A Statistical Physics Theory of Adaptive Human Behavior

Constantin Vaillant-Tenzer [1 3] Etienne Koechlin [1 2]

[1]Ecole Normale Supérieure - PSL [2]INSERM [3]Université Paris Cité

Introduction

Our aim is to understand how works **adaptivity** of human behavior.

Previous approaches

There were **Utility maximization** (Ferrari-Toniolo et al. 2021); **descriptive heuristic approaches** (Gigerenzer and Gaissmaier 2011, Newell 2005); and the elaboration of **the free energy principle**, inducing minimization of surprise (Friston 2009).

Postulates

Behavior is driven by learning world models but constrained by resources. There is an evolutionary sense to be satisfied to continue learning world models.

Information theory

For a statistical distribution characterized by the different world possible states $\theta \in \mathbb{R}^n$, we define the **statistical entropy** $H(\theta)$ as follows :

$$H(\theta) := -E(\theta) := -\sum_{\theta \in \mathbb{R}^n \,;\, \mathbb{P}(\theta) \neq 0} \ln(\mathbb{P}(\theta))\mathbb{P}(\theta)$$

The **information gain** I on the parameter θ, after the action a, is the value, where H is the statistical entropy :

$$I\left(\theta, \varphi(a)|a\right) = H\left(\theta|a\right) - H\left(\theta|(\varphi(a), a)\right)$$

The postulates in equations

Maximizing acquired information

$$\arg \max_{A \in \mathrm{Var}((\Omega, \mathcal{T}, \mathbb{P}), (\Omega, \mathcal{T}))} \left(\sum_{a \in A(\Omega)} I\left(\theta, \varphi(a)|a\right) \mathbb{P}(A = a) \right)$$

Constraint on resources

$$U - \sum_{a \in A(\Omega)} \varsigma(a)\mathbb{P}(a) \geq 0 \qquad (1)$$

With $\varsigma(a) = E_\theta R_\theta(\varphi(a)) - \omega(a)\mathbb{P}(a) - c_I I\left(\theta, \varphi(a)|a\right)$

Constraint on entropy

Regularization constraint on selection effort :

$$-\sum_{a \in A(\Omega)} \ln(\mathbb{P}(a))\mathbb{P}(a) - S \geq 0 \qquad (2)$$

General form of the solution

The **unique** distribution behavior function is, with $\lambda > 0$ and $\beta > 0$:

$$\forall a \in A(\Omega),\ \mathbb{P}(a) = \frac{1}{z} \exp\left(\beta\left(I(\theta, \varphi(a)|a) - \lambda\varsigma(a)\right)\right)$$

With $z = \sum_{a \in A(\Omega)} \exp\left(\beta\left(I(\theta, \varphi(a)|a) - \lambda_1\varsigma(a)\right)\right)$ is the partition function.

Exploration - exploitation parameter λ

We made different kinds of analytical and numerical computations. For β big enough (or S little enough, working for $S < \frac{\ln(N)}{3}$) :

$$\lambda \approx \frac{I(b) - \frac{\ln(z)}{\beta}}{\varsigma(b)} \approx \frac{I(b)(z-1)}{U - \varsigma(b)} = O\left(\frac{1}{U}\right)$$

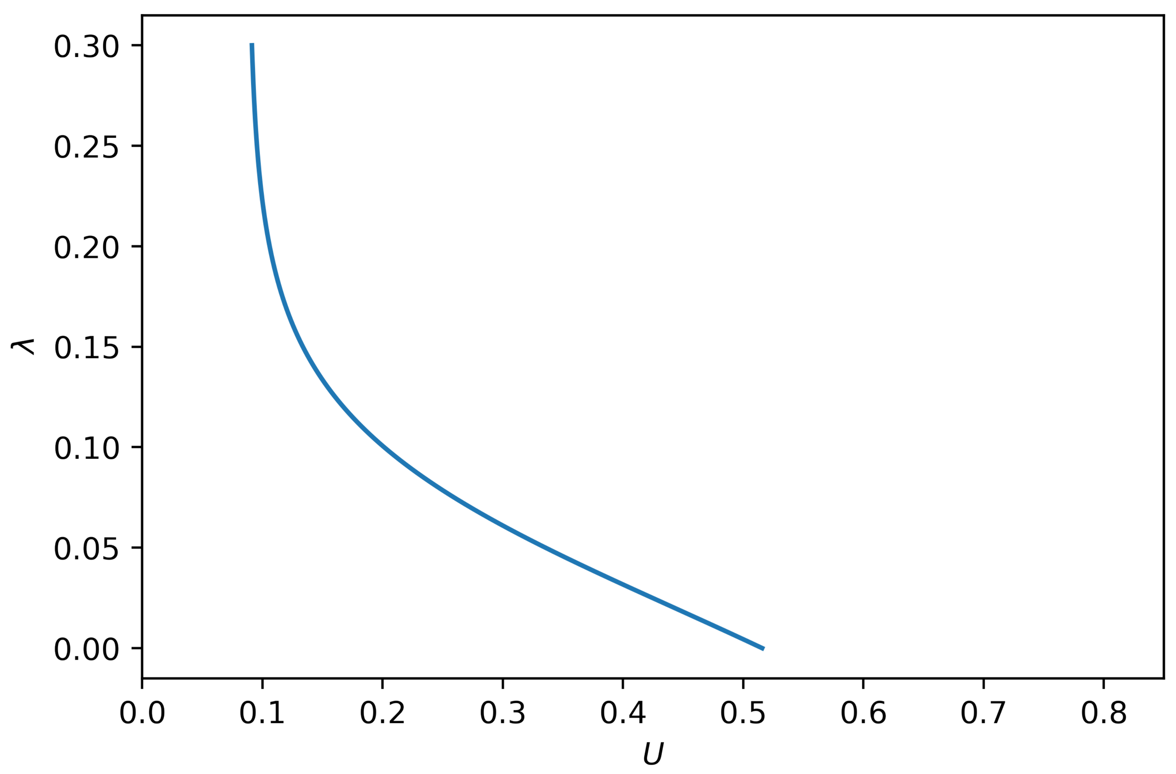

Figure 1. Lagrangian multipliers β and λ in function of the amount of resources U.

Generalization for continuous environments

This model may also represent motor actions. One wishes to find the probability measure m that maximizes :

$$\int_\Omega I\left(\theta, \varphi(a)|a\right) m(a) d\mu(a)$$

The constraint on a resource takes a continuous form and the constraint on entropy can be written trough Kullback-Libber divergence.

Application to bandits

This model can be experimentally applied and tested on bandits : a gambler arrives in a casino and has several slot machines, he/she does not know anything about. What choices would he/she make depending on his resources ? We can assume that we make information inferences using Dirichlet processes (Domenech, Rheims, and Koechlin 2020). We see that, as experimentally, even if they founded the most reliable slot machines, they will keep exploring.

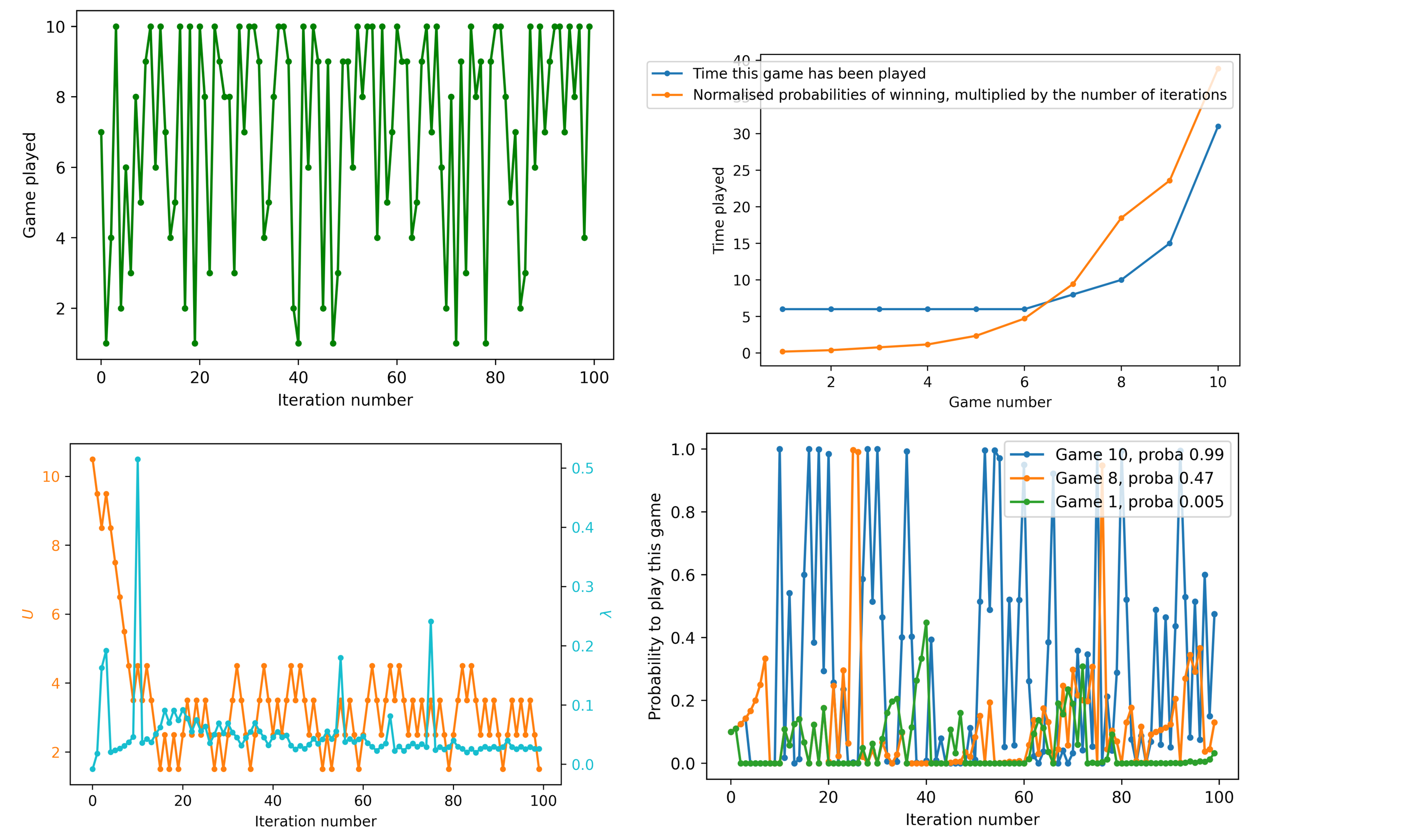

Figure 2. Action of a gambler in ten single-armed bandits (**top - left**) and number of time he plays each game (blue curve), compared to the number of time he/she would have played if he was playing proportionally to winning probabilities that he/she does not know (orange curve). Games are sorted from the least profitable (game 0) to the most profitable (game 9) (**top- right**). The **bottom-left** diagram represents evolution of the amount of resources U and of the multiplier λ and the **bottom-right** one the evolution of probabilities of choosing some games over time.

Next steps

There are still much things to do ! This includes **experimentally** test results, analyze data and **compare** with the other theories mentioned in introduction ; develop the **general version** of the model and test it and finally do **model based fMRI** to asses the neuroscience pertinence of our model.

References

Domenech, Philippe, Sylvain Rheims, and Etienne Koechlin (2020). “Neural mechanisms resolving exploitation-exploration dilemmas in the medial prefrontal cortex”. In: *Science* 369.6507, eabb0184.

Ferrari-Toniolo, Simone et al. (2021). “Nonhuman primates satisfy utility maximization in compliance with the continuity axiom of Expected Utility Theory”. In: *Journal of Neuroscience* 41.13, pp. 2964–2979.

Friston, Karl (2009). “The free-energy principle: a rough guide to the brain?” In: *Trends in cognitive sciences* 13.7, pp. 293–301.

Gigerenzer, Gerd and Wolfgang Gaissmaier (2011). “Heuristic decision making”. In: *Annual review of psychology* 62, pp. 451–482.

Newell, Ben R (2005). “Re-visions of rationality?” In: *Trends in cognitive sciences* 9.1, pp. 11–15.