Rationally Biased Learning

Michel De Lara
Cermics, École des Ponts ParisTech, France

July 4, 2017

Outline of the presentation

Humans display myriads of biases. Why?

When one saw the grass moving

When one saw the grass moving day after day

Killing three biases with one stone

Discussion: when optimal learning induces biases

Outline of the presentation

Humans display myriads of biases. Why?

When one saw the grass moving

When one saw the grass moving day after day

Killing three biases with one stone

Discussion: when optimal learning induces biases

Young children fear wild animals more than cars

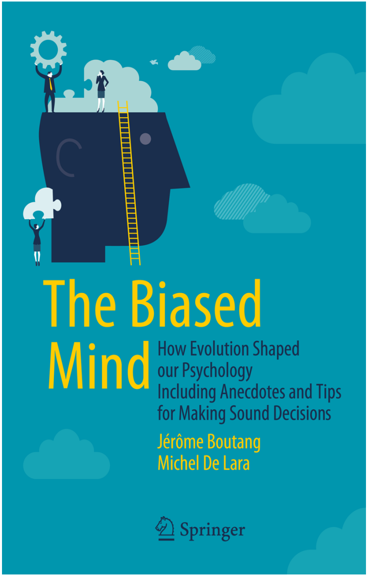
- Psychologist Adah Maurer's studied Chicago children fears
- almost all the 5- and 6-year-olds schoolchildren mentioned wild animals (most frequently snakes, lions, and tigers) in response to the question "What are the things to be afraid of?"
- "they do not (...) fear the things they have been taught to be careful about", say electric socket or cars

Humans display myriads of biases

- Children host innate preferences for savanna-type landscape
- We perceive sounds with increasing intensity
  as closer than they really are
- Men overperceive women neutral signals as sexual advances
- We all underestimate how long it takes to finish any project

- We are reluctant to change (status quo bias)
- We overweight the importance of salient events
- We overestimate low probabilities

A debate in the “heuristics and biases” literature opposes axiomatic and evolutionary validity



- Some behaviors are qualified of “bias” when they depart from given “rationality benchmarks” (like expected utility theory)
- Some suggest that those “so-called bias” were in fact advantageous in the type of environment where our ancestors lived and thrived (evolutionary validity)
- In this latter case, the benchmark should be a measure of fitness reflecting *survival and reproduction* (S & R) abilities

We will kill three birds (biases) with one stone (optimization with learning)

We propose a model of

- sequential optimization
- under uncertainty and learning

and we prove that the optimal behavior displays the following features of

- status quo bias
- salience effect
- overestimation of the probability of outcomes that are bad, unlikely and costly to avoid

Outline of the presentation

Humans display myriads of biases. Why?

When one saw the grass moving

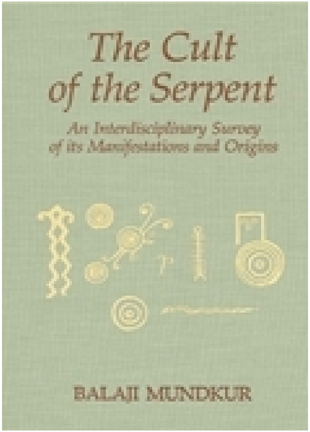
When one saw the grass moving day after day

Killing three biases with one stone

Discussion: when optimal learning induces biases

The story takes place in the savana landscape

Sinuous snakes haunt our minds



We start from a model of "paranoid optimist" by M. G. Haselton and D. Nettle

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

- Consider two possible states of Nature, denoted by $\{\mathtt{B}, \mathtt{G}\}$, that we materialize by
  - B: "a snake is in the grass" ("bad")
  - G: the contrary ("good")
- Consider two possible decisions, denoted by $\{\alpha, \varepsilon\}$, that we materialize by
  - α: avoid to cross the grass (and make a long detour)
  - ε: experiment/try to cross the grass

"Error management theory" amounts to minimize expected costs

- Now, assume that a probability is given on the states of Nature $\{\mathtt{B}, \mathtt{G}\}$:
  - a snake is in the grass with probability $p^{\mathtt{B}}$
  - and no snake is in the grass with probability $p^{\mathtt{G}}$
- In the mean, it is better to avoid than to experiment/try to cross the grass whenever

$$\underbrace{\mathcal{C}_\alpha}_{\text{cost of avoidance}} \quad < \quad \underbrace{p^{\mathtt{B}}\mathcal{C}^{\mathtt{B}} + p^{\mathtt{G}}\mathcal{C}^{\mathtt{G}}}_{\text{expected costs of experiment}}$$

Avoiding is optimal when the probability of the bad outcome exceeds a critical probability p^{c}

To minimize mean costs, better avoid crossing whenever

- the probability p^{B} that a snake is in the grass
- is greater than the critical probability p_c (avoidability index)

Optimal static decision rule

$$\text{avoid} \iff p^{\mathrm{B}} > p_c = \frac{\text{relative cost of avoidance}}{\text{relative cost of encounter}} = \frac{\mathcal{C}_\alpha - \mathcal{C}^{\mathtt{G}}}{\mathcal{C}^{\mathrm{B}} - \mathcal{C}^{\mathtt{G}}}$$

The higher is the cost of encounter
(in comparison to the cost to avoid), the more one avoids

The Haselton and Nettle model highlights the role of asymmetry in costs in biasing towards prudent behavior

- The conclusion of Haselton and Nettle is that
  - when errors are asymmetrical in cost
  - there is a tendency to favor false positive error (FP)
  - that is, to adopt a belief that is not in fact true (believing there is a snake, when this is not the case)
  - as in fire detectors biased towards false alarm (better evacuating a building in case of alarm, even false, than risking life by staying if detection fails)
- Asymmetries in cost have certainly tuned the human mind as a fire detector, explaining human penchants as diverse as: navigation bias, sound perception bias, disgust, agency bias, hostile attribution bias

Better treat a stick as a snake than the reverse!

On rabbits, foxes and snakes

- Joseph LeDoux, *The Emotional Brain*

  It is better to have treated a stick as a snake
  than not to have responded to a possible snake

On rabbits, foxes and snakes

- Joseph LeDoux, *The Emotional Brain*

  It is better to have treated a stick as a snake
  than not to have responded to a possible snake
- Richard Dawkins' so-called life-dinner principle

  The rabbit runs faster than the fox,
  because the rabbit is running for his life
  while the fox is only running for his dinner

Summary and roadmap

- The Haselton and Nettle model displays the following features
  - You have the choice between a safe (avoid) and a risky option (experiment/try)
  - You know the cost of the safe option (avoidance cost)
  - The risky option displays two outcomes, good and bad, and you know
    - the costs of the two outcomes
    - the probabilities of the two outcomes
- The conclusion: avoid if the probability of the bad outcome exceeds the costs ratio (of relative avoidance costs over relative bad outcome costs)
- Now, what happens when the probabilities are not known in advance?
- Answer: you can learn by observing frequencies, but you learn only by not avoiding, hence by taking risk

Outline of the presentation

Humans display myriads of biases. Why?

When one saw the grass moving

When one saw the grass moving day after day

Killing three biases with one stone

Discussion: when optimal learning induces biases

The "armed bandit problem" illustrates how acting affects learning and vice versa

- To implement the "critical probability rule" $\text{avoid} \iff p^{\mathrm{B}} > p_c$
- one needs to gauge the probability p^{B} that a snake is in the grass
- This estimate can be acquired by experience, by learning

Now, we emphasize that decisions are made sequentially

- We consider that decisions are made at discrete times $t = 0, 1, 2 \ldots$
- We define the history space $\mathbb{H}_\infty = \{\mathtt{B}, \mathtt{G}\}^{\mathbb{N}*}$ made of infinite sequences $\mathtt{GGGBG}\ldots$
- The state of Nature ($\mathtt{B}$ or $\mathtt{G}$) occuring at time $t = 0, 1, 2 \ldots$ is denoted by
$$X_{t+1} : \mathbb{H}_\infty \to \{\mathtt{B}, \mathtt{G}\}$$
and it is revealed at time $t + 1$ (hence the notation X_{t+1})
- At each time t, the decision-maker (DM) can
  - either "experiment/try" (decision ε), in which case the state of Nature ($\mathtt{B}$ or $\mathtt{G}$) is revealed and learned (hence available at $t + 1$ for informed decision-making)
  - or "avoid" (decision α), in which case the DM has no information about the state of Nature

Information about the current state of Nature depends on which decision is made

➡ Define the mapping $\mathcal{O} : \{\varepsilon, \alpha\} \times \{\mathtt{B}, \mathtt{G}\} \to \{\mathtt{B}, \mathtt{G}, \emptyset\}$

	"bad" ($\mathtt{B}$)	"good" ($\mathtt{G}$)
Avoid	no information ($\emptyset$)	no information ($\emptyset$)
Experiment/Try	"bad" ($\mathtt{B}$) is observed	"good" ($\mathtt{G}$) is observed

Table: Information according to decisions (rows) and states of Nature (columns)

➡ Thus, when the DM makes decision $v_t \in \{\varepsilon, \alpha\}$ at time t, at time $t+1$ he observes

$$Y_{t+1} = \mathcal{O}(v_t, X_{t+1})$$

Strategies depend upon observations

- Define the observation spaces at time $t = 0, 1, 2 \ldots$ by
  - $\mathbb{O}_0 = \{\emptyset\}$ (no observation at initial time $t = 0$)
  - $\mathbb{O}_t = \{\mathtt{B}, \mathtt{G}, \emptyset\}^t$ for $t = 1, 2 \ldots,$
    with typical element a sequence of t past observations
- A policy at time t is a mapping $\mathcal{S}_t : \mathbb{O}_t \to \{\varepsilon, \alpha\}$, that 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
- 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

$$\begin{array}{lrcl} \text{acting} & v_t = \mathcal{S}_t(Y_1, \ldots, Y_t) & & \in \{\varepsilon, \alpha\} \\ \text{learning} & Y_{t+1} = \mathcal{O}(v_t, X_{t+1}) & & \in \{\mathtt{B}, \mathtt{G}, \emptyset\} \end{array}$$

Humans display myriads of biases. Why?

When one saw the grass moving

When one saw the grass moving day after day

- A model of learning
- A fitness-based optimization problem
- The Gittins index optimal strategy (GIOS)

Killing three biases with one stone

- When learning stops, GIOS induces a status quo bias
- When learning stops, GIOS induces a salience bias
- When learning stops, GIOS induces probability overestimation

Discussion: when optimal learning induces biases

Payoffs depend on current decision and state of Nature

	"bad" state $\mathtt{B}$	"good" state $\mathtt{G}$
avoid (α)	avoidance $U(\alpha, \mathtt{B}) = \mathcal{U}_\alpha$	avoidance $U(\alpha, \mathtt{G}) = \mathcal{U}_\alpha$
experiment/try (ε)	encounter $U(\varepsilon, \mathtt{B}) = \mathcal{U}^{\mathtt{B}}$	$U(\varepsilon, \mathtt{G}) = \mathcal{U}^{\mathtt{G}}$

Table: Payoffs according to decisions
(rows "avoid" (α) or "experiment/try" (ε))
and states of Nature (columns "bad" $\mathtt{B}$ or "good" $\mathtt{G}$)

- The payoffs are ranked as

$$\underbrace{\mathcal{U}^{\mathtt{G}}}_{\text{base payoff}} > \overbrace{\mathcal{U}_\alpha}^{\text{avoidance payoff}} > \underbrace{\mathcal{U}^{\mathtt{B}}}_{\text{encounter payoff}}$$

- In an evolutionary interpretation,
payoffs are measured in "fitness"
(supposed to be cumulative, as an offspring stock)

We suppose that the DM maximizes discounted intertemporal payoff

We suppose that the DM

- knows the matrix $U(\cdot, \cdot)$ of payoffs
- evaluates his lifetime performance using strategy $\mathcal{S}$ by the discounted intertemporal payoff

$$J(\mathcal{S}, X(\cdot)) = \sum_{t=0}^{+\infty} \rho^t U(v_t, X_{t+1}) \ ,$$

where

$$\begin{aligned} v_t &= \mathcal{S}_t(Y_1, \ldots, Y_t) && \in \{\varepsilon, \alpha\} \\ Y_{t+1} &= \mathcal{O}(v_t, X_{t+1}) && \in \{\mathtt{B}, \mathtt{G}, \emptyset\} \end{aligned}$$

We interpret discounting till infinity as no discounting till Geometric distributed lifetime

- Let θ denote a random variable having geometric distribution

$$\mathbb{P}(\theta \geq t) = \rho^{t-1} \ , \quad t = 1, 2, 3 \ldots$$

- or, equivalently,

$$\mathbb{P}(\theta = t) = (1-\rho)\rho^{t-1} \ , \quad t = 1, 2, 3 \ldots$$

- Then, we have that

$$\underbrace{\mathbb{E}\Big[\sum_{t=0}^{\theta-1} U(v_t, X_{t+1})\Big]}_{\text{no discounting}} = \underbrace{\sum_{t=0}^{+\infty} \rho^t U(v_t, X_{t+1})}_{\text{discounting}}$$

We interpret the discount factor $\rho \in]0, 1[$ in term of mean lifetime of the DM

	discount factor	discount rate	mean time
discount factor	ρ	$r_e = \frac{1-\rho}{\rho}$	$\overline{\theta} = \frac{1}{1-\rho}$
discount rate	$\rho = \frac{1}{1+r_e}$	r_e	$\overline{\theta} = \frac{1+r_e}{r_e}$
mean time	$\rho = \frac{\overline{\theta}-1}{\overline{\theta}}$	$r_e = \frac{1}{\overline{\theta}-1}$	$\overline{\theta}$

discount factor ρ	discount rate r_e	mean time $\overline{\theta}$	survival $\mathbb{P}(\theta \geq \overline{\theta})$
0.9999	0.01%	10,000	0.368
0.9900	1%	100	0.370
0.9520	5%	21	0.377
0.9090	10%	11	0.386

We suppose that "bad" and "good" outcomes follow a Bernoulli process with probability $\overline{\mathbb{P}}$ on $\mathbb{H}_\infty$

- We suppose that Nature
  1. has selected probabilities $\overline{p}^{\mathtt{B}}$ and $\overline{p}^{\mathtt{G}}$ of "bad" and "good" outcomes
  2. then has let $(X_1, X_2, \ldots)$ be a sequence of independent Bernoulli trials
     - with the event $\{X_t = \mathtt{B}\}$ having probability $\overline{p}^{\mathtt{B}}$
     - with the event $\{X_t = \mathtt{G}\}$ having probability $\overline{p}^{\mathtt{G}}$

  thus yielding a (true) probability $\overline{\mathbb{P}}$ on the history space $\mathbb{H}_\infty$
- We suppose that the DM does *not* know the probabilities $\overline{p}^{\mathtt{B}}$ and $\overline{p}^{\mathtt{G}}$ (that define $\overline{\mathbb{P}}$)

We suppose that the DM has a prior $\mathbb{P}_{\pi_0}$ on $\mathbb{H}_\infty$

The DM makes the assumption that Nature

1. has selected the probabilities $p^{\mathtt{B}}$ and $p^{\mathtt{G}}$ of "bad" and "good" outcomes at random from the prior π_0, which is a distribution on the simplex $S_1 = \{p^{\mathtt{B}} \geq 0,\ p^{\mathtt{G}} \geq 0,\ p^{\mathtt{B}} + p^{\mathtt{G}} = 1\}$
2. then has let $(X_1, X_2, \dots)$ be a sequence of independent Bernoulli trials
   - with the event $\{X_t = \mathtt{B}\}$ having probability $p^{\mathtt{B}}$
   - with the event $\{X_t = \mathtt{G}\}$ having probability $p^{\mathtt{G}}$
3. so that the extended sample space $S_1 \times \mathbb{H}_\infty = S_1 \times \{\mathtt{B}, \mathtt{G}\}^{\mathbb{N}^*}$ is equipped with the probability $\pi_0(dp^{\mathtt{B}} dp^{\mathtt{G}}) \otimes \big(p^{\mathtt{B}}\delta^{\mathtt{B}} + p^{\mathtt{G}}\delta^{\mathtt{G}}\big)^{\otimes \mathbb{N}^*}$, whose marginal distribution on $\mathbb{H}_\infty$ we denote by $\mathbb{P}_{\pi_0}$

The DM can now formulate a discounted expected optimization problem

➩ We look for an optimal strategy $\mathcal{S}^\star$, solution of

$$\mathbb{E}^{\mathbb{P}_{\pi_0}}\left[J\big(\mathcal{S}^\star, X(\cdot)\big)\right] = \max_{\mathcal{S}} \mathbb{E}^{\mathbb{P}_{\pi_0}}\left[J\big(\mathcal{S}, X(\cdot)\big)\right]$$

➩ What are the features of an optimal strategy?

An optimal strategy can be searched for among state feedbacks

- Let $\Delta(S^1)$ denote the set of probability distributions on the simplex S^1
- It is well-known that an optimal strategy can be searched for among state feedbacks of the form

$$\mathcal{S}_t(Y_1, \dots, Y_t) = \widehat{\mathcal{S}}_t(\widehat{\pi}_t)$$

- with a state feedback

$$\widehat{\mathcal{S}}_t : \Delta(S^1) \to \{\varepsilon, \alpha\}$$

- and the information state $\widehat{\pi}_t \in \Delta(S^1)$, the conditional distribution, with respect to $Y_1, \dots, Y_t$ of the first coordinate mapping on $S_1 \times \mathbb{H}_\infty$

We introduce a relevant information state, the posterior

- The dynamics of the posterior $\widehat{\pi}_t \in \Delta(S^1)$ is given by

$$\widehat{\pi}_0 = \pi_0$$

and

$$\widehat{\pi}_{t+1} = \begin{cases} \widehat{\pi}_t & \text{if } Y_{t+1} = \emptyset \\ \theta^{\mathtt{B}}\widehat{\pi}_t & \text{if } Y_{t+1} = \mathtt{B} \\ \theta^{\mathtt{G}}\widehat{\pi}_t & \text{if } Y_{t+1} = \mathtt{G} \end{cases}$$

- where the mappings $\theta^{\mathtt{B}}, \theta^{\mathtt{G}} : \Delta(S^1) \to \Delta(S^1)$ are given by

$$(\theta^{\mathtt{B}}\pi)(dp^{\mathtt{B}}dp^{\mathtt{G}}) \propto p^{\mathtt{B}}\pi(dp^{\mathtt{B}}dp^{\mathtt{G}})$$
$$(\theta^{\mathtt{G}}\pi)(dp^{\mathtt{B}}dp^{\mathtt{G}}) \propto p^{\mathtt{G}}\pi(dp^{\mathtt{B}}dp^{\mathtt{G}})$$

The maximum is achieved by a Gittins index strategy

Gittins index optimal strategy

There exists a function $\mathcal{I} : \Delta(S^1) \to \mathbb{R}$ (called the Gittins index) — which depends on the discount factor ρ and on the payoff U — such that the following strategy is optimal

- if $\mathcal{I}(\widehat{\pi}_t) < \mathcal{U}_\alpha$ (index $<$ sure payoff), then select decision α ("avoid")
- if $\mathcal{I}(\widehat{\pi}_t) > \mathcal{U}_\alpha$ (index $>$ sure payoff), then select decision ε ("experiment")
- if $\mathcal{I}(\widehat{\pi}_t) = \mathcal{U}_\alpha$ (index $=$ sure payoff), then select indifferently decision α or decision ε

We call the strategy Gittins index optimal strategy (GIOS) and GIOS DM a decision-maker who adopts the Gittins index optimal strategy

The first time τ when the GIOS DM avoids plays a pivotal role

First avoidance time

We denote by τ the first time t, if it exists, when the GIOS DM avoids

$$\tau = \inf\{t = 0, 1, 2 \ldots \mid \mathcal{I}(\widehat{\pi}_t) < \mathcal{U}_\alpha\}$$

The DM that follows the GIOS strategy switches at most one time from experimenting to avoiding (I)

- Infinite learning:
  if $\tau = +\infty$, that is,
  if $\mathcal{I}(\widehat{\pi}_t) \geq \mathcal{U}_\alpha$ for all times $t = 0, 1, 2 \ldots$,
  the GIOS DM never avoids
- No learning:
  if $\tau = 0$, that is, if $\mathcal{I}(\pi_0) < \mathcal{U}_\alpha$,
  the GIOS DM avoids from the start

The DM that follows the GIOS strategy switches at most one time from experimenting to avoiding (II)

- Finite learning:
  if $1 \leq \tau < +\infty$, the GIOS DM
  - experiments from $t = 0$ to $\tau - 1$, that is, as long as
    $$\mathcal{I}(\widehat{\pi}_t) \geq \mathcal{U}_\alpha$$
  - then switches to avoiding at time $t = \tau$, that is, as soon as
    $$\mathcal{I}(\widehat{\pi}_t) < \mathcal{U}_\alpha$$
  - and then keeps avoiding for all times
    (indeed, once the GIOS DM avoids,
    he no longer updates the posterior $\widehat{\pi}_t$,
    so that he keeps avoiding)

Outline of the presentation

Humans display myriads of biases. Why?

When one saw the grass moving

When one saw the grass moving day after day

Killing three biases with one stone

Discussion: when optimal learning induces biases

Humans display myriads of biases. Why?

When one saw the grass moving

When one saw the grass moving day after day
- A model of learning
- A fitness-based optimization problem
- The Gittins index optimal strategy (GIOS)

Killing three biases with one stone
- When learning stops, GIOS induces a status quo bias
- When learning stops, GIOS induces a salience bias
- When learning stops, GIOS induces probability overestimation

Discussion: when optimal learning induces biases

We recover the status quo bias

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

- The optimal rule states that,
  once the GIOS DM selects the "avoid" option,
  he will never more experiment,
  hence no longer updates the posterior $\widehat{\pi}_t$,
  so that he keeps avoiding
- Status quo bias: once stuck in a risk avoidance attitude,
  experimenting brings no benefit

We recover the salience bias

- The Gittins index function $\mathcal{I} : \Delta(S^1) \to \mathbb{R}$ has the property

$$\mathcal{I} \circ \theta^{\mathtt{G}} \geq \mathcal{I}$$

  that is, the index increases when the posterior changes following a "good" outcome, so that

$$\mathcal{I}(\widehat{\pi}_t) \geq \mathcal{U}_\alpha \Rightarrow \mathcal{I}(\theta^{\mathtt{G}}\widehat{\pi}_t) \geq \mathcal{I}(\widehat{\pi}_t) \geq \mathcal{U}_\alpha$$

- As a consequence

$$t < \tau \text{ and } Y_{t+1} = \mathtt{G} \Rightarrow t + 1 < \tau$$

We recover the salience bias

- This "stay-with-a-winner" characteristics of GIOS makes that, when learning stops, the only switch from experimenting to avoiding occurs when a bad outcome materializes
- Salience bias: behaviour change occurs only when a bad outcome materializes

How does the GIOS DM estimate the unknown probability $\overline{p}^{\mathrm{B}}$?

We introduce the numbers N_t^{B} and N_t^{G} of "bad" and "good" encounters up to time t

$$N_0^{\mathrm{B}} = N_0^{\mathrm{G}} = 0$$

$$N_t^{\mathrm{B}} = \sum_{s=1}^{t} \mathbf{1}_{\{Y_s = \mathrm{B}\}} \ , \quad N_t^{\mathrm{G}} = \sum_{s=1}^{t} \mathbf{1}_{\{Y_s = \mathrm{G}\}} \ , \quad t = 1, 2 \ldots$$

We suppose that the DM's prior π_0 is a beta distribution on the simplex S_1

- We suppose that the prior π_0 is a beta distribution

$$\pi_0 = \beta(n_0^{\mathtt{B}}, n_0^{\mathtt{G}}) \propto (p^{\mathtt{B}})^{n_0^{\mathtt{B}}-1}(p^{\mathtt{G}})^{n_0^{\mathtt{G}}-1}$$

  where $n_0^{\mathtt{B}} > 0$ and $n_0^{\mathtt{G}} > 0$
- We can establish that the posterior $\widehat{\pi}_t$ is the beta distribution

$$\widehat{\pi}_t = \beta(n_0^{\mathtt{B}} + N_t^{\mathtt{B}}, n_0^{\mathtt{G}} + N_t^{\mathtt{G}}) \propto (p^{\mathtt{B}})^{n_0^{\mathtt{B}}+N_t^{\mathtt{B}}-1}(p^{\mathtt{G}})^{n_0^{\mathtt{G}}+N_t^{\mathtt{G}}-1}$$

How does the GIOS DM estimate the unknown probability $\overline{p}^{\mathtt{B}}$?

We introduce an estimator $\widehat{p}_t^{\mathtt{B}}$ of the "bad" outcome probability

$$\begin{aligned}\widehat{p}_t^{\mathtt{B}} &= \frac{n_0^{\mathtt{B}} + \text{ number of "bad" encounters up to time } t}{n_0^{\mathtt{B}} + n_0^{\mathtt{G}} + \text{ number of tries up to time } t} \\ &= \frac{n_0^{\mathtt{B}} + N_t^{\mathtt{B}}}{n_0^{\mathtt{B}} + N_t^{\mathtt{B}} + n_0^{\mathtt{G}} + N_t^{\mathtt{G}}}\end{aligned}$$

Here is how the GIOS DM estimates the unknown probability $\overline{p}^{\mathrm{B}}$ of the "bad" outcome

- *Infinite learning:* if $\tau = +\infty$, experiment goes on forever and the GIOS DM estimates $\overline{p}^{\mathrm{B}}$ by the limit

$$\lim_{t \to +\infty} \widehat{p}_t^{\mathrm{B}} = \overline{p}^{\mathrm{B}}$$

  due to the Law of Large Numbers under the (true) probability $\overline{\mathbb{P}}$

- *Finite learning and No learning:* if $0 \leq \tau < +\infty$, the GIOS DM estimates $\overline{p}^{\mathrm{B}}$ by the probability estimator $\widehat{p}_\tau^{\mathrm{B}}$, which satisfies

$$p_c < \widehat{p}_\tau^{\mathrm{B}}$$

$$
\begin{aligned}
& \frac{n_0^{\mathtt{B}} + N_\tau^{\mathtt{B}}}{n_0^{\mathtt{B}} + N_\tau^{\mathtt{B}} + n_0^{\mathtt{G}} + N_\tau^{\mathtt{G}}} \mathcal{U}^{\mathtt{B}} + \frac{n_0^{\mathtt{G}} + N_\tau^{\mathtt{G}}}{n_0^{\mathtt{B}} + N_\tau^{\mathtt{B}} + n_0^{\mathtt{G}} + N_\tau^{\mathtt{G}}} \mathcal{U}^{\mathtt{G}} \\
\leq \quad & \mathcal{I}\big(\beta(n_0^{\mathtt{B}} + N_\tau^{\mathtt{B}}, n_0^{\mathtt{G}} + N_\tau^{\mathtt{G}})\big) < \mathcal{U}_\alpha \\
\Rightarrow & \\
& p_c < \widehat{p}_\tau^{\mathtt{B}}
\end{aligned}
$$

Accuracy is not necessarily optimal

- Notice that, when $\tau < +\infty$, the GIOS DM no longer updates the probability estimator of the "bad" outcome after the finite time τ
- Hence, the DM does not retrieve asymptotically the true probability $\overline{p}^{B}$
- The observation that accuracy is not necessarily optimal — the optimal strategy does not necessarily lead to rightly evaluate the unknown probability — had already been made by economist Michael Rothschild, and is coined the Incomplete Learning Theorem in his paper "A two-armed bandit theory of market pricing." *Journal of Economic Theory*, 9(2):185–202, October 1974
- Our contribution is demonstrating overestimation, that is, we establish a biased form of inaccuracy

We recall the definition of the critical probability p_c

Critical probability / Avoidability index

We define the critical probability (avoidability index)
by the following ratio
of avoidance disutility over "bad" encounter disutility

$$p_c = \frac{\mathcal{U}^{\mathtt{G}} - \mathcal{U}_{\alpha}}{\mathcal{U}^{\mathtt{G}} - \mathcal{U}^{\mathtt{B}}} = \frac{\text{disutility of avoidance}}{\text{disutility of encounter}} \in]0,1[$$

- A low index p_c means that
  it is cheap to avoid the bad outcome
- An index p_c close to 1 means that
  the bad outcome can only be avoided at high costs

We consider outcomes that are bad, unlikely and costly to avoid

We consider the case where

$$\underbrace{\overline{p}^{\mathrm{B}}}_{\text{true probability}} \leq \underbrace{p_c}_{\text{avoidability index}}$$

- Therefore, with foresight, it is worth taking the risk because the (true) probability of the bad outcome is so low that the expected payoff outweighs the sure payoff of avoiding
- Without foresight, we show that it is optimal to take the risk,
  - either forever — obtaining an accurate probability estimate of $\overline{p}^{\mathrm{B}}$ in the long run
  - or up to a finite time — ending up with a probability estimate that is consistently distorted towards overestimating the probability $\overline{p}^{\mathrm{B}}$ of the bad and unlikely outcome

Theorem (Biased Learning Theorem)

Suppose that the "bad" outcome B *is unlikely, in the sense that*

$$\overline{p}^{\mathrm{B}} \leq p_c$$

A decision-maker who follows the Gittins index optimal strategy

- *either experiments forever,*
  and he accurately estimates *asymptotically*
  the (true) probability of the bad and unlikely outcome B
- *or experiments during a finite number of periods (possibly 0)*
  and, when the experiment phase ends, he overererestimates
  the (true) probability of the bad and unlikely outcome B

We consider cases where, with foresight, it would be worth taking the risk

Case $\overline{p}^{\mathrm{B}} \leq p_c$	estimate of $\overline{p}^{\mathrm{B}}$	comment
$\tau = +\infty$	$\lim_{t \to +\infty} \widehat{p}_t^{\mathrm{B}} = \overline{p}^{\mathrm{B}}$	exact estimation of $\overline{p}^{\mathrm{B}}$
$0 \leq \tau < +\infty$	$\widehat{p}_\tau^{\mathrm{B}} > p_c \geq \overline{p}^{\mathrm{B}}$	overestimation of $\overline{p}^{\mathrm{B}}$

Outline of the presentation

Humans display myriads of biases. Why?

When one saw the grass moving

When one saw the grass moving day after day

Killing three biases with one stone

Discussion: when optimal learning induces biases

In many situations, probabilities are not known but learnt

- The recent nuclear accident in Japan (2011) has led many countries to stop nuclear energy
- This sharp switch may be interpreted as the stopping of an experiment phase where the probability of nuclear accidents has been progressively learnt

We suggest that rational overestimation of loss probability might enlighten some economic puzzles

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

- 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
- However, this analysis is done *ex post* under risk, whereas decision-makers take 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

Conclusion: are our innate biases the consequences of optimal strategies, selected through evolution?

- In many situations of uncertainty, probabilities are not known, but can be learned
- In our setting, we show that optimal learning can stop before full learning and that, under general assumptions,

1. When learning stops,
   we overestimate the probability of an outcome that is bad, unlikely and costly to avoid
2. When learning stops,
   we do it necessarily when a bad outcome materializes
3. When learning stops,
   we will not learn again and we stick to the (no risk) status quo

We are biased by design

THANK YOU :-)