves to the next iteration and tries another tweak. These steps are repeated until convergence is likely. The results are presented in Figure 1. Notice that the optimal utility function for all variants of the environment $E_7$ is linear. As the agents in this environment are supposed to maximize the expected amount of resources, risk neutrality is exactly what we expect.

A number of utility functions are discontinuous. In environments $E_1$, $E_3$, and $E_5$ the utility function equals zero for $x < 0$ and jumps to a positive value at $x = 0$. I attribute these discontinuities to the discontinuity of $\Phi$: for small $\varepsilon > 0$ the difference between $x = \varepsilon$ and $x = -\varepsilon$ is the difference between life and death. In the case of $E_1$ and $E_2$ the source of discontinuity in the utility function is the discontinuity of $f$: gambles that lend us just above zero are highly desirable in comparison to gambles that lend us just below zero, even if there is no starvation possible.

The easiest way to investigate risk preferences for a given utility function is to visually inspect it for local concavities and convexities. This simple method allows us to identify three sources of risk preferences. The possibility of starvation ($E_1$) and the diminishing returns to having more resources ($E_5$ and $E_6$) result in local concavity for $x \geq 0$ and can be interpreted as sources of traditional risk aversion. The possibility of losing or gaining social status ($E_2$) results in S-shaped, reference-dependent utility function.

Visually inspecting utility function for their local curvature disregards important features of these utility functions that may affect which gambles are rejected and which are accepted. Discontinuities and flat areas of a utility function are certain to affect risk aversion, at least in their neighborhood. The investigation of risk preferences embodied in such utility functions requires thus more careful analysis. In addition to having these pernicious features, environments $E_1$ and $E_2$ are especially interesting due to their relationship with the literature. $E_1$ resembles a typical model from optimal foraging theory and $E_2$ seems to provide evolutionary reasons for the S-shaped value function from the Prospect Theory.

The traditional Arrow-Pratt measure of absolute risk aversion $ARA(x) = -u''(x)/u'(x)$ is in this context impractical. First, it is undefined for flat areas of utility function. Second, it disregards points of discontinuity even though they surely affect which gambles are accepted. Finally, even an attempt to calculate it for strictly increasing and continuous parts of utility functions is plagued with difficulties. Since it is obtained using stochastic optimization, the estimated value of utility function $\hat{u}(x)$ equals the true value plus a random error: $\hat{u}(x) = u(x) + \varepsilon$. Using first differences to calculate first-order derivatives is thus prone to large errors, which become even larger when the procedure is repeated to obtain second-order derivatives. Alternatively, one could approximate pieces of utility functions by fitting the data with some parametrized function and then calculate derivatives of that function. This approach however also turns out to be unreliable. To see why, consider any constant absolute risk aversion (CARA) utility function of the form $u(x) = a + be^{cx}$. Trying to approximate a piece of this function with a polynomial results in Arrow-Pratt absolute risk aversion being a rational function, which is guaranteed not to be constant.

Instead of absolute risk aversion, I use an ad-hoc measure which I call fractional risk aversion (FRA), defined as a probability that a randomly drawn gamble is rejected in favor of the status quo. Such measure requires population of the gambles to be known in advance. This criterion is met by the environments $E_1 - E_7$ and the population of gambles is the same in a given variant of all of the environments. For the variant I of each environment we can use the following formula:

$$FRA(x) = \int_0^{+\infty} \int_0^{+\infty} \int_0^{1} \frac{1}{2}\Big(1 + \operatorname{sgn}\big(u(x) - pu(x+g) - (1-p)u(x-l)\big)\Big)\, dp \frac{dl}{e^l} \frac{dg}{e^g}$$

where $\operatorname{sgn}(\cdot)$ is a standard sign function and $u(x)$ is the utility function. Note that for all $x$, $0 \leq FRA(x) \leq 1$. $FRA(x) = 0.5$ for risk neutral and indifferent individuals (with a linear and flat utility function respectively). Finally, FRA is constant if the utility function is CARA.

Value of FRA depends directly on the value of a utility function. Unlike with ARA, with FRA it is practical to use a parametrized functional form and fit it with numerically obtained approximations of the utility function. I use exponential functions to fit utility function of both $E_1$ and $E_2$. The goodness of fit and additional experiments with $E_5$ suggest the following conjecture: for the variant I, the optimal utility functions in $E_1$ and $E_2$ are piecewise of the functional form $a + be^{cx}$ and the utility function in $E_5$ is piecewise of the functional form $a + be^{cx} + dx$.

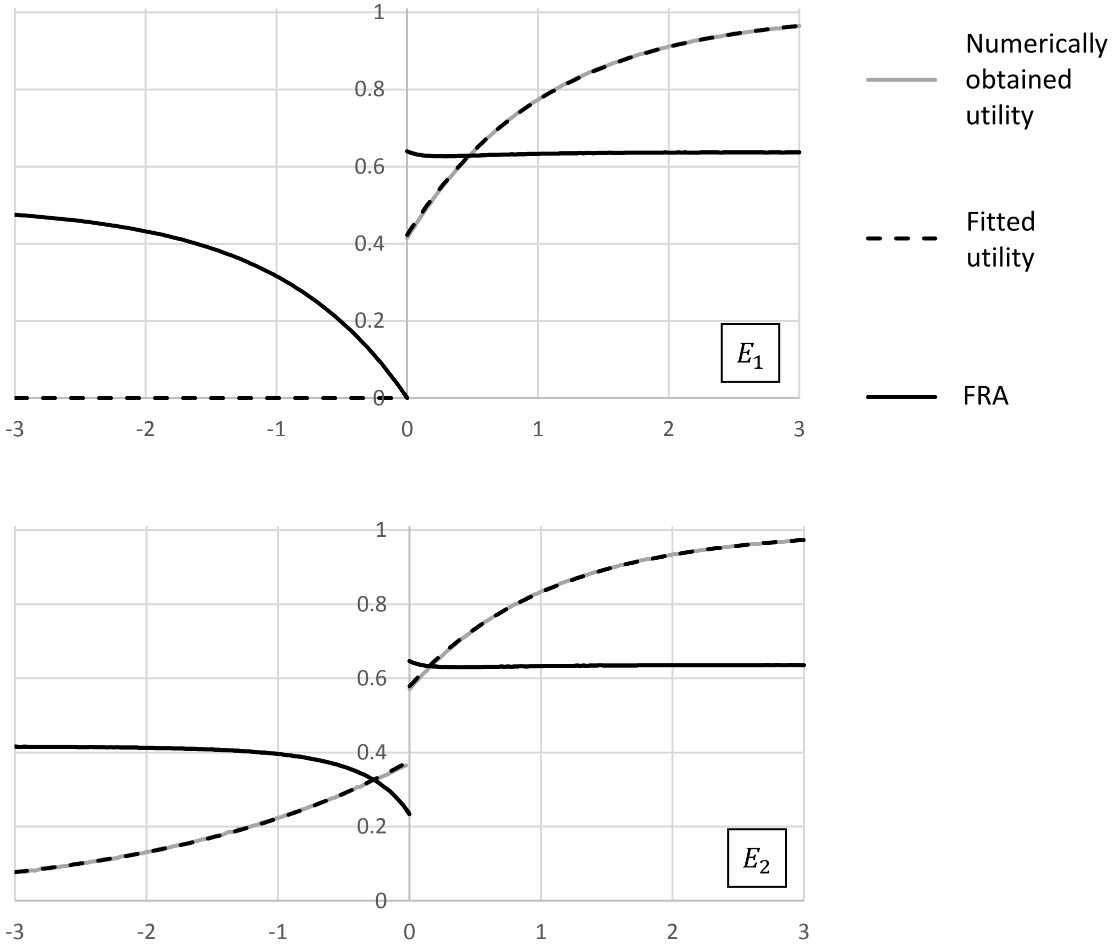

**Figure 2.** Fractional risk aversion (FRA) for the variant I of the environments $E_1$ and $E_2$.

Using fitted functional forms rather than numerical estimates smoothens the utility functions making the results less reliable on the precision of stochastic optimization methods. Moreover, it speeds up the Monte Carlo evaluation of integrals. The results are presented in Figure 2. In both cases FRA confirms results of the visual inspection. However, it also suggests that the risk aversion drops as we approach the point of discontinuity from the left. This is an important feature which is not immediately obvious after visually inspecting local curvature. I discuss this result further in Section 4.

Finally, let us consider environments $E_8$ and $E_9$, similar to the environments considered so far. An individual living in $E_8$ or $E_9$ is still characterized only by the amount of resources she has at the beginning of period $t$: $x_t \in R$. In every period, the individual consumes $d_t \sim \mathrm{U}[0,1]$. In every period, the individual is provided with a foraging opportunity characterized by a potential gain $g_t$, loss $l_t$, and a probability of success $p_t$, where $g_t, l_t \sim \mathrm{U}[0,2]$ and $p_t \sim \mathrm{U}[0,1]$. The individual lives for $T$ periods. After $T$ periods, a mating season occurs. If the individual enters the mating period with sufficient amount of resources ($x_{T+1} \geq K$, where $K > 0$), then her performance equals 1 otherwise ($x_{T+1} < K$) it equals 0. Moreover, the individual living in $E_8$ faces a possibility of starvation. Her performance equals 0 if $x_t < 0$ for any $t$. In $E_9$ there is no starvation possible.

An agent living in such environments does not have a single utility function to guide her decisions. To see why, suppose that after today's consumption the individual has $K + 1$ unit of resources. If the individual is at the beginning of her lifespan, it makes sense for her to accept some foraging opportunities that can improve her situation. However, if the individual is in period $T - 1$, she will surely survive to the mating season even if she does not engage in any foraging activity. In other words, if she refuses to accept any gambles, she has a probability 1 of success. Accepting any gambles reduces her expected performance. As a result, for a given amount of resources, certain gambles will be accepted in the initial periods but rejected towards the end.

Therefore, not in all stochastic environments natural selection will automatically equip an individual with a single utility function. The construction of a utility functions in the environments $E_1 - E_7$ rests on a proposition that a single utility function indeed yields optimal decisions in these environments. This proposition is proven in Section 3.

## 3. Analytical results

In this section I construct a general model of behavior in a stochastic environment (MBSE) and prove that under stationarity of the environment, a single utility function is the optimal decision policy. I also show how to calculate the value of such utility function based on expected performance. These results imply that using a single utility function is indeed optimal for environments $E_1 - E_7$. It also allows to analytically identify properties of some of the optimal utility functions.

Consider an agent operating in a stochastic environment with discrete time $t = 1,2, \ldots$. Agent's state in period $t$ is described by $x_t \in X$, where $X$ is the state space. Let the sequence of states $h^T = (x_1, x_2, \ldots, x_T) \in X^T$ be called agent's life history up to period $T$. Denote $H = \bigcup_{T=1}^{\infty} X^T$ as a set of all potential life histories.

At the beginning of each period $T$, the process (life of the agent) may get terminated. There exists a probability distribution $q_1(\cdot\,|h^T)$ over the set $\{0,1\}$, where 1 indicates process termination. Let random variable $\phi_T$ be a process termination indicator. Then, $q_1(\phi_T = 1|h^T)$ is the probability that the process gets terminated when life history $h^T$ is reached and $q_1(\phi_T = 0|h^T)$ is the probability that the process continues.

A prospect $p$ is a probability distribution over $X$. Let $P$ denote the set of all prospects that can occur. In every period, the agent has to choose from a choice set $c_T$ that contains a collection of prospects. Let $C \subset 2^P$ be a set of all available choice sets. For every life history $h^T \in H$, there exists a probability distribution $q_2(\cdot\,|h^T)$ over $C$. That is probability of obtaining particular choice set depends on the agent's life history.

The agent is equipped with a decision policy, which is a function $D: H \times C \to 2^P$ taking a life history and a choice set as an argument and returning a set of preferred prospects: $D(h^T, c_T) \subset c_T$. Moreover, there exists conditional probability distribution $q_3(\cdot\,|D(h^T, c_T))$ over $P$ such that probability of selecting any of the prospects from $D(h^T, c_T)$ is equally likely. In other words, the agent determines which prospects from $c_T$ are preferred, and then uses a fair coin toss (or similar procedure in case of more than two preferred prospects) to determine which of these equally preferred prospects to eventually select. Let $p_T \in D(h^T, c_T)$ denote the prospect selected in period $T$.

The selected prospect $p_T$ obtains and the initial state of the period $T+1$ is determined: $x_{T+1} \sim p_T$ (state $x_1$ is determined according to given probability distribution $p_0$). $q_1, q_2, q_3,$ and $p_T$ can be combined into joint distribution. That is, given the decision policy $D$, there exists a probability distribution $q_4(\cdot\,|h^T)$ over $\{0,1\} \times C \times P \times X$ which is based on $q_1, q_2, q_3,$ and $p_T$ and which jointly determines $\phi_T$, $c_T$, $p_T$, and $x_{T+1}$. $\{0,1\} \times C \times P \times X$ is a period-specific sample sub-space, and $\Omega = X \times (\{0,1\} \times C \times P \times X)^\infty$ is the entire sample space of a stochastic environment. An elementary event of this space is

$$\left(\begin{pmatrix}0\\0\\0\\x_1\end{pmatrix}, \begin{pmatrix}\phi_1\\c_1\\p_1\\x_2\end{pmatrix}, \begin{pmatrix}\phi_2\\c_2\\p_2\\x_3\end{pmatrix}, \dots\right) = \omega \in \Omega$$

The existence of the probability measure over $\Omega$ is generally not guaranteed and requires certain assumptions about the involved sets and functions, most importantly their measurability. For the sake of result's generality, I will assume that measurability conditions are satisfied and that related questions need to be addressed separately. Thus, I assume that appropriate $\sigma$-algebra over $\Omega$ and appropriate probability measure $\mu_D$ (which depends on the adopted decision rule $D$) exist and so the probability space is well defined.

The performance function is $f: H \to R$. Moreover, let a random variable $\phi = \min\{T: \phi_T = 1\}$ indicate the number of the period when process terminates. Then, random variable $g = f(h^\phi)$ is the value of the performance function at the moment of process termination. Fitness $F$ is the expected value of the performance function given the adopted decision rule, that is

$$F(D) = E_D g = \int_\Omega g d\mu_D = \int_\Omega f(h^\phi) d\mu_D$$

Assuming that $F(D)$ exists, an optimal policy is a decision policy $D^*$ such that

$$D^* \in \arg\max_D F(D)$$

A utility function policy is a stationary decision policy for which there exist a function $u: X \to R$ (called a utility function) such that for any $T$ the decision policy returns prospect $p \in c_T$ only if for all other prospects $q \in c_T$

$$\int_X u(x_{T+1}) dp(x_{T+1}) \geq \int_X u(x_{T+1}) dq(x_{T+1})$$

where $p(x_{T+1})$ indicates that $x_{T+1}$ is the variable of integration (I will keep using this somewhat sloppy notation for simplicity). Denote $v(h^T, D)$ to be expected value of the performance function given the life history $h^T$. That is

$$v(h^T, D) = E_D[g|h^T] = \int\limits_{\Omega_{h^T}} g d\mu_D^{h^T}$$

Where $\Omega_{h^T} = \{\omega \in \Omega: (x_1(\omega), \dots, x_T(\omega)) = h^T\}$ and $\mu_D^{h^T}$ is the conditional probability measure on $\Omega$ induced by decision policy $D$ and normalized so that $\mu_D^{h^T}(\Omega_{h^T}) = 1$. If $F(D)$ exists than $v(h^T, D)$ exists almost surely (with probability 1).

**Lemma 1 (Bellman optimality principle)**. $D^* \in \arg\max_D v(h^T, D)$ almost surely.

Lemma 1 indicates that the optimal policy does not have to be well-behaved on a measure-zero set. This is because optimality is defined in terms of an integral (expected value). Also, Lemma 1 suggests conditions for the preference between prospects: a prospect that yields inferior expected value of the performance function cannot be chosen over a prospect that yields superior one. This is made explicit in Lemma 2.

**Lemma 2 (almost sure rationality).** An optimal decision policy almost surely selects a prospect $p \in c_T$ only if for any $q \in c_T$

$$\begin{aligned} E_p v(h^{T+1}, D^*) &= \int_X v(h^{T+1}, D^*) dp(x_{T+1}) \geq \int_X v(h^{T+1}, D^*) dq(x_{T+1}) \\ &= E_q v(h^{T+1}, D^*). \end{aligned}$$

Lemma 2 does not guarantee that once we reach history $h^T$ the agent almost surely selects a prospect that maximizes expected performance function. In fact, after reaching particular history $h^T$, the agent can do anything, unless probability of reaching $h^T$ is positive (which is never the case if the underlying probability distributions are continuous).

Implication of Lemma 2 is thus weaker: with probability 1 a life history obtains such that in each period agent choses a prospect with the highest expected performance.

Let us say that a stochastic environment is stationary if and only if probability of process termination and probability distribution over choice sets depend only on the current state, that is $q_1(\cdot\,|h^T) = q_1(\cdot\,|x_T)$ and $q_2(\cdot\,|h^T) = q_2(\cdot\,|x_T)$ where $h^T = (x_1, \ldots, x_T)$.

**Theorem (optimality of utility function policies).** Assume that a stationary stochastic environment has performance function $f(h^T) = \sum_{t=1}^{T} r(x_t) + R(x_T)$. Then, as long as optimal policy $D^*$ exists, the expected value of the performance function is almost surely $v(h^T, D^*) = \sum_{t=1}^{T-1} r(x_t) + \tilde{v}(x_T)$ for $T > 1$ and $v\big((x_1), D^*\big) = \tilde{v}(x_1)$ for $T = 1$. The optimal policy is almost surely a utility function policy, with a utility function $u(x) = a + b\tilde{v}(x)$ for any $a \in R$ and $b > 0$.

Let us call $u_0(x) = \tilde{v}(x)$ the baseline utility function. Two useful corollaries follow immediately from the Theorem and are presented without formal proof.

**Corollary 1.** Assume that a stationary stochastic environment has performance function $f(h^T) = R(x_T)$. Then, as long as optimal decision policy $D^*$ exists, almost surely $v(h^T, D^*) = \tilde{v}(x_T)$ and the optimal policy is a utility function policy.

**Corollary 2.** Assume that a stationary stochastic environment has performance function $f(h^T) = -T$. Then, as long as optimal decision policy $D^*$ exists, almost surely $v(h^T, D^*) = 1 - T + \tilde{v}(x_T)$ and the optimal policy is a utility function policy.

Corollary 1 can be immediately applied to environments $E_1 - E_7$ to show that their optimal decision policy (if one exists) must be a utility function policy. Environments $E_8$ and $E_9$ are not stationary, because probability of process termination depends on the number of the current period rather than the current state. Corollary 2 can be interpreted in the following way: if agent’s objective in a stationary stochastic environment is to achieve a goal as quickly as possible, then it is optimal to use a utility function policy. The value of baseline utility function equals negative of the expected number of periods before reaching the goal. In this case, achieving goal may be interpreted as reaching certain subset of states $G \subset X$ for which probability of process termination is 1.

The Theorem and both corollaries assume that solution exists and then provide its properties. Proving that the solution exists needs to be done separately. Proposition 1 proves existence by directly demonstrating the solution for the environment $E_7$. Proposition 2 does not prove existence but proves some properties of the optimal utility function in the environment $E_2$. The purpose of these two propositions is to show how to apply the Theorem and to boost credibility of numerical results presented in section 2.

**Proposition 1.** Let $X = R$, $f(h^T) = x_T$, $\Phi(h^T) = p_0 \in (0,1)$, and in every period $c_t$ contains two elements: (1) $x_t - d_t$ with probability 1 and (2) $x_t - d_t + g_t$ with probability $p_t$ and $x_t - d_t - l_t$ with probability $1 - p_t$.

a) If $d_t \sim \mathrm{U}[0,2\bar{c}]$, $p_t \sim \mathrm{U}[0,1]$, and $l_t, g_t \sim \mathrm{U}[0,2]$ are independently distributed, then the baseline utility function $u_0(x) = \left(\frac{1}{6}(\ln(16) - 1) - \bar{c}\right)\frac{1-p_0}{p_0} + x$ is the optimal decision policy.

b) If $d_t \sim \mathrm{Exp}\left(\frac{1}{\bar{c}}\right)$, $p_t \sim \mathrm{U}[0,1]$, and $l_t, g_t \sim \mathrm{Exp}(1)$ are independently distributed, then the baseline utility function $u_0(x) = \left(\frac{1}{3} - \bar{c}\right)\frac{1-p_0}{p_0} + x$ is the optimal decision policy.

| Variant | Slope | | Intercept | | | $1 - R^2$ |
|---|---|---|---|---|---|---|
| | $1 - \hat{a}$ | St.Dev. | $b_0$ | $b_0 - \hat{b}$ | St.Dev. | |
| I | 4.016E-5 | 5.712E-5 | 2/15 | 4.755E-3 | 9.934E-5 | 7.798E-7 |
| II | 7.918E-5 | 4.553E-5 | -4/15 | 2.898E-3 | 7.919E-5 | 4.955E-7 |
| III | -2.504E-5 | 5.750E-5 | 17/90 | 3.346E-3 | 1.000E-4 | 7.902E-7 |
| IV | -1.511E-4 | 4.341E-5 | -1.827E-2 | 3.415E-3 | 7.551E-5 | 4.503E-7 |

**Table II.** Estimation of the linear utility functions in the four variants of $E_7$. $\hat{a}$ and $\hat{b}$ are the fitted values. Corresponding values calculated analytically are 1 and $b_0$ respectively. To avoid superfluous 0s and 9s I report $1 - \hat{a}$, $b_0 - \hat{b}$, and $1 - R^2$ instead of $\hat{a}$, $\hat{b}$, and $R^2$. Scientific notation is used for the same reason. Each estimation is based on 241 observations.

To see whether numerical results in Section 2 agree with the theory I use OLS to fit $y = ax + b$ for each of the four variants of $E_7$. The results are presented in Table II. The numerically calculated utility functions are very close to 45-degree lines. Their intercepts are consistently underestimated. Although the magnitude of the differences between theoretical and estimated utility values is relatively small (around 0.1% for $x = 3$), the difference is in all cases statistically significant with t-statistic exceeding 30. I suspect that the source of this bias is in weak convergence properties of the stochastic optimization algorithm. Hence, similar bias is to be expected for other utility functions in Figure 1.

**Proposition 2.** Let $X = R$, $\Phi(h^T) = p_0 \in (0,1)$, and in every period $c_t$ contains two elements: (1) $x_t - d_t$ with probability 1 and (2) $x_t - d_t + g_t$ with probability $p_t$ and $x_t - d_t - l_t$ with probability $1 - p_t$ where $d_t \sim \text{Exp}\left(\frac{1}{\bar{c}}\right)$, $p_t \sim \text{U}[0,1]$, and $l_t, g_t \sim \text{Exp}(1)$ are independently distributed. Also, let $f(h^T) = \mathbf{1}(x_T \geq 0)$. Then, assuming the solution exists, the optimal decision policy can be almost surely described by a baseline utility function $u_0(x)$ satisfying the following conditions:

a) $\lim\limits_{x\to-\infty} u_0(x) = 0$, $\lim\limits_{x\to+\infty} u_0(x) = 1$, and $0 < u_0(x) < 1$.

b) $u_0(x)$ is strictly increasing.

c) $u_0(x)$ is continuous everywhere except for $x = 0$ and $u_0(0) - \lim\limits_{x\to 0^-} u_0(x) = p_0$.

Unlike Proposition 1, Proposition 2 does not allow us to directly test whether results of the numerical procedure are correct (except for the size of the gap at $x = 0$). Instead, it allows us to visually inspect the estimated utility functions and validate their general properties. Furthermore, Proposition 2 does not prove the existence of the optimal policy. Existence of optimal policies for environments $E_1 - E_6$ is a conjecture.

## 4. Discussion

I start discussion with a purely mathematical aspect. MBSE can be converted into a Markov decision process (MDP). The Theorem solves a problem typical for the MDP literature: it proves that optimal decision policy is stationary and shows a way to find it numerically. In addition to applications specified in Section 2, the Theorem can be applied to a class of traditional MDP problems.

The second part of discussion is concerned with interpretation. Fitness and utility are two different concepts even though an agent seems to maximize both of them simultaneously. Also, the stochasticity of environment should be interpreted in terms of subjective probabilities. Validity of MBSE depends on whether subjective probabilities are a good way to model incomplete information.

The third part deals with the main result of this article and its implications. MBSE challenges a popular normative assumption that individuals use a single utility function. Moreover, it shows how maximization of expected utility is methodologically reducible to maximization of fitness and that utility may be empirically measurable and may have well defined units.

Finally, I discuss the models presented in Section 2. MBSE provides a single framework for various optimal foraging models and evolutionary analysis of risk aversion. The models allow to identify three sources of human risk preferences: possibility of death, diminishing returns to having more resources, and possibility of gaining or losing social status. They also rationalize globally concave utility functions, predict higher risk-loving for individuals just below reference point than those far below it, and suggest that risk aversion may be an evolutionary mismatch.

### *4.1. Relation to Markov decision processes*

MBSE can be reformulated as a total-reward MDP. Original state space can be combined with the space of choice sets and a terminal state in order to form an MDP-like state space. Each element of this new state space is then associated with an original state and a single choice set. Action space can be built out of all subsets of $P$. After agent selects an action, the transition probability into a subsequent state depends on: (1) original transition probabilities, (2) probability of process termination (transition into the terminal state), and (3) probability distribution over choice sets (to determine the choice set component of the subsequent MDP-like state). Upon reaching the terminal state, there is a final reward (equal to the agent's performance) and further transition probabilities maintain the process in the terminal state.

The Theorem can be crudely restated in the language of MDPs as: "for a certain stationary process, the optimal policy is stationary and can be expressed in the form of a utility function." The literature on MDPs has a lot of results concerning existence of optimal stationary policies and ways to calculate them. Stationary policies are not in general optimal for finite-horizon models (cf. $E_8$ and $E_9$). Infinite-horizon models are most extensively studied for finite or countable state and action spaces. Infinite-horizon models can be generally divided into three groups: discounted-reward, total-reward, and average-reward problems. We are interested in infinite-horizon, total-reward, uncountable state MDPs. The analysis thereof has been focused mostly on two further categories: positive models and negative models, both of which impose restrictions on rewards. These restrictions are in general not satisfied by our stochastic environments (e.g. rewards in $E_7$ are unbounded). The remaining literature consists of several papers concerned mostly with $\varepsilon$-optimality of stationary strategies. There does not seem to exist a result that could be directly applied in place of the Theorem. Feinberg and Schwartz (2002) and Puterman (2005) provide in-depth introductions to MDPs and extensive literature reviews.

The Theorem can potentially have other MDP-related applications in addition to those presented in Section 2. As an example, consider a simple 4x3 environment with a sequential decision problem described by Russell and Norvig (2010, p. 646). In this environment, an agent is located on a 4 by 3 lattice. In every period, the agent has to choose the direction of movement. However, the actual direction of movement can differ from the chosen one. The actual direction of movement is what the agent chose with probability 0.8 and can deviate by 90 degrees clockwise or counterclockwise with probability 0.1 each. If the actual direction of movement leads the agent out of the lattice or onto an obstacle, no movement occurs. Otherwise, agent moves to the adjacent lattice node. In addition to an obstacle, the lattice contains two goals. Upon reaching any of them, the process terminates. One of the goals yields a reward of +1 and the other yields a reward of -1. Moreover, every period subtracts 0.04 from the final reward. The objective is to find a decision policy that maximizes the expected reward. This simple 4x3 environment satisfies the conditions of the Theorem. It immediately follows that the optimal decision policy is stationary (if it exists) and is in the form of a utility function, a fact otherwise well known.

Finally, the Theorem may improve algorithms looking for numerical approximations of optimal policies for problems with uncountable state spaces. According to the Theorem, the value of the utility function for a given state can be calculated based on the expected performance when starting in this state. This implies that $u_0(x) = 0$ for $x < 0$ in $E_1$, $E_3$, and $E_5$. It also implies that there may be no need to look for the entire utility function simultaneously. In fact, for the problems of Section 2, instead of approximating the entire utility function using high-density piecewise linear functions, I use low-density piecewise linear functions to find the maximum expected performance for a particular state. Then, I repeat this process for interesting states. The pictures in Figure 1 would be much coarser (with just a few clearly visible linear pieces) if I did not apply the Theorem while writing the optimization algorithm. It is also worth noting that after using this procedure, the estimated utility function can be further improved by policy iterations. For example, a single policy iteration would entirely remove the bias (described in Section 3) with which I estimate the utility function for the environment $E_7$.

### *4.2. Interpretation of utility, fitness, and stochasticity*

A contemporary formal notion of utility function most often follows Von Neumann and Morgenstern (1964). According to this view, a utility function can be derived from preferences and can be used to efficiently determine which states of the world or gambles over the states of the world would be selected by a rational individual. It is often argued (or assumed) that agents have such a utility function and that it helps them make a decision whenever they face a choice between prospects. For examples see Friedman and Savage (1948) or Russel and Norvig (2010, p. 651). Meanwhile, the same authors often define utility as a function of policy. The utility function no longer helps to compare states of the world. Instead, it helps to compare decision policies in the sequential decision-making processes. This can be seen for example in Dubins and Savage (1965, p.25) or Russel and Norvig (2010, p. 647). Numerous other authors follow suit and some authors (e.g. Real and Caraco, 1986, Cooper, 1987, De Freja, 2009, or Kenrick et al., 2009) explicitly equate fitness with utility.

MBSE shows how this “let us call a utility whatever is being currently maximized” approach can lead to confusion. Following this approach, MBSE has two distinct objects that should be called a utility function. One is the way of making single decisions – this is what I call a utility function in this article. The other is the assessment of a decision policy

– something that is also often called fitness. Hence, following this terminology, not only we have a single term to describe two distinct objects but we also have two terms to describe the same object. For that reasons, I decided to use term utility function only in its traditional, non-sequential meaning, when the argument is a state of the world or a consumption bundle. Fitness on the other hand is a function of chosen utility function(s) defined as expected level of overall lifetime success. Rayo and Becker (2007) as well as Robson and Samuelson (2011) seem to follow similar distinction.

The following list summarizes differences between fitness and utility:

1) Fitness is a function of decision policy. Utility is a function of world state.
2) Expected Utility Theory of von Neumann and Morgenstern (1964) does not apply to fitness, since notion of probability distributions over decision policies is hardly ever an issue. On the other hand, the notion of probability distributions over world states is often very useful.
3) Fitness is maximized by the agent's designer (or the principal): natural selection in case of living organisms or a human engineer in case of artificial agents. Utility is maximized by the agent.
4) Since it is defined as expected value, fitness can be empirically measured only for a population. If utility can be empirically measured, then the measurement involves a single individual.

There is a link between the two quantities which can be explained with introduction of the third term: performance. This term is well established in the theory of artificial intelligence (see Russel and Norvig, 2010, p. 37). Performance is the quantity that can be empirically measured after observing agent's behavior. In case of an animal, it may be a number of descendants and in case of an artificial agent it can be the quantity or quality of agent's output.

By definition, rational agent maximizes her expected performance. Expected performance of an agent can be empirically estimated at any stage of agent's operation, after observing actual performance of identical agents under similar circumstances. Fitness is the a priori expected performance which takes into account all possible circumstances that can occur during agent's operation. Finally, utility informs the agent which states should she prefer in order to maximize her expected performance. Utility can be equal to expected performance, but it can also be its positive affine transformation so that there is a single utility function even though expected performance for a given state varies from period to period (as in Corollary 2).

Let us now discuss stochasticity of the environment. MBSE seemingly assumes that the environment works according to laws of probability and the agent uses a decision-making mechanism compatible with these laws. In reality, aside from some quantum effects, the physical world can be well approximated as deterministic. An agent with precise information on initial conditions and cognitive ability to model physical processes sufficiently quickly should be able to predict the outcome of a coin toss with very high degree of certainty (Strzałko et al., 2008). In fact, the perceived randomness of a coin toss comes mainly from cognitive ineptitude of humans. The same applies to the vast majority of other apparently random events.

The stochasticity of environment comes not from its intrinsic randomness but from the incomplete information an agent has about the environment. That is, the underlying assumption is not that the environment is random, but that an agent uses subjective probability to model her incomplete information about the environment and that she does it accurately. This assumption may be correct only as long as the optimal way to represent uncertainty in an evolved or designed agent are indeed subjective probabilities. Russell and Norvig (2010, chapters 13 and 14) provide a brief but informative overview of the literature on the use of subjective probabilities and alternative theories.

### *4.3. Main results*

MBSE aims at being as general as possible. In every period, individual is provided with a choice set containing probability distributions over the state space. State space is arbitrary: it can be a finite set, a countable set, or an uncountable set like $R^n$. Choice sets can also be finite, countable or uncountable. A state can reflect variables both observable and unobservable to the agent. Available choice sets and performance of the agent depend on her entire history in an arbitrary way. This setup allows for a high degree of flexibility and permits construction of detailed models for a large class of problems actual agents may face.

Any rational agent, by definition, maximizes her expected performance (cf. Lemma 2). A rational agent uses a single utility function if the environment she was designed for satisfies certain conditions. Some of these conditions are identified by the Theorem: the environment must be stationary. Since these are only sufficient condition, there may still be other circumstances under which single utility function is optimal. For example, in a finite-horizon environment like $E_9$, a single utility function is optimal if the objective is to maximize the expected amount of resources ($x_{T+1}$).

Violating stationarity can lead to a situation when a single utility function is not an optimal decision policy. Potential causes for a non-stationary optimal policy include performance function and probability of process termination changing over time (the latter exemplified by $E_8$ and $E_9$). Stationarity is a restrictive condition and such violations are likely to occur in sophisticated natural environments. This suggests the following non-stationarity conjecture: human and animal utility functions change over time.

The idea that agents have a single utility function only under special circumstances has a number of implications. Insofar as natural selection is efficient in creating rational agents, MBSE can be treated as positive model of behavior that offers testable predictions. Humans and other animals should have time-invariant preferences in stationary environments. But, more interestingly, according to the non-stationarity conjecture, humans and animals should have variable preferences in non-stationary environments. Examples of such environments are easy to find. For a human, consider a person whose objective is not to go into red between two paychecks. For an animal, consider a small bird or mammal that needs to accumulate enough calories during a day to survive the subsequent winter night. These situations can be modeled by $E_8$. In both of them, risk preference of a rational agent changes with time.

Many normative models of behavior assume that individuals use certain utility function to make decisions. Examples from economics include a typical intertemporal consumption problem, $\max_{c} \sum_{t=0}^{\infty} \beta^t u(c_t)$, and attempts to explain human attitudes to risk with a single utility function of money or wealth (e.g. Friedman and Savage, 1948). MBSE implies that this assumption cannot be unconditionally maintained. Using a single utility function is in general not normatively correct and thus requires justification. One potential justification, provided by the Theorem, is that the behavior under consideration occurs in a stationary environment. The optimal policy can be also approximated by a single utility function if such approximation does not significantly affect results.

A potential strength of MBSE relative to traditional normative and positive models of behavior is its immediate methodological reducibility. Consistent methodological reductionism leads to a conclusion that all observable phenomena can be ultimately explained by the fundamental forces of physics. Behavior of humans and other animals is no exception. Anderson (1972) remarks that “the workings of our minds and bodies, and of all the animate or inanimate matter of which we have any detailed knowledge, are assumed to be controlled by the same set of fundamental laws, which except under certain extreme conditions we feel we know pretty well.” Anderson presents a hierarchy of sciences in which, roughly speaking, physics is the foundation of chemistry, chemistry is the foundation of biology, biology is the foundation of psychology, and psychology is the foundation of social sciences.

Methodological reducibility ensures that the assumptions used to develop a theory have robust foundations. If an explanation of a phenomenon cannot be ultimately reduced to the fundamental forces of physics, then it likely contains unjustifiable assumptions and the scientific status of such explanation is dubious. This is especially important in case of theories about human behavior, which too often hinge on intuitions and preconceptions about human nature (examples abound in social sciences and philosophy). Thus, the scientific status of the theory that people act as if they were maximizing their expected utility depends on whether such theory can be derived from basic laws of physics.

In this research project, I show how maximization of expected utility may be explained by the tendency of natural selection to maximize fitness. This provides a reductionist link between one of the most prominent theories used in economics and one of the most prominent theories in theoretical biology (for a reductionist link between fitness optimization and even more basic laws see Grafen, 2002). This research project is not the first attempt at providing such a link, some of the previous work include Karni and Schmeidler (1986), Cooper (1987), Robson (1996), and Robson (2001b). I would like to extend this link by supposing the following conjecture: all optimizing behavior people engage in can be ultimately derived from the optimizing nature of natural selection.

Another interesting implication of MBSE comes from that fact that the Theorem identifies the value of baseline utility function. The prevailing viewpoint among welfare economists is that utility cannot be measured and interpersonal comparisons of utility are impossible. The two main reasons behind this stance is, first, that a utility function as a decision-making mechanism under uncertainty can be uniquely described only up to a positive affine transformation and, second, that even if the value of a utility functions can be uniquely determined, it is not known a priori how to weight the utility of one person against the utility of another. As Arrow (1963) puts it: “It requires a definite value judgement not derivable from individual sensations to make the utilities of different individuals dimensionally compatible and still a further value judgement to aggregate them according to any particular mathematical formula.”

MBSE allows us to look at the first of the arguments against cardinal utility from a new angle. Consider a population of individuals living in the stochastic environment $E_1$ described in Section 2. An astute observer should quickly realize that performance of an individual depends on her ability to procreate before she starves. Such an observer would soon notice that, on average, 77% of individuals having one unit of resources procreate and that 91% of individuals having two units of resources procreate. An individual having one unit of food is thus clearly worse off than an individual who has two units of food. Corollary 1 suggests that the baseline utility equals probability of survival. The value of baseline utility is thus in this case empirically measurable, has precise interpretation, and allows for meaningful comparisons across individuals.

On the other hand, the baseline utility function obtained in an environment satisfying premises of Corollary 2 does not allow for such comparisons. An individual who has been working on her task for 10 periods is clearly worse off than a similar individual who has been working for only 5 periods, even if the expected remaining time (i.e. the value of the baseline utility function) is the same for both individuals. As a result, the value of utility function does not seem to be appropriate for comparisons across individuals.

However, there exists another potentially more suitable variable: expected performance. For the individuals inhabiting the same environment, expected performance can be, in principle, estimated with empirical data and has the same interpretation and the same units for all individuals. In special cases, like those described by Corollary 1, expected performance equals baseline utility.

Measurement of utility or expected performance may prove to be hard in practice and is related to the very complicated questions of what contributes to and what reduces fitness. There are also other difficulties with using expected performance to identify individual welfare. One such difficulty is that humans evolved in an environment quite different to the one present today. Let us call the pre-industrial environment $E_0$ and the post-industrial environment $E_+$. The two environments differ both in their state spaces and in available choice sets. A decision policy that was created to maximize fitness in $E_0$ can be different than a decision policy that would have maximized fitness in $E_+$. This is why the people of today often make choices that in obvious way hamper their reproductive success, for example by accepting risk of smoking or obesity or giving up reproduction for career reasons – a phenomenon called evolutionary mismatch (Robson and Samuelson, 2011). As a result, even if practically measurable, the value of performance function of an individual may not be well correlated with the fulfilment of her revealed preferences.

Finally, the problem of aggregation of utilities or expected performances remains untouched by this analysis. It is easy to imagine that in a multi-agent system, maximizing the sum of expected performances of agents would tautologically maximize the expected performance of the entire system. As long as the definition of biological fitness is “ability to propagate own genes,” application of such procedure to humans would result in maximizing overall propagation of human genome in the Universe. Whether this is the desired outcome I dare not to speculate.

### *4.4. Example foraging models*

Optimal foraging theory is concerned with decisions animals make while searching for food. An important part of this theory is concerned with decisions under risk (see Kacelnik and Bateson, 1996, for a literature review). Real and Caraco (1986) review a series of experiments in which animals have to choose between steady sources of food in which one visit yields constant return and sources of food in which return is a random variable with the expected value equal to or exceeding that of the steady supply. The experiments were done using bees, wasps, various species of birds, shrews, and rats. All animals exhibited risk aversion in normal conditions. Experiments with birds and shrews also indicated that subjects were risk-loving if they were headed towards starvation (or were below threshold required to achieve other important biological goal like seasonal migration).

The environments $E_1 - E_9$ are explicitly optimal foraging models. $E_1$ can be used to explain the behavior of birds and shrews summarized by Real and Caraco (1986). Small size of these animals indicates real possibility of starvation if they fail to acquire any resources in several subsequent foraging endeavors (the idea that maximization of probability of survival results in animal risk aversion was suggested by Caraco, 1980). The solution to $E_1$ is consistent with empirical observations: such animals should be risk-averse in general but risk-loving if their energy balance goes below zero. Similarly, $E_2$ can be reinterpreted to reflect the situation of migratory birds. In this interpretation, state of the agent is the animal's weight relative to the weight needed for migration. The performance function is zero if migration does not occur and one if it does occur. Therefore, the animal is risk-loving until its weight crosses the threshold necessary for migration – then it becomes risk-averse, another prediction confirmed by empirical evidence.

Although foraging efforts seem to be far removed from the reality of modern human life, the human decision mechanism responsible for accumulating resources likely evolved in conditions when optimal foraging was critical for survival. Foraging models may be thus a good guide in an effort to find origins of human preferences. Based on this idea, MBSE provides a single framework for deriving various commonly discussed utility functions of money and identifying potential sources of risk preferences. Simple risk aversion may be led by the possibility of losing resources necessary for survival ($E_1$) or by diminishing returns to having more resources in terms of expected performance ($E_5$ and $E_6$). S-shaped risk preferences can be explained by possibility of losing social status by an individual above the reference point and willingness to accept risk to gain social status if the individuals is below it ($E_2$).

All these results have been already discussed in the literature. Borch (1966) investigates a simple stochastic environment in which optimal utility function of an agent is concave due to possibility of death. Borch interprets the agent to be a firm, whose bankruptcy is here equivalent to death, and the performance measure is the total discounted amount of dividend paid. The relationship between concavity of performance function and concavity of utility function has been considered by Robson (1996). Rubin and Paul (1979) explain how competition for status and sexual selection could produce risk seeking in young and risk aversion in older males. Rayo and Becker (2007) show how individuals with limited perception can evolve an S-shaped, reference-dependent value function. Robson (2001a) provides a literature review of evolutionary accounts of risk aversion. More recent literature review, but slightly less focused on theoretical models was written by Collins et al. (2016).

The analysis carried out in Section 2 not only demonstrates how to apply the Theorem in order to derive various evolutionary results within a single framework, but also has interesting implications on its own. First, environments $E_1$ and $E_2$ show why a rational individual would use a globally concave utility function, even though such functions are sometimes thought to have absurd level of risk aversion for large gambles (Rabin, 2000). For example, a sensible person may want reject a bet which yields -\$20,000 with probability 50% or \$1 billion with probability 50%, if the -\$20,000 results in significant permanent reduction in her performance (bankruptcy, homelessness, permanent loss of social status, loss of a mating partner, loss of custody of children, early death due to inadequate health care, etc.).

Second, analysis of risk profile generated by environment $E_2$ suggests that individuals just below reference point are more risk loving than individuals far below it.

This result is a direct consequence of discontinuity in the utility function: sharp increase induces individuals to accept gambles that can land them on the other side of the threshold despite their low expected value. This constitutes a significant empirically verifiable departure from traditional Prospect Theory, which assumes that value function is continuous at the reference point (Kahneman and Tversky, 1979).

Finally, collecting all evolutionary reasons for human risk preferences allows to reevaluate them as "deeply rational." While regular rationality means simply acting according to one's consistent preferences, deep rationality refers to preferences that serve a higher purpose – maximization of fitness. In the language of MBSE, decisions which are deeply rational are the decisions that maximize expected performance.

Kenrick et al. (2009) present risk aversion as one of the candidates for deeply rational preferences. To the contrary, I would like to consider a conjecture that the level of risk aversion in modern humans is an evolutionary mismatch (for more potential economics-related evolutionary mismatches see Rubin, 1982). Typical evolutionary explanations of risk aversion revolve around features of the environment that are no longer present in the modern society. Social safety nets make sure that individuals do not starve in absence of resources. Availability of credit allows individuals to temporarily go into negative equity. Fertility rates are often negatively correlated with the resourcefulness of the household members and their social status. On the other hand, excessive risk aversion can potentially decrease the amount of investment undertaken by individuals, not only contradicting their own stated preferences for wealth but also hampering economic growth and welfare of the entire society. Therefore, if this conjecture is true, it is more appropriate to call human risk aversion "deeply irrational."

## Appendix

**Proof of Lemma 1.** If, in every period, probability that $D^* \in \arg\max_D v(h^T, D)$ is one, then the proof is over. Assume otherwise: there exists a period $T$ in which $\mu_{D^*}\left(\left\{\omega \in \Omega : D^* \notin \arg\max_D v(h^T, D)\right\}\right) > 0$. Let us denote the first such a period as $T_0$.

$\mu_{D^*}\left(\left\{\omega \in \Omega: D^* \notin \arg\max_{D} v(h^{T_0}, D)\right\}\right) > 0$ implies that there exists a set of $T_0$-long life histories $H_0 \subset H$ such that $\forall h^{T_0} \in H_0, D^* \notin \arg\max_{D} v(h^{T_0}, D)$ and $\mu_{D^*}(\{\omega \in \Omega: h^{T_0} \in H_0\}) > 0$. We can construct a decision rule $D'$ that improves on $D^*$ so that $\forall h^{T_0} \in H_0, v(h^{T_0}, D') = v(h^{T_0}, D^*) + \varepsilon(h^{T_0})$ where $\varepsilon(h^{T_0}) > 0$, but $D'(h^T, c_T) = D^*(h^T, c_T)$ for those life histories that are not in $H_0$. The measures $\mu_{D'}$ and $\mu_{D^*}$ are equal for all subsets of $\Omega$ for which $h^{T_0} \notin H_0$ as well as for subsets of $\Omega$ for which $h^{T_0} \in H_0$ but $x_t$ for $t > T_0$ are not specified – it is because these histories were reached using only those parts of policies $D'$ and $D^*$ that coincide. Denote $\Omega_0 = \{\omega \in \Omega: h^{T_0} \in H_0\}$.

$$F(D') - F(D^*) = \int_{\Omega} g d\mu_{D'} - \int_{\Omega} g d\mu_{D^*} = \int_{\Omega\backslash\Omega_0} g d(\mu_{D'} - \mu_{D^*}) + \int_{\Omega_0} g d\mu_{D'} - \int_{\Omega_0} g d\mu_{D^*}$$

Since $\mu_{D'} = \mu_{D^*}$ on $\Omega\backslash\Omega_0$ we can eliminate the first integral. We can also separate variables in the remaining two integrals. Let $\mu_D(H') = \mu_D(\{\omega \in \Omega: h^T \in H'\})$ for some $H' \subset H_0$. Then

$$F(D') - F(D^*) = \int_{H_0} \left( \int_{\Omega_{h^{T_0}}} g d\mu_{D'}^{h^{T_0}} \right) d\mu_{D'} - \int_{H_0} \left( \int_{\Omega_{h^{T_0}}} g d\mu_{D^*}^{h^{T_0}} \right) d\mu_{D^*}$$

$$= \int_{H_0} v(h^{T_0}, D') d\mu_{D'} - \int_{H_0} v(h^{T_0}, D^*) d\mu_{D^*}$$

But for $H' \subset H_0$ we have $\mu_{D^*}(\{\omega \in \Omega: h^T \in H'\}) = \mu_{D'}(\{\omega \in \Omega: h^T \in H'\})$ so

$$F(D') - F(D^*) = \int_{H_0} \left( v(h^{T_0}, D') - v(h^{T_0}, D^*) \right) d\mu_{D^*} = \int_{H_0} \varepsilon(h^{T_0}) d\mu_{D^*} > 0$$

As a result, $F(D') - F(D^*)$ can be simplified down to an integral of a positive function over a positive-measure set, and so $F(D') - F(D^*) > 0$ which contradicts that $D^*$ is optimal. ■

**Proof of Lemma 2.** If, in every period, probability that $\int_X v(h^{T+1}, D^*) dp(x_{T+1}) \geq \int_X v(h^{T+1}, D^*) dq(x_{T+1})$ is one, then the proof is over. Assume otherwise: there exists a period $T$ in which $\mu_{D^*}\left(\left\{\omega \in \Omega : \exists q \in c_T \wedge \int_X v(h^{T+1}, D^*)dp_T(x_{T+1}) < \int_X v(h^{T+1}, D^*)dq(x_{T+1})\right\}\right) > 0$. Let us denote the first such a period as $T_0$.

Similarly to the proof of Lemma 1, let us identify a positive-measure set of elementary events for which it is possible to construct a decision policy which improves on the original decision policy. Currently, we are interested in elementary events for which there are choice sets $c_{T_0}$ such that with positive probability an inferior prospect is selected. Let $\Omega_0 = \left\{\omega \in \Omega : \exists q \in c_{T_0} \wedge \int_X v(h^{T_0+1}, D^*)d\left[q\left(x_{T_0+1}\right) - p_{T_0}\left(x_{T_0+1}\right)\right] > 0\right\}$. By assumption, $\mu_{D^*}(\Omega_0) > 0$. Let us now construct the improved decision policy $D'$ which for $h^{T_0}$ and $c_{T_0}$ in $\Omega_0$ rather than $p_T$ selects the $q$ implicitly defined in the above definition of $\Omega_0$ and coincides with $D^*$ otherwise. Let us also formalize the relationship between $v(h^T, D)$ and $v(h^{T+1}, D)$:

$$v(h^T, D)$$
$$= q_1(1, h^T)f(h^T)$$
$$+ q_1(0, h^T)\int_C \int_{D(h^T, c_T)} \int_X v(h^{T+1}, D)dp_T(x_{T+1})\, dq_3\left(p_T \middle| D(h^T, c_T)\right) dq_2(c_T|h^T)$$

Notice that

$$\int_X v(h^{T_0+1}, D')dq\left(x_{T_0+1}\right) \geqq \int_X v(h^{T_0+1}, D^*)dp_{T_0}\left(x_{T_0+1}\right)$$

with a strict inequality on a positive-measure set. Similarly

$$\int_{D'(h^{T_0}, c_{T_0})} \int_X v(h^{T_0+1}, D')dq\left(x_{T_0+1}\right) dq_3\left(q \middle| D'\left(h^{T_0}, c_{T_0}\right)\right)$$

$$\geqq \int_{D^*(h^{T_0}, c_{T_0})} \int_X v(h^{T_0+1}, D^*)dp_{T_0}\left(x_{T_0+1}\right) dq_3\left(p_{T_0} \middle| D^*\left(h^{T_0}, c_{T_0}\right)\right)$$

holds with strict inequality on a positive measure set because the outer integral is simply the expected value with respect to the probability measure on the set of preferred prospects. Since $q_1$, $q_2$, and $f$ are the same for the two decision rules, we finally obtain

$$v(h^{T_0}, D') > v(h^{T_0}, D^*)$$

on a positive-measure set, which violates implication of Lemma 1. ■

**Proof of Theorem.** Let us conduct a proof by contradiction and assume that there does not exist a function $\tilde{v}(x)$ such that $\mu_{D^*}(\{\omega \in \Omega: v(h^T, D^*) = \sum_{t=1}^{T-1} r(x_t) + \tilde{v}(x_T)\}) = 1$.

Notice that by definition,

$$v(h^T, D^*) = \int_{\Omega_{h^T}} g d\mu_{D^*}^{h^T} = \int_{\Omega_{h^T}} f(h^\phi) d\mu_{D^*}^{h^T} = \int_{\Omega_{h^T}} \left[\sum_{t=1}^{\phi} r(x_t) + R(x_\phi)\right] d\mu_{D^*}^{h^T}$$

By assumption, $\phi \geq T$: we are interested in calculating $v(h^T, D^*)$ only if the process has not terminated before reaching $T$. $x_t$ are determined for $t \leq T$ in the subspace $\Omega_{h^T}$, thus:

$$v(h^T, D^*) = \sum_{t=1}^{T-1} r(x_t) + \int_{\Omega_{h^T}} \left[\sum_{t=T}^{\phi} r(x_t) + R(x_\phi)\right] d\mu_{D^*}^{h^T}$$

Hence, our current working assumption is equivalent to:

$$\forall \tilde{v}(x_T), \qquad \mu_{D^*}\left(\left\{\omega \in \Omega: \int_{\Omega_{h^T(\omega)}} \left[\sum_{t=T}^{\phi} r(x_t) + R(x_\phi)\right] d\mu_{D^*}^{h^T(\omega)} = \tilde{v}(x_T)\right\}\right) < 1$$

In other words there must exist a positive-measure set of life histories $h^T$ for which the expression $\int_{\Omega_{h^T}} \left[\sum_{t=T}^{\phi} r(x_t) + R(x_\phi)\right] d\mu_{D^*}^{h^T}$ is not a function of $x_T$, that is it can take different values for the same values of $x_T$.

Consider two life histories $h^{T_1}$ and $g^{T_2}$, for which terminal state is the same, such that

$$\int_{\Omega_{h^{T_1}}} \left[\sum_{t=T_1}^{\phi} r(x_t) + R(x_\phi)\right] d\mu_{D^*}^{h^{T_1}} < \int_{\Omega_{g^{T_2}}} \left[\sum_{t=T_2}^{\phi} r(x_t) + R(x_\phi)\right] d\mu_{D^*}^{g^{T_2}}$$

Note that there must exist a positive-measure set of such life histories $h^{T_1}$ for which exist a corresponding history $g^{T_2}$ and which do not derive from a shorter life history in this set (a longer history derives from a shorter history if the longer history starts with the shorter history). Let us now devise an improved decision policy $D'$. For any life history $\alpha$ which derives from $h^{T_1}$ (that is $\alpha$ arises by concatenating $h^{T_1}$ with some sequence $(x_1, \dots, x_N)$ of length $N = 0,1,2, \dots$) let us take a corresponding life history $\beta$ which derives from $g^{T_2}$ (that is $h^{T_1}$ and $g^{T_2}$ have the same terminal element and $\beta$ is derived from $g^{T_2}$ by concatenating it with the same sequence $(x_1, \dots, x_N)$ we used to derive $\alpha$ from $h^{T_1}$). The new decision policy is such that $D'(\alpha, c_T) = D^*(\beta, c_T)$ for all $\alpha$ that can be derived from all $h^{T_1}$ and the decision policies are equal for the remaining life histories. In other words, roughly speaking, the improved decision policy uses superior parts of the original decision policy and applies them wherever original decision policy is inferior to itself.

Given the stationarity and the same terminal element of $h^{T_1}$ and $g^{T_2}$, the corresponding conditional probability distribution governing the process are the same: $q_1(\cdot\,|\alpha) = q_1(\cdot\,|\beta)$ and $q_2(\cdot\,|\alpha) = q_2(\cdot\,|\beta)$. The measure induced by $D'$ on $\Omega_{h^{T_1}}$ is equivalent to the measure induced by $D^*$ on $\Omega_{g^{T_2}}$. Hence

$$\int_{\Omega_{h^{T_1}}} \left[\sum_{t=T_1}^{\phi} r(x_t) + R(x_\phi)\right] d\mu_{D'}^{h^{T_1}} = \int_{\Omega_{g^{T_2}}} \left[\sum_{t=T_2}^{\phi} r(x_t) + R(x_\phi)\right] d\mu_{D^*}^{g^{T_2}}$$

$$> \int_{\Omega_{h^{T_1}}} \left[\sum_{t=T_1}^{\phi} r(x_t) + R(x_\phi)\right] d\mu_{D^*}^{h^{T_1}}$$

That is $v(h^{T_1}, D') > v(h^{T_1}, D^*)$ and since this strict inequality holds for a positive-measure set of life histories $h^{T_1}$, the implication of Lemma 1 is violated. As a result, there must exist a function $\tilde{v}(x)$ such that $\mu_{D^*}(\{\omega \in \Omega: v(h^T, D^*) = \sum_{t=1}^{T-1} r(x_t) + \tilde{v}(x_T)\}) = 1$. Also, note that for $T = 1$ we have $\tilde{v}(x_1) = v(h^1, D^*)$.

Finally, according to Lemma 2, almost surely, a prospect $p \in c_T$ is selected only if for any prospect $q \in c_T$ we have

$$\int\limits_X v(h^{T+1}, D^*) dp(x_{T+1}) \geq \int\limits_X v(h^{T+1}, D^*) dq(x_{T+1})$$

$$\Leftrightarrow \int\limits_X \left[\sum_{t=1}^{T} r(x_t) + \tilde{v}(x_{T+1})\right] dp(x_{T+1}) \geq \int\limits_X \left[\sum_{t=1}^{T} r(x_t) + \tilde{v}(x_{T+1})\right] dq(x_{T+1})$$

$$\Leftrightarrow \sum_{t=1}^{T} r(x_t) + \int\limits_X \tilde{v}(x_{T+1}) dp(x_{T+1}) \geq \sum_{t=1}^{T} r(x_t) + \int\limits_X \tilde{v}(x_{T+1}) dq(x_{T+1})$$

$$\Leftrightarrow \int\limits_X \tilde{v}(x_{T+1}) dp(x_{T+1}) \geq \int\limits_X \tilde{v}(x_{T+1}) dq(x_{T+1})$$

Denote $u(x) = a + b\tilde{v}(x)$ for any $a \in R$ and $b > 0$. Then prospect $p \in c_T$ is selected only if for any prospect $q \in c_T$

$$\int\limits_X u(x_{T+1}) dp(x_{T+1}) \geq \int\limits_X u(x_{T+1}) dq(x_{T+1})$$

which satisfies the definition of a utility function policy and concludes the proof. ■

**Proof of Proposition 1.** In both cases (a) and (b) the stochastic environment satisfies conditions of Corollary 1. Therefore if optimal decision policy exists, it almost surely must be a utility function. Proving that solution exists is done by demonstrating the solution.

According to the Theorem, baseline utility function $u_0(x)$ is the expected value of the performance function if the initial state is $x$. Consider a shifted utility function $u_a(x) = u_0(x - a)$. An agent using $u_a$ makes the same decisions as an agent using $u_0$, except that she perceives all prospects to be shifted by $a$. As a result, the distribution of $x_{T_1+\Delta T}$ given $x_{T_1} = a$ for the agent using $u_a$ is the same as the distribution of $x_{T_2+\Delta T} + a$ given $x_{T_2} = 0$ for the agent using $u_0$.

Let $u_0(x_0) = u_0(0) + x_0 + \Delta u$. Assume $\Delta u < 0$. Consider an agent using shifted utility function $u_{x_0}$. Given $x_{T_1} = x_0$ the distribution of $x_{T_1+\Delta T}$ for the agent using $u_{x_0}$ is the same as the distribution of $x_{T_2+\Delta T} + x_0$ for the agent using $u_0$ and starting at $x_{T_2} = 0$. Hence $u_{x_0}(x_0) = u_0(0) + x_0 > u_0(x_0)$ which almost surely violates Lemma 1. Assume

$\Delta u > 0$. Consider an agent using $u_{-x_0}$. Given $x_{T_1} = 0$, the distribution of $x_{T_1+\Delta T}$ for this agent is the same as the distribution of $x_{T_2+\Delta T} - x_0$ for the agent using $u_0$ if $x_{T_2} = x_0$. Hence $u_{-x_0}(0) = u_0(x_0) - x_0 = u_0(0) + \Delta u > u_0(0)$ which almost surely violates Lemma 1. Thus, $\Delta u = 0$ and the in both cases (a) and (b) the baseline utility function is of the form $u_0(x) = a + x$.

Now, assume that there exists a better, possibly non-stationary decision policy. Since $u_0(x) = a + x$ is equivalent to risk-neutrality in every period, such a policy would have to give up risk-neutrality with positive probability. Therefore, such policy would have to either accept gambles with negative expected value or reject gambles with positive expected value. This obviously reduces fitness and so such a decision policy cannot exist. Hence, being almost surely risk-neutral is the optimal decision policy.

Since prospects are defined in terms of gains and losses with respect to the current state, the distribution of those gains and losses is invariant, and the preference between prospects does not depend on the current state. The expected gain between two consecutive states (given process has not terminated) is constant. Let denote this gain as $\Delta x - \bar{c}$.

$$u_0(0) = 0p_0 + (\Delta x - \bar{c})p_0(1 - p_0) + 2(\Delta x - \bar{c})p_0(1 - p_0)^2 + 3(\Delta x - \bar{c})p_0(1 - p_0)^3$$
$$+ \cdots = (\Delta x - \bar{c})(1 - p_0)\sum_{i=0}^{\infty} ip_0(1 - p_0)^{i-1} = (\Delta x - \bar{c})(1 - p_0)\frac{1}{p_0}$$

The average gain $\Delta x - \bar{c}$ can be separated into two components: average consumption $\bar{c} = Ec$ and average gain from choosing profitable gambles. Since the utility function is almost surely linear, the agent chooses a gamble if and only if $p(x - c + g) + (1 - p)(x - c - l) > x - c \Leftrightarrow pg - (1 - p)l > 0 \Leftrightarrow p > \frac{l}{g+l}$. Given $l$ and $g$, the expected gain is thus:

$$\int_{\frac{l}{l+g}}^{1} [pg + (1 - p)l]dp = \frac{g^2}{2(g + l)}$$

a) Assume that $l, g \sim U[0,2]$. Then

$$\Delta x = \int_0^2 \int_0^2 \frac{g^2}{2(g+l)} \frac{dg}{2} \frac{dl}{2} = \frac{1}{6}(\ln(16) - 1)$$

As a result, $u_0(0) = \left(\frac{1}{6}(\ln(16) - 1) - \bar{c}\right)(1 - p_0)\frac{1}{p_0} + x$.

b) Assume that $l, g \sim \text{Exp}(1)$. Then

$$\Delta x = \int_0^{+\infty} \int_0^{+\infty} \frac{g^2}{2(g+l)} \frac{dg}{e^g} \frac{dl}{e^l} = \frac{1}{3}$$

As a result, $u_0(0) = \left(\frac{1}{3} - \bar{c}\right)(1 - p_0)\frac{1}{p_0} + x$. ■

**Proof of Proposition 2.** The described stochastic environment satisfies conditions of Corollary 1. Thus, $u_0(x) = E\left[f\left(h^{\Phi}\right)\middle|x_1 = x\right] = P\left(x_\phi \geq 1|x_1 = x\right)$. Consider any $x_0 \in R$ and $\varepsilon > 0$ and let us assume that the current state is $x_0$. Assuming process does not terminate in the current period, we can partition the area where process terminates into three parts. Let $p_- = P\left(x_\phi < -\varepsilon|x_T = x_0\right)$, $p_\varepsilon = P\left(-\varepsilon \leq x_\phi < 0\middle|x_T = x_0\right)$, and $p_+ = P\left(0 \leq x_\phi\middle|x_T = x_0\right)$. Then $u_0(x_0) = p_0\mathbf{1}(x_0 \geq 0) + (1 - p_0)p_+$.

Consider a shifted utility function $u_\varepsilon(x) = u_0(x - \varepsilon)$. An agent using $u_\varepsilon$ makes the same decisions as an agent using $u_0$, except that she perceives all prospects to be shifted by $\varepsilon$. As a result, the distribution of $x_{T_1+\Delta T}$ given $x_{T_1} = \varepsilon$ for the agent using $u_\varepsilon$ is the same as the distribution of $x_{T_2+\Delta T} + \varepsilon$ given $x_{T_2} = 0$ for the agent using $u_0$.

In accordance with Lemma 1, $u_0(x_0 + \varepsilon) \geq P\left(0 \leq x_\phi\middle|x_T = x_0 + \varepsilon \wedge D = u_\varepsilon\right)$. Since $P\left(0 \leq x_\phi\middle|x_T = x_0 + \varepsilon \wedge D = u_\varepsilon\right) = p_0\mathbf{1}(x_0 + \varepsilon \geq 0) + (1 - p_0)(p_+ + p_\varepsilon) \geq p_0\mathbf{1}(x_0 \geq 0) + (1 - p_0)p_+ = u_0(x_0)$, we have $u_0(x_0 + \varepsilon) \geq u_0(x_0)$, that is $u_0(x)$ is a non-decreasing function of $x$.

Let us prove now that $\lim_{x\to+\infty} u_0(x) = 1$. Assume that the current state is $x_0 \gg 0$. $u_0(x_0)$ is the probability that the process terminates at $x_\phi \geq 0$ and $1 - u_0(x_0)$ is the probability that the process terminates at $x_\phi < 0$. Let us bound $1 - u_0(x_0)$ from above and show that this bound goes to zero as $x_0$ goes to plus infinity.

To bound $1 - u_0(x_0)$ from above, let us assume that the agent always chooses the gamble instead of secure state, always loses, and when her state reaches drops below zero, the process terminates immediately. Denote probability of $x_\phi < 0$ under such circumstances as $\tilde{p}$.

$$\begin{aligned}\tilde{p} = (1-p_0)P(l_1 + c_1 > x_0) + (1-p_0)^2 P(l_1 + l_2 + c_1 + c_2 > x_0) \\ + (1-p_0)^3 P(l_1 + l_2 + l_3 + c_1 + c_2 + c_3) + \cdots \\ = \sum_{i=1}^{+\infty} (1-p_0)^i P\left(\sum_{k=1}^{i} l_k + \sum_{k=1}^{i} c_k > x_0\right) \\ = \sum_{i=1}^{+\infty} (1-p_0)^i P\left(G_i^L + G_i^C > x_0\right)\end{aligned}$$

where $l_i \sim \mathrm{Exp}(1)$, and $c_i \sim \mathrm{Exp}\left(\frac{1}{\bar{c}}\right)$ are independently distributed for all $i$, $\sum_{k=1}^{i} l_k = G_i^L \sim \Gamma(i, 1)$, and $\sum_{k=1}^{i} c_k = G_i^C \sim \Gamma\left(i, \frac{1}{\bar{c}}\right)$. $\Gamma$ denotes the Gamma distribution.

Let us show that $\tilde{p}$ can be arbitrarily close to zero. Observe that $\forall n \geq 1, \forall \varepsilon > 0, \exists x_0 : \forall x > x_0, \forall i \leq n, P\left(G_i^L + G_i^C > x_0\right) < \varepsilon$. Hence

$$\begin{aligned}\tilde{p} = \sum_{i=1}^{n} (1-p_0)^i P\left(G_i^L + G_i^C > x_0\right) + \sum_{i=n+1}^{+\infty} (1-p_0)^i P\left(G_i^L + G_i^C > x_0\right) \\ \leq \sum_{i=1}^{n} (1-p_0)^i \varepsilon + \sum_{i=n+1}^{+\infty} (1-p_0)^i \\ = \varepsilon(1-p_0) \sum_{i=0}^{n-1} (1-p_0)^i + (1-p_0)^{n+1} \sum_{i=0}^{+\infty} (1-p_0)^i \\ = \varepsilon(1-p_0) \frac{1-(1-p_0)^n}{p_0} + \frac{(1-p_0)^{n+1}}{p_0}\end{aligned}$$

We can choose $\varepsilon$ and $n$ so that $\tilde{p}$ is arbitrarily close to zero. And since $\tilde{p} \geq 1 - u_0(x_0)$ and $u_0(x_0)$ is non-decreasing, we have $\lim_{x \to +\infty} u_0(x) = 1$.

It is possible to prove that $\lim_{x \to -\infty} u_0(x) = 0$ analogously by bounding $u_0(x_0)$ from above given $x_0 \ll 0$ and assuming that agent's consumption is zero, she always chooses to gamble and wins, and the process terminates as soon as the state reaches or exceeds zero.

Assume $u_0(x_0) = 0$ for some $x_0$. Then $u_0(x_0 - c) = u_0(x_0 - c - l) = 0$ for any $c$ and $l$, since $u_0$ is non-decreasing. However, since $\lim_{x \to +\infty} u_0(x) = 1$ and $g \sim \text{Exp}(1)$, $P(x_0 - c + g \geq 0) > 0$ which implies $u_0(x_0) > 0$. Now assume $u_0(x_0) = 1$ for some $x_0$. Let us say that for the next $2\left\lceil\frac{x_0}{\bar{c}}\right\rceil$ periods the agent gets $c \in (2\bar{c}.3\bar{c}), g \in (0, \bar{c}), l \in (0, \bar{c}), p \in [0,1]$ and then process terminates. There is a positive probability that this will happen and since in each period the state decreases by at least $\bar{c}$, the process must terminate at $x_\phi < 0$ with positive probability which contradicts $u_0(x_0) = 1$. This concludes proof of the part (a) of the proposition.

Now, let us prove that for any $T > 0$ the distribution of $x_T$ is such that $P(a < x_T < b) > 0$ for any $a < b$ and that $P(x_T = a) = 0$ for any $a$. In other words, $x_T$ is a continuous random variable with no mass points and the support spanning entire $R$.

Consider an arbitrary state $x_T$ and an arbitrary interval $(a, b)$. Because $\lim_{x \to -\infty} u_0(x) = 0$ and $u_0(x_T) > 0$, there must exist an interval $(u, v)$ such that $v < a$, $u_0(v) < u_0(a)$, $v < x_T$, and $P\big(x_T - c \in (u, v)\big) > 0$. For each $x_T - c \in (u, v)$ we have $P\big(x_T - c + g \in (a, b)\big) > 0$. Moreover, for each $x_T - c \in (u, v)$ and $x_T - c + g \in (a, b)$ we have $P\big(pu_0(x_T - c + g) + (1 - p)u_0(x_T - c - l) > u_0(x_T - c)\big) > 0$. The agent prefers to take the gamble for any $l$ and sufficiently large $p$, since $u_0(x_T - c + g) \geq u_0(a) > u_0(v) \geq u_0(x_T - c)$. As a result, $P\big(x_{T+1} \in (a, b)\big) > 0$ and by induction for any $T > 0$ and any $a < b$ we have $P(a < x_T < b) > 0$.

Consider any $x_T$. Since $c$, $p$, $l$, and $g$ have independent continuous distributions without mass points, $x_{T+1}$ also must have a continuous distribution without mass points. Thus, by induction, regardless of the distribution of $x_1$, for any $T > 1$, $x_T$ has a continuous distribution with no mass points.

Recall the argument that $u_0(x)$ is non-decreasing at the beginning of this proof. Recall that $p_\varepsilon = P\big(-\varepsilon \leq x_\phi < 0 \big| x_T = x_0\big)$. Since for any $T > 1$, $P(-\varepsilon \leq x_T < 0) > 0$, hence $p_\varepsilon > 0$. In turn $u_0(x_0 + \varepsilon) \geq P\big(0 \leq x_\phi \big| x_T = x_0 + \varepsilon \wedge D = u_\varepsilon\big) = p_0\mathbf{1}(x_0 + \varepsilon \geq 0) + (1 - p_0)(p_+ + p_\varepsilon) > p_0\mathbf{1}(x_0 \geq 0) + (1 - p_0)p_+ = u_0(x_0)$ that is $u_0(x_0)$ is strictly increasing. This concludes proof of the part (b) of the proposition.

Consider any $x_0$ and $\varepsilon > 0$ sufficiently small such that if $x_0 > 0$ then $x_0 - \varepsilon > 0$. Assuming process does not terminate in the current period, we can partition the area where it terminates into three parts. Let $p_- = P\big(x_\phi < 0 \big| x_T = x_0\big)$, $p_\varepsilon = P\big(0 \leq x_\phi < \varepsilon \big| x_T = x_0\big)$, and $p_+ = P\big(\varepsilon \leq x_\phi \big| x_T = x_0\big)$. Then $u_0(x_0) = p_0\mathbf{1}(x_0 \geq 0) + (1 - p_0)(p_\varepsilon + p_+)$.

Let us use a shifted utility function $u_{-\varepsilon}(x) = u_0(x + \varepsilon)$. The distribution of $x_\phi$ for agent using $u_{-\varepsilon}$ given $x_T = x_0 - \varepsilon$ is the same as the distribution of $x_\phi - \varepsilon$ for the agent using $u_0$ given $x_T = x_0$. Hence, $P\big(0 \leq x_\phi \big| x_T = x_0 - \varepsilon \wedge D = u_{-\varepsilon}\big) = p_0\mathbf{1}(x_0 - \varepsilon \geq 0) + (1 - p_0)p_+$. According to Lemma 1, $u_0(x - \varepsilon) \geq P\big(0 \leq x_\phi \big| x_T = x_0 - \varepsilon \wedge D = u_{-\varepsilon}\big)$. Thus

$$\begin{aligned} &u_0(x_0) - u_0(x - \varepsilon) \\ &\qquad\qquad \leq p_0\mathbf{1}(x_0 \geq 0) + (1 - p_0)(p_\varepsilon + p_+) - p_0\mathbf{1}(x_0 - \varepsilon \geq 0) - (1 - p_0)p_+ \\ &\qquad\qquad = p_0[\mathbf{1}(x_0 \geq 0) - \mathbf{1}(x_0 - \varepsilon \geq 0)] + (1 - p_0)p_\varepsilon \end{aligned}$$

If $x_0$ and $x_0 - \varepsilon$ have the same sign, that is if $x_0 \neq 0$, then $\mathbf{1}(x_0 \geq 0) = \mathbf{1}(x_0 - \varepsilon \geq 0)$ and $u_0(x_0) - u_0(x_0 - \varepsilon) \leq (1 - p_0)p_\varepsilon$. Since there are no mass points in the distribution of $x_\phi$ for $\phi > T$, we have $p_\varepsilon \to 0$ as $\varepsilon \to 0$. As a result, for any $x_0 \neq 0$, $\lim_{\varepsilon \to 0} u_0(x_0) - u_0(x_0 - \varepsilon) = 0$, that is $u_0(x_0)$ is continuous.

If $x_0$ and $x_0 - \varepsilon$ have opposite signs, that is if $x_0 = 0$, then $\mathbf{1}(x_0 \geq 0) = 1$ and $\mathbf{1}(x_0 - \varepsilon \geq 0) = 0$, and $u_0(0) - u_0(-\varepsilon) \leq p_0 + (1 - p_0)p_\varepsilon$. Thus $\lim_{\varepsilon \to 0} u_0(0) - u_0(-\varepsilon) \leq p_0$.

Consider now alternative partitioning. Let $p_- = P\big(x_\phi < -\varepsilon \big| x_T = -\varepsilon\big)$, $p_\varepsilon = P\big(-\varepsilon \leq x_\phi < 0 \big| x_T = -\varepsilon\big)$, and $p_+ = P\big(0 \leq x_\phi \big| x_T = -\varepsilon\big)$. Then, $u_0(-\varepsilon) = (1 -$

$p_0)p_+$. Let us use a shifted utility function $u_\varepsilon(x) = u_0(x-\varepsilon)$. $P\left(x_\phi \geq 0 \middle| x_T = 0 \wedge D = u_\varepsilon\right) = p_0 + (1-p_0)(p_\varepsilon + p_+)$. According to Lemma 1, $u_0(0) \geq P\left(x_\phi \geq 0 \middle| x_T = 0 \wedge D = u_\varepsilon\right)$. Thus, $u_0(0) - u_0(-\varepsilon) \geq p_0 + (1-p_0)p_\varepsilon$. As a result, $\lim_{\varepsilon\to 0} u_0(0) - u_0(-\varepsilon) \geq p_0$.

In summary, $p_0 \leq \lim_{\varepsilon\to 0} u_0(0) - u_0(-\varepsilon) \leq p_0$, and so $\lim_{\varepsilon\to 0} u_0(0) - u_0(-\varepsilon) = p_0$. This concludes proof of the part (c) of the proposition. ■

**References**

Anderson, P. (1972): “More is different,” *Science,* 177 (4047), 393-396.

Arrow, K. (1963): *Social choice and individual values*. New Haven: Yale University Press.

Borch, K. (1966): “A utility function derived from a survival game,” *Management science,* 12 (8), B287-B295.

Buss, D. (2016): *Evolutionary psychology: the new science of the mind*. New York: Routledge.

Caraco, T. (1980): “On foraging time allocation in a stochastic environment,” *Ecology,* 61 (1), 119-128.

Collins, J., B. Baer, and E. Weber (2016): “Evolutionary biology in economics: a review,” *Economic Record*, (92) 297, 291-312.

Cooper, W. (1987): “Decision theory as a branch of evolutionary theory: a biological derivation of the Savage axioms,” *Psychological review*, 94 (4), 395-411.

De Fraja, G. (2009) “The origin of utility: sexual selection and conspicuous consumption,” *Journal of economic behavior and organization*, 72 (1), 51-69.

Dubins, L., and L. Savage (1965): *How to gamble if you must: inequalities for stochastic processes,* New York: McGraw-Hill.

Feinberg, E., and A. Shwartz (2002): *Handbook of Markov decision processes: methods and applications,* Boston: Kluwer Academic Publishers.

Friedman, M., and L. Savage (1948): “The utility analysis of choices involving risk,” *The journal of political economy,* 56 (4), 279-304.

Grafen, A. (2002): “A first formal link between the price equation and an optimization program,” *Journal of theoretical biology*, 217 (1), 75-91.

Kacelnik, A. and M. Bateson (1996): "Risky theories: the effects of variance on foraging decisions," *American zoologist*, 36 (4), p. 402-434.

Kahneman, D., and A. Tversky (1979): "Prospect theory: an analysis of decision under risk," *Econometrica*, 47 (2), 263-292.

Karni, E., and D. Schmeidler (1986): "Self-preservation as a foundation of rational behavior under risk," *Journal of economic behavior and organization*, 7 (1), 71-81.

Kenrick, D., V. Griskevicius, J. Sundie, N. Li, Y. Li, and S. Neuberg (2009): "Deep rationality: the evolutionary economics of decision making," *Social cognition*, 27 (5), 764-785.

Pierce, B. (2012): *Genetics: a conceptual approach*. New York: W.H. Freeman and Company.

Puterman, L. (2005): *Markov decision processes: discrete stochastic dynamic programming*. Hoboken: John Wiley & Sons.

Rabin, M. (2000): "Risk aversion and expected utility theory: a calibration theorem," *Econometrica*, 68 (5), 1281-1292.

Rayo, L. and G. Becker (2007): "Evolutionary efficiency and happiness," *Journal of Political Economy*, 115 (2), 302-337.

Real, L. and T. Caraco (1986): "Risk and foraging in stochastic environments," *Annual review of ecology and systematics*, 17, 371-390.

Robson, A. (1996): "A biological basis for expected and non-expected utility," *Journal of economic theory*, 68 (2), 397-424.

Robson, A. (2001a): "The biological basis of economic behavior," *Journal of economic literature*, 39 (1), 11-33.

Robson, A. (2001b): "Why would nature give individuals utility functions?" *Journal of political economy*, 109 (4), 900-914.

Robson, A. and L. Samuelson (2011): The evolutionary foundations of preferences. In J. Benhabib, A. Bisin, and M. Jackson (Eds.), *Handbook of Social Economics*. San Diego: Elsevier.

Rubin, P. (1982): “Evolved ethics and efficient ethics,” *Journal of economic behavior and organization*, 3 (2-3), 161-174.

Rubin, P. and C. Paul (1979): “An evolutionary model of taste for risk,” *Economic inquiry*, 17 (4), 585-596.

Russell, S., and P. Norvig (2010): *Artificial intelligence: a modern approach*, Upper Saddle River: Prentice Hall.

Spall, J. (2003): *Introduction to stochastic search and optimization estimation, simulation, and control*. Hoboken: Wiley-Interscience.

Strzałko, J., J. Grabski, A. Stefański, P. Perlikowski, and T. Kapitaniak (2008): „Dynamics of coin tossing is predictable,” *Physics reports*, 469 (2), 59-92.

Von Neumann, J., and O. Morgenstern (1964): *Theory of games and economic behavior*. New York: Jon Wiley & Sons.

Wiessner, P. (2002): “Hunting, healing, and hxaro exchange: a long-term perspective on !Kung (Ju/’hoansi) large-game hunting,” *Evolution and Human Behavior*, 23 (6), 407-436.