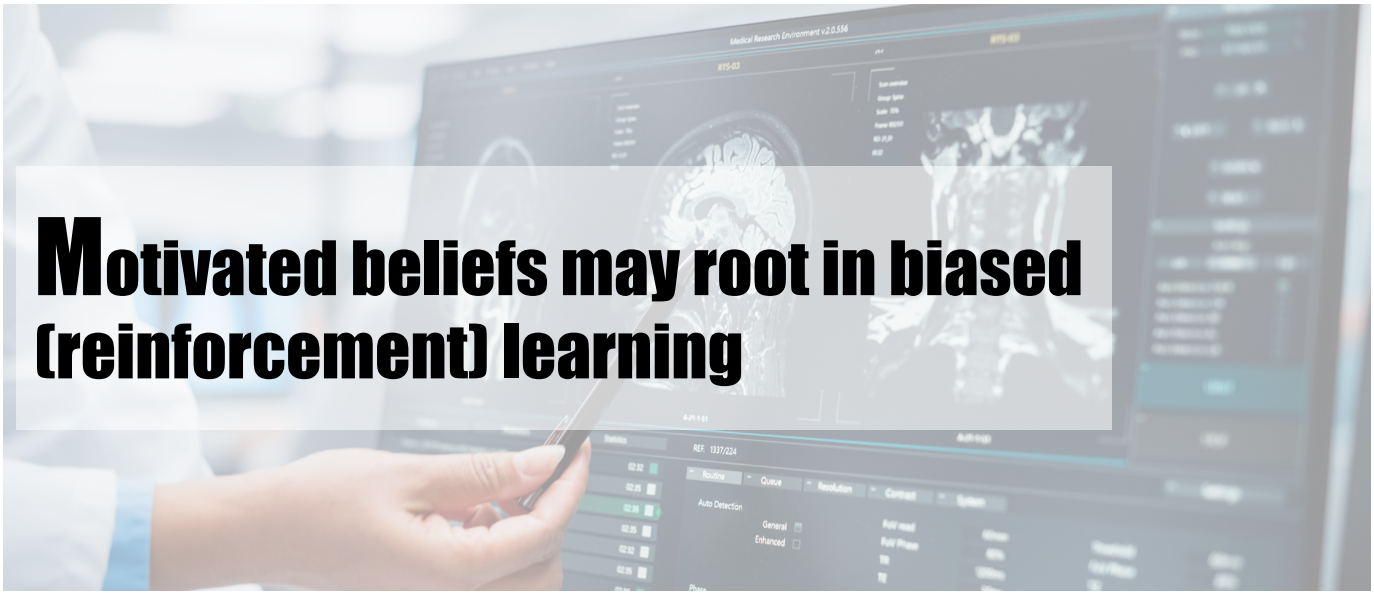
CHAIR
HERMÈS PARIS
PARIS SCHOOL OF ECONOMICS
ÉCOLE D'ÉCONOMIE DE PARIS
Opening economics

By
Maël Lebreton
& Stefano Palminteri



Motivated beliefs may root in biased (reinforcement) learning

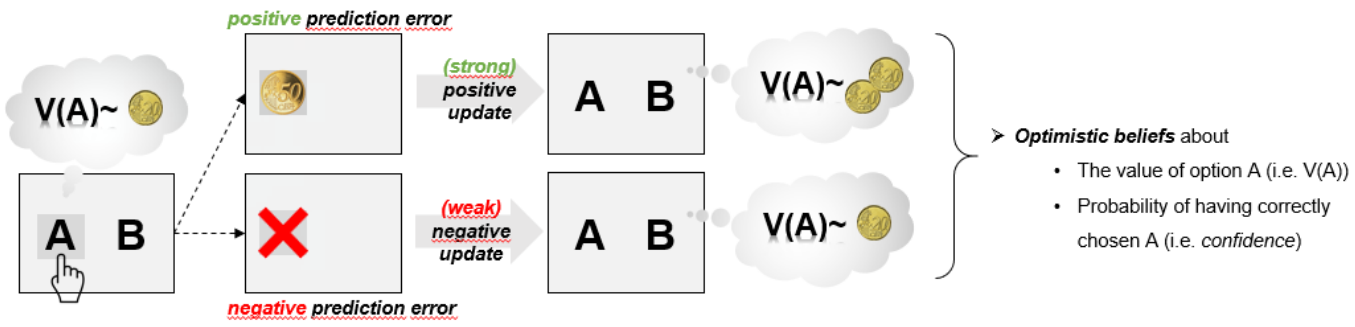
Since the second half of the 20th Century and the emergence of the fields of behavioural and experimental economics, economists and psychologist have documented a flurry of instances where human behaviour depart from rational expectations. A whole class of those decision and cognitive biases seem to specifically affect our beliefs, i.e. our judgments and assumptions about our environment, ourselves, and our behaviour. More specifically, empirical investigations consistently report that human decision makers tend to form, hold and actively maintain beliefs in part because they attach value to them. Such instances of *motivated beliefs* include over-optimism, i.e. the tendency to overestimate the probability of occurrence of positive events and, reciprocally, to underestimate the probability of occurrence of painful or negative events; and overconfidence, i.e. the tendency to overestimate our general abilities and qualities, as well as the fitness of our behaviour, actions and judgments. Critically, overoptimistic and overconfident tendencies are responsible from considerable societal costs, as they indiscriminately contribute to suboptimal decision-making in the behaviour of consumers, investors and top corporate executives, but also medical practitioners or political decision-makers.

Current dominant theories of *motivated beliefs* suggest that over optimism and overconfidence are somewhat specific to elaborated, semantic beliefs about oneself and the world, and that they created and maintained by a process of reality denial, whereby individuals do not update these beliefs properly in response to bad news. Over the last 5 years, in collaboration with the Human Reinforcement-Learning Lab led by Dr. Stefano Palminteri at the *Département d'études cognitives* of the École normale supérieure, we published a series of empirical and theoretical studies exploring if and how motivated-belief-like patterns can also be observed in basic choice behaviour, and whether they could be caused by a similar asymmetry in the processing of good news versus bad news.

In a first study, published in *Nature Human Behavior*, we collected and analyzed the behavior of human participants in an instrumental *reinforcement-learning* task, akin to the one developed to understand animals' conditioned behavior at the beginning of the 20th Century (Lefebvre *et al.*, 2017). Reinforcement-learning characterizes the process through which we learn, by

trial-and-errors, to repeat actions that bring us rewards and to avoid action that bring us pain or losses. Reinforcement-learning is therefore a very elementary cognitive process which governs our behavior, which is essential for survival, and which is therefore shared with almost all animals (from worms and flies to mammals). In our reinforcement-learning task, human participants were repeatedly faced with pairs of abstract cues, each probabilistically paired with monetary outcomes. They could maximize their monetary outcome by learning through trial-and errors to preferentially choose the cue paired with the most profitable outcomes distribution. To describe and analyze participants' choice behavior, we used a reinforcement-learning model, inspired from neuroscience and computer science, that can efficiently solve these learning task. The model assigns expected values to the available options, which are updated through a simple error-correction learning mechanism: upon the reception of an outcome, this outcome is compared to its prior expectation, generating a so-called prediction error. The prediction error, weighted by a key parameter called *learning rate*, is then used to update (or correct) the outcome expected value. Critically, we found that the choices of most participants are best described and explained by a model that possesses *two different learning rates*: one that controls the updating process after a positive prediction error (i.e. a better than expected outcome) and another that controls the updating process after a negative prediction error (i.e. a worse than expected outcome). When comparing these two *learning rates*, we observed that the positive *learning rate* had significantly higher values than the negative *learning rate*. In other terms, even in this very basic task featuring binary choices and basic rewards, participants behave in a similar way to the reality denial pattern: they fail to update their behavior properly in response to bad news, and reciprocally over-react in response to good news. This good news-bad news asymmetry generates over-optimistic expectations about the rewards associated with chosen options, and about the probability of making the right choices *(Figure 1)*. This result showed, for the first time to our knowledge, that over-optimism also exists in the realm of a very elementary cognitive process: the reinforcement-learning operations that governs our basic reward seeking behavior.

Figure 1.
A reinforcement-learning mechanism that generates optimistic beliefs and over-confidence.



Note: When choosing between two options (A or B) whose value is uncertain, an agent bases its choice on its estimation of the option's expected value (V(A) and V(B)). After the choice, the outcome can be better or worse than expected, leading to positive or negative prediction-errors. These prediction-errors are used to update the option's expected values, but in an asymmetrical way: the update is much stronger in the positive than the negative case, leading to a potential overestimation of the chosen option's value, and henceforth of the belief.

We comprehensively reviewed the empirical reinforcement-learning literature, to evaluate if this asymmetric good news-bad news updating pattern could be observed and generalized in other contexts (Palminteri & Lebreton, 2022). Our literature review first revealed that the asymmetric updating pattern had been present (though largely overlooked), in a large variety of reinforcement learning task in humans: higher learning-rates after positive than negative feedback can be observed in reinforcement-learning tasks where participants have to accumulate monetary gains but also in task where they have to avoid monetary losses, or even electric shocks. The learning-rate asymmetry was also observed under different regimes of outcome probabilities: high and low outcome probabilities, as well as stable or volatile regimes. We also leveraged the fact that our experimental paradigm is simple enough to be routinely used in non-human animals to identify studies where non-human primates and rodents also unambiguously exhibited the asymmetric value-updating pattern. In summary, our literature review, ultimately published in *Trends in Cognitive Sciences*, uncovered that a significant body of empirical studies in humans and animals is consistent with the idea that the asymmetries that affect high-level belief updates are shared with more elementary forms of updates, notably in reinforcement-learning. This suggests that *motivated beliefs* could be the result of a hard-wired asymmetry embedded in our neural architecture.

In our most recent study, we collected and analysed several datasets where participants performed a reinforcement-learning task, and also had to indicate their confidence in their choices (Salem-Garcia *et al.*, 2023). Among other results, we found that asymmetric updating tendencies were associated with overconfidence: the more individuals over-reacted in response to good news and underreacted to bad news, the more they were unreasonably confident in their choices. This confirms that a key link exists between high-level *motivated beliefs* about oneself (overconfidence) and the elementary asymmetric updating observed in reinforcement-learning, and suggests that both phenomena could share similar neuro-computational architecture and mechanisms. These results, recently published in *Psychological Review*, constitute additional evidence in favor of the *hard-wired* hypothesis about the origin and source of *motivated beliefs*.

In conclusion, this line of research –at the crossroads of psychology, economics and neurosciences–, demonstrates that the asymmetric information update mechanism leading to *motivated belief* is more general than initially thought in the economics literature. Not only do these results challenge the current and dominant theories of *motivated beliefs*, but they also open the way to the elucidation of their computational and neural mechanisms.

References

Bénabou R. & Tirole J., 2016, "Mindful Economics: The Production, Consumption, and Value of Beliefs", *Journal of Economic Perspectives*, 30(3), 141–164.

Kunda Z., 1990, "The Case for Motivated Reasoning", *Psychological Bulletin*, 108(3), 480.

Lefebvre G., Lebreton M., Meyniel F., Bourgeois-Gironde S. & Palminteri, S., 2017, "Behavioural and neural characterization of optimistic reinforcement learning", *Nature Human Behaviour*, 1(4), Article 4.

Loewenstein G. & Molnar A., 2018, "The Renaissance of Belief-Based Utility in Economics", *Nature Human Behaviour*, 2(3), 166–167.

Palminteri S. & Lebreton M., 2022, "The computational roots of positivity and confirmation biases in reinforcement learning", *Trends in Cognitive Sciences*, 26(7), 607–621.

Salem-Garcia N., Palminteri S. & Lebreton M., 2023, "Linking confidence biases to reinforcement-learning processes", *Psychological Review*, 130(4), 1017–1043.

Sharot T., Korn C.W. & Dolan R.J., 2011, "How unrealistic optimism is maintained in the face of reality", *Nature Neuroscience*, 14(11), Article 11.

Zimmermann F., 2020, "The Dynamics of Motivated Beliefs", *American Economic Review*, 110(2), 337–361.

- **Maël Lebreton** is senior researcher at the Paris School of Economics and head of the ERC Starting Grant INFORL.
- **Stefano Palminteri** is researcher at the National Institute of Health and Medical Research (INSERM) and at the École normale supérieure.