When biases under risk are optimal under uncertainty and learning. Overestimation of low probabilities and status quo bias

Michel De Lara*

October 29, 2011

Abstract

In experimental economics, it is observed that individuals having to choose between loteries behave as if they overweigh small probabilities. Somewhat related is the observation that low probability-vivid consequences events are often given a weight that is much higher than their observed frequencies (like plane accidents, for instance). Our claim is that such biases may result from the observation that, under uncertainty, probabilities are not given but are sequentially learnt by trials. We show that overweighting of low probability events may be a byproduct of optimal sequential decision-making under uncertainty and learning. Our model extends an analysis made by Haselton and Nettle (*The paranoid optimist: An integrative evolutionary model of cognitive biases*, 2006) to the situation of dynamic learning exhibiting a multi-armed bandit structure. A decision maker (DM) does not know *a priori* the objective probability driving the ocurrence of a "bad" event, but learns it *a posteriori* by experience (trials). However, the DM can also decide to avoid being confronted possibly to such bad event, but at the price of an avoidance cost. We define a critical probability by the ratio between costs of encounter and costs of avoidance. We show that, for a rare event with probability lower than the critical probability, following an optimal strategy leads to either experimenting forever and rightly evaluating the unknown probability or experimenting followed by forever avoiding and overestimating the unknown probability. We conclude that behaviors considered as *biases under risk – overweighting of small probabilities* and *status quo bias* – are features of *optimal behavior under uncertainty and learning*. As far as minimizing mean discounted costs enhances fitness (better survival and reproduction), natural selection may have favored such biases in individuals.

*Université Paris–Est, CERMICS, 6–8 avenue Blaise Pascal, 77455 Marne la Vallée Cedex 2, France. delara@cermics.enpc.fr

1 Introduction

A debate opposes two conceptions in the "heuristics and biases" literature (Kahneman, Slovic, and Tversky, 1982; T. Gilovich, 2002): on the one hand, some behaviors are qualified of "bias" when they depart from given "rationality benchmarks" (like expected utility theory); on the other hand, some scholars (see (Gigerenzer, 2004, 2008; Hutchinson and Gigerenzer, 2005)) claim that those "so-called bias" were in fact advantageous in the type of environment where our ancestors lived and thrived (ecological validity). In this latter case, the benchmark should be a measure of fitness reflecting *survival and reproduction* (S & R) abilities.

An example of bias is the following. In experimental economics, it is observed that individuals having to choose between loteries behave as if they overweight small probabilities. A certain number of theories – Kahneman and Tversky'*prospect theory* (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992), Lopes'*security/potential and aspiration* theory (Lopes, 1996; Lopes and Oden, 1999), etc. – even produce curves of S-shaped probability deformations (see Figure 1). Somewhat related is the observation that low probability-vivid

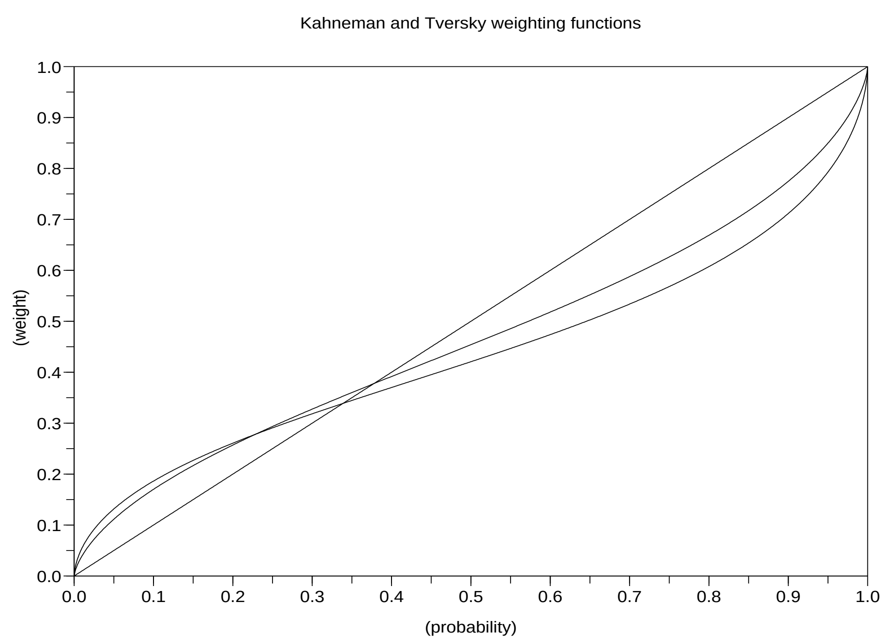

Figure 1: Kahneman and Tversky probability weighting functions (the two curves correspond to two domains, of losses and of gain). In the left hand side, the curves are above the identity diagonal, exhibiting overweighting of low probabilities.

consequences events – like a plane crash, or winning the jackpot in a casino – are often given a weight that is judged excessive with respect to "rationality".

We propose a model of decision under uncertainty and learning with a benchmark for which we prove that the optimal behavior has the following features: overweighting of low probability events, and status quo bias. This model is inspired by one given by Haselton and Nettle in (Haselton and Nettle, 2006). To reach his destination, an individual has two options: a short risky way passes through a grassy land possibly hiding a poisonous snake inflicting grave pains, while a safe way makes a long costly detour. The analysis in (Haselton and Nettle, 2006) is made supposing known the probability that a snake is in the grass. We propose an argument based on learning, where this latter probability is unknown but may be learnt by choosing the risky way.

A decision maker (DM) does not know *a priori* the objective probability driving the ocurrence of a "bad" event, but learns it by experience (trials). However, the DM can also decide to avoid being confronted possibly to such bad event, but at the price of an avoidance cost. We define a critical probability related to costs of encounter and avoidance. We show that, for a "bad" event with probability lower than the critical probability, following an optimal strategy leads to overestimate the unknown probability. The observation that the optimal strategy does not necessarily lead to rightly evaluate the unknown probability has already been made by economists (Rothschild, 1974; Easley and Kiefer, 1988; Brezzi and Lai, 2000), and is coined the *incomplete learning theorem*. Our contribution is exhibiting overestimation. In the evolutionary literature, the discussion bears more on the precision of Bayesian estimates of the unknown probability (Trimmer, Houston, Marshall, Mendl, Paul, and McNamara, 2011) than on optimality benchmarks and resulting optimal strategies.

In Sect. 2, we recall an example from (Haselton and Nettle, 2006) where the probability driving the ocurrence of a "bad" event is supposed to be known. In Sect. 3, we extend this example to the case of unknown probability, and we exhibit the structure of a strategy that minimizes mean discounted costs. Supposing that these latter costs are correlated with fitness (the lower the costs, the higher the fitness), we conclude in Sect. 4 that natural selection may have favored individuals which attribute to low probability events a weight larger than their probability. In addition, we show that the so-called *status quo bias* is a feature of the optimal strategy.

2 When Cro-Magnon saw the grass moving

In (Haselton and Nettle, 2006), the following situation is examined. Consider two possible states of the world, denoted by $\{\mathtt{s}, \overline{\mathtt{s}}\}$, that we materialize by $\mathtt{s} =$ "a snake is in the grass", and by $\overline{\mathtt{s}}$ the contrary. Now, suppose that the grass is moving. According to our belief in what makes the grass moving – either $\mathtt{S} =$ "a snake is believed to be in the grass", or the contrary $\overline{\mathtt{S}}$ – two decisions are possible. Supposing a snake in the grass leads to avoid (A) the place and make a (costly) detour. On the other hand, passing through ("experimenting" T) is less costly if the snake is not, but can be lethal if it is.

Paraphrasing (Haselton and Nettle, 2006), a belief can be adopted when it is in fact true (a true positive or TP), or it cannot be adopted and not be true (a true negative or TN). Then there are two possible errors. A false positive (FP) error occurs when a person adopts a belief that is not in fact true (believing there is a snake, when this is not the case), and a false negative (FN) occurs when a person fails to adopt a belief that is true. This is summarized in Table 1:

- assuming there is a snake ($\mathtt{S}$) induces avoidance (A) with cost cTP – like losing time in making a detour – let there indeed be a snake ($\mathtt{s}$, true positive TP) or not ($\overline{\mathtt{s}}$, false positive FP);
- assuming there is no snake ($\overline{\mathtt{S}}$) induces a cost cFN of encounter – being bitten, with painful and incapacitating consequences – if there indeed is a snake ($\mathtt{s}$, false negative FN) and a zero cost if not ($\overline{\mathtt{s}}$, true negative TN).

	snake ($\mathtt{s}$)	no snake ($\overline{\mathtt{s}}$)
SNAKE ($\mathtt{S}$)	cTP= cost of avoidance	cFP= cost of avoidance
NO SNAKE ($\overline{\mathtt{S}}$)	cFN= cost of encounter	cTN=0

Table 1: Costs according to decisions (rows) and states of the world (columns)

Now, assume that a probability is given on $\{\mathtt{s}, \overline{\mathtt{s}}\}$: a snake is in the grass with probability $p_{\mathtt{s}}$, and no snake is in the grass with probability $p_{\overline{\mathtt{s}}}$. If the above situation repeats itself independently, the empirical mean costs (over repetitions) are approximated by the theoretical expected costs by the law of large numbers:

- assuming there is a snake ($\mathtt{S}$) induces the same cost whatever the state of nature, which is the cost of avoidance cTP;
- assuming there is no snake ($\overline{\mathtt{S}}$) induces a cost of encounter cFN with probability $p_{\mathtt{s}}$ and a zero cost else, hence a mean cost $p_{\mathtt{s}} \times c$FN.

In the mean, it is better to believe $\mathtt{S}$ rather than $\overline{\mathtt{S}}$ if

$$\underbrace{p_{\mathtt{s}} c\mathrm{TP} + p_{\overline{\mathtt{s}}} c\mathrm{FP}}_{\text{expected costs assuming a snake in the grass}} < p_{\mathtt{s}} c\mathrm{FN} + p_{\overline{\mathtt{s}}} c\mathrm{TN} \; . \tag{1}$$

In other words, it is better to believe that there is a snake ($\mathtt{S}$) than not ($\overline{\mathtt{S}}$) if the probability $p_{\mathtt{s}}$ that a snake is in the grass is greater than the *critical probability* p_c, that we define as follows:

$$p_{\mathtt{s}} > p_c := \frac{\text{cost of avoidance}}{\text{cost of encounter}} = \frac{c\mathrm{TP} - c\mathrm{TN}}{c\mathrm{FN} - c\mathrm{TN}} \; . \tag{2}$$

As a consequence, the lower the cost of avoidance and the higher the cost of encounter, the lower the critical probability p_c hence the better to believe $\mathtt{S}$ rather than $\overline{\mathtt{S}}$. The conclusion of Haselton and Nettle is that, when errors are asymmetrical in cost,[1] there is a tendency to

[1] Such asymmetry in costs is manifest in the so-called *life-dinner* principle of Richard Dawkins – "The rabbit runs faster than the fox, because the rabbit is running for his life while the fox is only running for his dinner" – and can exert a strong selection pressure (Dawkins and Krebs, 1979).

favor false positive error (FP), that is, adopting a belief that is not in fact true (believing there is a snake, when this is not the case).

Notice that the optimal decision rule depends on two quantities attached to the situation: one is the probability $p_{\mathtt{s}}$ that a snake is in the grass; the other is the ratio p_c of two costs, which, being less than 1, we interpret as a probability. While the costs may be learnt by a single encounter, the probability $p_{\mathtt{s}}$ of encounter is usually learnt as a frequency resulting from multiple encounters. This is what we discuss in the next section.

3 A model of learning and optimal strategies

Now, to "implement" the rule avoid/take risk depending on beliefs, one needs to have an idea of the probability $p_{\mathtt{s}}$ that a snake is in the grass. This idea is acquired by experience, by learning. In fact, there is a mix of learning and of acting, and this situation is examplified in the famous "multi-armed bandit problem" (Gittins, 1979).

3.1 A model of learning

On the one hand, suppose the existence of two states of the world, either "bad" ($\mathtt{B}$) or "good" ($\mathtt{G}$). On the other hand, two decisions are possible, either "avoid" ($\mathcal{A}$) or "try"/ "experiment"/ "learn" ($\mathcal{T}$). To give an example, consider the "free rider" problem: suppose you take public transport, and your decisions are either cheating ("try") or pay your ticket ("avoid" cheating); "bad" state is when you encounter a controller[2] and "good" state when you do not.

Consider that time $t = 0, 1, 2\ldots$ is discrete. Define the history space by $\mathcal{H}_\infty := \{\mathtt{B}, \mathtt{G}\}^{\mathbb{N}*}$ with typical element an infinite sequence of elements in $\{\mathtt{B}, \mathtt{G}\}$. Denote by $X_{t+1} : \mathcal{H}_\infty \to \{\mathtt{B}, \mathtt{G}\}$ the state of nature at time $t = 0, 1, 2\ldots$ At each period t, the DM can either "try" (decision $\mathcal{T}$)– in which case the state of the world $\mathtt{B}$ or $\mathtt{G}$ is revealed and learnt – or "avoid" (decision $\mathcal{A}$) – in which case the DM has no information about the state of the world. Define the mapping $\Phi : \{\mathcal{T}, \mathcal{A}\} \times \{\mathtt{B}, \mathtt{G}\} \to \{\mathtt{B}, \mathtt{G}, \infty\}$ by $\Phi(\mathcal{T}, \mathtt{B}) = \mathtt{B}$, $\Phi(\mathcal{T}, \mathtt{G}) = \mathtt{G}$ and $\Phi(\mathcal{A}, \mathtt{B}) = \Phi(\mathcal{A}, \mathtt{G}) = \infty$. Thus, the observation at time $t = 0, 1, 2\ldots$ if the DM takes decision $v_t \in \{\mathcal{T}, \mathcal{A}\}$ is $Y_{t+1} := \Phi(v_t, X_{t+1})$, where ∞ corresponds to no information.

3.2 Payoffs

The payoffs depend both on the decision and on the state of the world as in Table 2.

We assume that the payoffs attached to the couple (action,state) in Table 2 are ranked as follows:

$$V(\mathtt{G}) > V(\mathcal{A}) > V(\mathtt{B}) \; . \tag{3}$$

The payoffs may be seen as opposites of costs; in that case, with obvious notations $V(\mathtt{B}) = -\mathcal{C}_{\mathtt{B}}$, $V(\mathtt{G}) = -\mathcal{C}_{\mathtt{G}}$ and $V(\mathcal{A}) = -\mathcal{C}_{\mathcal{A}}$. Going on with the "free rider" problem, the so-called

[2]Nothing "bad" in encountering a controller! This is just terminology.

	"bad" state $\mathtt{B}$	"good" state $\mathtt{G}$
avoid ($\mathcal{A}$)	avoidance payoff $V(\mathcal{A})$	avoidance payoff $V(\mathcal{A})$
try ($\mathcal{T}$)	encounter payoff $V(\mathtt{B})$	base payoff $V(\mathtt{G})$

Table 2: Payoffs according to decisions (rows "avoid" ($\mathcal{A}$) or "try" ($\mathcal{T}$)) and states of the world (columns "bad" $\mathtt{B}$ or "good" $\mathtt{G}$)

cost of avoidance is the price of a ticket, say $\mathcal{C}_{\mathcal{A}} = 1.70$ euros inner Paris. If you cheat and do not pay the ticket, the "good" cost is $\mathcal{C}_{\mathtt{G}} = 0$ and the "bad" cost is a fine of $\mathcal{C}_{\mathtt{B}} = 50$ euros. In an evolutionary interpretation, payoffs are measured in "fitness" unit.

3.3 Strategies

Our learning assumption is that the DM is not visionary: he cannot know the future in advance, neither can he know the state of the world if he decides to avoid. We allow the visionary to accumulate past observations; therefore the decision v_t at time can only be a function of $Y_1, \ldots, Y_t$ (the initial decision v_0 is taken without information).

Define the observation spaces at time $t = 0, 1, 2 \ldots$ by $\mathcal{O}_0 := \{\infty\}$ (no observation at initial time $t = 0$) and $\mathcal{O}_t := \{\mathtt{B}, \mathtt{G}, \infty\}^t$ for $t = 1, 2 \ldots$, with typical element a sequence of t observations. A *policy* at time t is a mapping $\mathcal{S}_t : \mathcal{O}_t \to \{\mathcal{T}, \mathcal{A}\}$. A policy tells the DM what is his next action in view of past observations. A *strategy* $\mathcal{S}$ is a sequence $\mathcal{S}_0, \mathcal{S}_1 \ldots$ of policies for all times t.

Given the sequence $X(\cdot) := (X_1, X_2, \ldots)$ of states of nature at time $t = 0, 1, 2 \ldots$, and given a strategy $\mathcal{S}$, decisions and observations are inductively given by

$$v_t = \mathcal{S}_t(Y_1, \ldots, Y_t) \in \{\mathcal{T}, \mathcal{A}\} \text{ and } Y_{t+1} = \Phi(X_{t+1}, v_t) \in \{\mathtt{B}, \mathtt{G}, \infty\} .$$

We suppose that the decision maker (DM) applying a strategy knows the three payoffs in (3) because he has experienced the three situations in Table 2. Indeed, once the bad event has been experienced by the DM, we suppose that he knows its cost; then, an avoidance option is looked after, and its cost is also known. Knowing the payoffs, a DM can evaluate his lifetime performance using strategy $\mathcal{S}$ by the discounted intertemporal payoff

$$J\big(\mathcal{S}, X(\cdot)\big) := \sum_{t=0}^{+\infty} \rho^t V\big(\Phi(v_t, X_{t+1})\big) . \tag{4}$$

The discount factor $\rho \in [0, 1[$ can be interpreted as the parameter in the Geometric distribution supposed to be followed by the DM's lifetime. Since the payoff (4) is contingent on the unknown history $X(\cdot) = (X_1, X_2, \ldots)$, it is practically impossible that a strategy performs better than another for all histories. In this context where costs/utilies are known to the DM, but no probabilities are assigned to "bad" and "good" events, how can we say that a strategy is "better" than another?

3.4 Mother Nature's benchmark

Suppose that Mother Nature

1. selects the probabilities $p_{\mathtt{B}}$ and $p_{\mathtt{G}}$ of "bad" and "good" events at random from the prior π_0, which is a beta distribution $\beta(n^0_{\mathtt{B}}, n^0_{\mathtt{G}}) \propto p_{\mathtt{B}}^{n^0_{\mathtt{B}}-1} p_{\mathtt{G}}^{n^0_{\mathtt{G}}-1}$ on the one-dimensional simplex $S_1 = \{p_{\mathtt{B}} \geq 0,\ p_{\mathtt{G}} \geq 0,\ p_{\mathtt{B}} + p_{\mathtt{G}} = 1\}$, where $n^0_{\mathtt{B}}$ and $n^0_{\mathtt{G}}$ are positive integers;

2. then lets $(X_1, X_2, \ldots)$ be a sequence of independent Bernoulli trials with $\{X_t = \mathtt{B}\}$ having probability $p_{\mathtt{B}}$ and $\{X_t = \mathtt{G}\}$ having probability $p_{\mathtt{G}}$.

This corresponds to the "universe" $\Omega = S_1 \times \{\mathtt{B}, \mathtt{G}\}^{\mathbb{N}^*}$, being equipped with the probability[3] $\mathbb{P} := \pi_0(dp_{\mathtt{B}}, dp_{\mathtt{G}}) \otimes [p_{\mathtt{B}}\delta_{\mathtt{B}} + p_{\mathtt{G}}\delta_{\mathtt{G}}]^{\otimes \mathbb{N}^*}$.

For the DM who lives in the "sub-universe" $\{(p_{\mathtt{B}}, p_{\mathtt{G}})\} \times \{\mathtt{B}, \mathtt{G}\}^{\mathbb{N}^*}$, the occurence of "bad" and "good" events looks like following a sequence of independent Bernoulli trials with probabilities $p_{\mathtt{B}}$ and $p_{\mathtt{G}}$ respectively. These probabilities are fixed, but are unknown to the DM. Only by deciding to "try" can the DM infer the state of the world, hence evaluate the frequencies of bad and good events.

We shall make the assumption that Mother Nature compares two strategies with respect to the benchmark $\mathbb{E}^{\mathbb{P}}\big[J\big(\mathcal{S}, X(\cdot)\big)\big]$. A strategy yields higher fitness than another if, in the mean with respect to probability $\mathbb{P}$, the payoff (4) is higher.

3.5 Optimal strategy

An optimal strategy solves

$$\max_{\mathcal{S}} \mathbb{E}^{\mathbb{P}}\big[J\big(\mathcal{S}, X(\cdot)\big)\big] \ . \tag{5}$$

It is well known that the optimum is achieved by a a so-called *Gittins index* strategy (Gittins, 1989; Berry and Fristedt, 1985; Bertsekas, 2000) as follows.

Let us define $N^{\mathtt{B}}_t := \sum_{s=1}^t \mathbf{1}_{\{Y_s = \mathtt{B}\}}$ the number of "bad" encounters up to time t, and the same for $N^{\mathtt{G}}_t$. Observe that $N^{\mathtt{B}}_t + N^{\mathtt{G}}_t \leq t$, with $N^{\mathtt{B}}_t + N^{\mathtt{G}}_t = t$ if and only if the DM never avoided before t. There exists a function $\mathcal{I} : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ (called an index), which depends only on the discount factor ρ and on the utility V, such that the following strategy is optimal:

- if $\mathcal{I}(n^0_{\mathtt{B}} + N^{\mathtt{B}}_t, n^0_{\mathtt{G}} + N^{\mathtt{G}}_t) > V(\mathcal{A})$, then select decision $\mathcal{T}$ (try);
- if $\mathcal{I}(n^0_{\mathtt{B}} + N^{\mathtt{B}}_t, n^0_{\mathtt{G}} + N^{\mathtt{G}}_t) < V(\mathcal{A})$, then select decision $\mathcal{A}$ (avoid);
- if $\mathcal{I}(n^0_{\mathtt{B}} + N^{\mathtt{B}}_t, n^0_{\mathtt{G}} + N^{\mathtt{G}}_t) = V(\mathcal{A})$, then select indifferently decision $\mathcal{T}$ or decision $\mathcal{A}$.

Recall that $n^0_{\mathtt{B}}$ and $n^0_{\mathtt{G}}$ are the positive integers in the prior beta distribution $\pi_0 = \beta(n^0_{\mathtt{B}}, n^0_{\mathtt{G}})$ on the one-dimensional simplex S_1.

[3] For example, the event GGGBG has the probability $\mathbb{P}\{\mathtt{GGGBG}\} = \int_{S_1} \pi_0(dp_{\mathtt{B}}, dp_{\mathtt{G}}) p_{\mathtt{B}} p_{\mathtt{G}}^4 = \frac{1}{B(n^0_{\mathtt{B}}, n^0_{\mathtt{G}})} \int_0^1 p^{n^0_{\mathtt{B}}-1}(1-p)^{n^0_{\mathtt{G}}-1} p(1-p)^4 dp$ where $B(n^0_{\mathtt{B}}, n^0_{\mathtt{G}})$ is the normalizing factor of the beta distribution.

3.6 Almost optimal strategy (AOS) and critical probability

In general, there is no analytic expression of the index function $\mathcal{I}$. However, it is established that $\mathcal{I}(N^{\mathtt{B}}, N^{\mathtt{G}}) \geq \frac{N^{\mathtt{B}}}{N^{\mathtt{B}}+N^{\mathtt{G}}} V(\mathtt{B}) + \frac{N^{\mathtt{G}}}{N^{\mathtt{B}}+N^{\mathtt{G}}} V(\mathtt{G})$. In (Chancelier, Lara, and de Palma, 2007), our numerical simulations reveal that the above optimal strategy can reasonably be approximated by the following simple one.

We define the *critical probability* by the following ratio of avoidance cost over "bad" encounter cost:

$$p_c := \frac{V(\mathtt{G}) - V(\mathcal{A})}{V(\mathtt{G}) - V(\mathtt{B})} = \frac{\text{cost of avoidance}}{\text{cost of encounter}} \in]0, 1[\ . \tag{6}$$

All things being equal, the "worse" a bad event (that is, high bad costs), the lower the critical probability. In the "free rider" problem, the critical probability is $p_c = \frac{1.70}{50} = 3.4\%$.

Let us define an estimator $\widehat{p}_{\mathtt{B}}(t)$ of the "bad" event probability by[4]

$$\widehat{p}_{\mathtt{B}}(t) := \frac{n^0_{\mathtt{B}} + N^{\mathtt{B}}_t}{n^0_{\mathtt{B}} + N^{\mathtt{B}}_t + n^0_{\mathtt{G}} + N^{\mathtt{G}}_t} = \frac{n^0_{\mathtt{B}} + \text{ number of ''bad'' encounters up to time } t}{n^0_{\mathtt{B}} + n^0_{\mathtt{G}} + \text{ number of tries up to time } t} \ . \tag{7}$$

Now, if we replace $\mathcal{I}(n^0_{\mathtt{B}} + N^{\mathtt{B}}_t, n^0_{\mathtt{G}} + N^{\mathtt{G}}_t)$ in the above optimal strategy by its lower approximation $\frac{n^0_{\mathtt{B}}+N^{\mathtt{B}}_t}{n^0_{\mathtt{B}}+N^{\mathtt{B}}_t+n^0_{\mathtt{G}}+N^{\mathtt{G}}_t} V(\mathtt{B}) + \frac{n^0_{\mathtt{G}}+N^{\mathtt{G}}_t}{n^0_{\mathtt{B}}+N^{\mathtt{B}}_t+n^0_{\mathtt{G}}+N^{\mathtt{G}}_t} V(\mathtt{G})$, we obtain the almost optimal strategy (AOS)

- select decision $\mathcal{T}$ and experiment when $\widehat{p}_{\mathtt{B}}(t) \leq p_c$;
- select decision $\mathcal{A}$ and avoid when $\widehat{p}_{\mathtt{B}}(t) > p_c$.

Thus, the DM experiments whenever the probability estimator (7) is less than the critical probability (6). In the sequel, we shall improperly label this simple strategy AOS as *optimal*, though it is not optimal, but because it is close to, and because the qualitative features of both rules are alike, so that our conclusions are not affected. Notice that, when the optimal DM avoids, so does the almost optimal one: thus, the AOS is more conservative than the optimal, in that it more often avoids.

3.7 Features of the AOS

Denote by τ the first time t, if it exists, where the probability estimator $\widehat{p}_{\mathtt{B}}(t)$ of the "bad" event exceeds the critical probability p_c:

$$\tau = \inf\{t = 1, 2 \ldots \mid \widehat{p}_{\mathtt{B}}(t) > p_c\} \ . \tag{8}$$

In case, $\widehat{p}_{\mathtt{B}}(t) \leq p_c$ for all times $t = 1, 2 \ldots$, the convention is $\tau = +\infty$. Now, the AOS leads to the following behavior:

- "experiment" from times $t = 0$ up to τ[5], that is as long as $\widehat{p}_{\mathtt{B}}(t) \leq p_c$,

[4] It can easily be shown that $\widehat{p}_{\mathtt{B}}(t)$ $\mathbb{P}$-almost surely converges towards the random variable $p_{\mathtt{B}}$ on Ω.

[5] Notice that, when $\tau = +\infty$, the DM indefinitely experiments. Of course, this is an ideal situation made possible because the model assumes that the life of the DM follows a Geometric distribution.

- "avoid" for all periods after time $t = \tau + 1$, that is as soon as $\widehat{p}_{\mathtt{B}}(t) > p_c$ (indeed, once the DM avoids, he no longer updates the probability estimator (7) of the "bad" event, so that he keeps avoiding).

Additionaly to providing almost optimal performance, the AOS also leads to an estimate of the unknown probability $p_{\mathtt{B}}$ as follows (see also Table 3).

- If $\tau < +\infty$ (that is, if there exists a time t where the probability estimator $\widehat{p}_{\mathtt{B}}(t)$ exceeds the critical probability p_c), at the first period of avoidance $t = \tau$, the DM estimates the unknown probability $p_{\mathtt{B}}$ of the "bad" event by the probability estimator $\widehat{p}_{\mathtt{B}}(\tau) > p_c$. This situation occurs with probability $\mathbb{P}\{\exists t, \widehat{p}_{\mathtt{B}}(t) \geq p_c\}$.

- If $\tau = +\infty$ (that is, if the probability estimator $\widehat{p}_{\mathtt{B}}(t)$ never exceeds the critical probability p_c), experiment goes on forever and the estimate of $p_{\mathtt{B}}$ is given by the limit $\lim_{t\to+\infty} \widehat{p}_{\mathtt{B}}(t) = p_{\mathtt{B}}$ by the law of large numbers[6]. This situation occurs with probability $\mathbb{P}\{\forall t, \widehat{p}_{\mathtt{B}}(t) < p_c\}$.

Thus, when experiment goes on forever, the DM learns (asymptotically) the "true" and unknown $p_{\mathtt{B}}$. On the other hand, once avoidance occurs, $p_{\mathtt{B}}$ is estimated by the empirical frequency $\widehat{p}_{\mathtt{B}}(\tau) > p_c$, which is greater than p_c by the very property of the AOS.

	"avoid" $\widehat{p}_{\mathtt{B}}(0) > p_c$	"experiment" then "avoid" $\exists t = 1, 2 \ldots, \widehat{p}_{\mathtt{B}}(t) > p_c$	always "experiment" $\forall t, \widehat{p}_{\mathtt{B}}(t) \leq p_c$
estimate of $p_{\mathtt{B}}$	$\widehat{p}_{\mathtt{B}}(0) = \frac{n^0_{\mathtt{B}}}{n^0_{\mathtt{B}}+n^0_{\mathtt{G}}}$	$\widehat{p}_{\mathtt{B}}(\tau)$	$p_{\mathtt{B}}$
(comment)	($\widehat{p}_{\mathtt{B}}(0) > p_c$)	($\widehat{p}_{\mathtt{B}}(\tau) > p_c$)	(exact)

Table 3: Estimates of the probability $p_{\mathtt{B}}$ of the "bad" event depending on how long lasts the learning phase

On the subset $\{p_{\mathtt{B}} > p_c\}$ of Ω, an infinite experiment phase has zero probability by the law of large numbers. Therefore, avoidance occurs with probability one. This is intuitive: if the "bad" event has high probability and high costs with respect to avoidance costs (hence low critical probability), better avoid. Notice that the probability of a high probability-high consequences event is systematically estimated by the critical probability p_c (see the first two rows of Table 3). Since both $p_{\mathtt{B}}$ and the estimate are above p_c, we cannot conclude on over or under estimation.

On the subset $\{p_{\mathtt{B}} \leq p_c\}$ of Ω, an infinite experiment is possible, as well as experiment followed by avoidance. Therefore, Table 3 shows that rare events (those for which their probability is less than the critical probability) are either rightly evaluated or over-estimated by the critical probability p_c. Notice that the overestimation depends upon the payoffs (costs); this could be tested in experiments.

[6] By this, me mean that $\mathbb{P}\{\tau = +\infty \text{ and } \lim_{t\to+\infty} \widehat{p}_{\mathtt{B}}(t) = p_{\mathtt{B}}\} = \mathbb{P}\{\tau = +\infty\}$.

4 Learning, overestimation of low probabilities and status quo bias

From the above analysis, we now bring out qualitative features of the simple strategy described in Sect. 3, which provide insights into which attitudes natural selection may have favored in humans having to decide under uncertainty and learning. These insights send us back to the introductory recall on the debate about benchmarks for qualifying biases. We suppose that, if a strategy provides higher mean discounted payoff (lower mean discounted costs) than another, it provides higher fitness (better survival and reproduction).

4.1 A selective advantage to overweigh the probability of low probability events

Say that an event is rare if it occurs with a probability $p_{\mathtt{B}}$ lower than its critical probability p_c: $p_{\mathtt{B}} < p_c$. Natural selection has favored the following attitude in front of low probability events: attribute them a weight larger than their probability. Indeed, an individual adopting a strategy maximizing fitness exhibits

- either an infinite experiment phase, leading to learn the *exact value* $p_{\mathtt{B}}$,
- or a possible experiment phase followed, ending in learning the *over-estimate* $\widehat{p}_{\mathtt{B}}(\tau) \geq p_c > p_{\mathtt{B}}$.

As a consequence, low (loss) probabilities are either justly evaluated or are over-weighted.[7]

4.2 An argument in favor of the status quos bias

In addition to favoring overweighting the probability of low probability events, the above formal analysis provides theoretical support to the so-called *status quo bias* (Samuelson and Zeckhauser, 1988). Indeed, first observe that the lower the cost of avoidance and the higher the cost of a "bad" encounter, the lower the critical probability p_c and the less tendency to experiment; this is quite intuitive. Second, the optimal rule states that, once the DM selects the "avoid" option, he will never more experiment. Thus, once stuck in a risk avoidance attitude, there is no benefit in experimenting.

5 Conclusion

We have shown that overestimation of low probability events may be a byproduct of optimal strategies with respect to an intertemporal evolutionary benchmark in the face of uncertainty and learning (and not risk). Identifying the benchmark with fitness, this may explain that

[7]Symmetrically, common good events are either justly evaluated or are under-weighted.

natural selection has favored humans to attribute to rare events an importance larger than their probability.

Our formal analysis also provides theoretical support to the *availability bias* for vivid events (Kahneman and Tversky, 1974). The hypothesis that probabilities of vivid events are overweighted because such events are better recalled now becomes: vivid events are better recalled in order to give them more weight in decision-making because this is optimal from dynamic decision under uncertainty and learning point of view.

There are many situations where probabilities are not known but learnt. The equity premium puzzle comes from the observation that bonds are underweighted in portfolios, despite the empirical fact that stocks have outperformed bonds over the last century in the USA by a large margin (Mehra and Prescott, 1985). However, this analysis is done *ex post* under risk, while decision-makers take their decisions day by day under uncertainty, and sequentially learn about the probability of stocks loss. *Ex ante*, the overweighting of sure bonds might possibly be explained by optimal strategy under uncertainty and learning.

Acknowledgments. The author thanks Daniel Nettle, Newcastle University, and Jean-Marc Tallon, Université Paris 1, for their comments.

References

D. A. Berry and B. Fristedt. *Bandit problems: sequential allocation of experiments*. Chapman and Hall, 1985.

D. P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, Belmont, Massachusets, second edition, 2000. Volumes 1 and 2.

Monica Brezzi and Tze Leung Lai. Incomplete learning from endogenous data in dynamic allocation. *Econometrica*, 68(6):1511–1516, 2000. ISSN 1468-0262.

J.-P. Chancelier, M. De Lara, and A. de Palma. Risk aversion, road choice and the one-armed bandit problem. *Transportation Science*, 41(1):1–14, February 2007.

Richard Dawkins and J. R. Krebs. Arms races between and within species. *Proceedings of the Royal Society of London, Series* B, 205:489–511, 1979.

David Easley and Nicholas M Kiefer. Controlling a stochastic process with unknown parameters. *Econometrica*, 56(5):1045–64, September 1988.

Gerd Gigerenzer. Fast and frugal heuristics: The tools of bounded rationality. In Derek J. Koehler and Nigel Harvey, editors, *Blackwell handbook of judgement and decision making*, pages 62–88. Blackwell Publishing, Oxford, 2004.

Gerd Gigerenzer. Why heuristics work. *Perspectives on psychological science*, 3(1):20–29, 2008.

J. C. Gittins. Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society. Series B*, 41(2):148–177, 1979.

J. C. Gittins. *Multi-armed Bandit Allocation Indices*. Wiley, New York, 1989.

M. G. Haselton and D. Nettle. The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review*, 10(1):47–66, 2006.

John M. C. Hutchinson and Gerd Gigerenzer. Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet. *Behavioural Processes*, 69(2):97–124, 2005.

D. Kahneman and A. Tversky. Judgment under uncertainty: heuristics and biases. *Science*, 185:1124–1131, 1974.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292, 1979.

Daniel Kahneman, Paul Slovic, and Amos Tversky. *Judgment under Uncertainty : Heuristics and Biases*. Cambridge University Press, April 1982.

Lola L. Lopes. When time is of the essence: Averaging, aspiration, and the short run. *Organizational Behavior and Human Decision Processes*, 65(3):179–189, March 1996.

Lola L. Lopes and Gregg C. Oden. The role of aspiration level in risky choice: a comparison of cumulative prospect theory and SP/A theory. *J. Math. Psychol.*, 43(2):286–313, 1999.

Rajnish Mehra and Edward C. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics*, 15(2):145 – 161, 1985.

Michael Rothschild. A two-armed bandit theory of market pricing. *Journal of Economic Theory*, 9(2):185–202, October 1974.

William Samuelson and Richard Zeckhauser. Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1):7–59, March 1988.

D. Kahneman T. Gilovich, D. Griffin, editor. *Heuristics and Biases. The Psychology of Intuitive Judgement*. Cambridge University Press, 2002.

Pete Trimmer, Alasdair Houston, James Marshall, Mike Mendl, Elizabeth Paul, and John McNamara. Decision-making under uncertainty: biases and bayesians. *Animal Cognition*, 14:465–476, 2011.

Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, October 1992.