Surprise-based learning: Neuromodulation by surprise in multi-factor learning rules

Mohammad Javad Faraji[1], Kerstin Preuschoff[1], and Wulfram Gerstner[1]

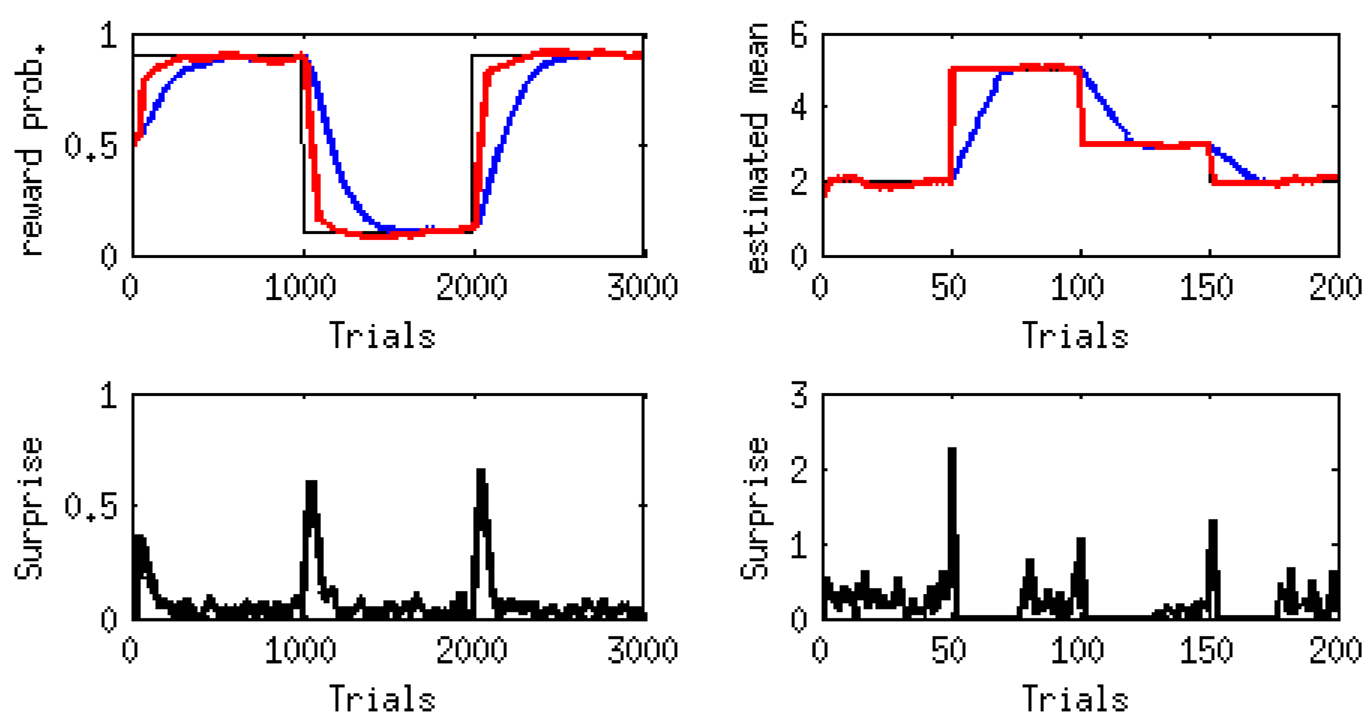
1. School of Life Sciences, Brain Mind Institute & School of Computer and Communication Sciences, EPFL, Switzerland

Abstract

The role of surprise in synaptic learning rules in neural networks is largely undetermined. We address (1) how surprise affects learning and (2) how surprise signals are reflected in neural networks. We show that surprise in principle can improve learning by

- modulating the learning rate,
- regulating the exploration-exploitation trade off, and
- generating new environmental states.

Modulating learning rate by surprise



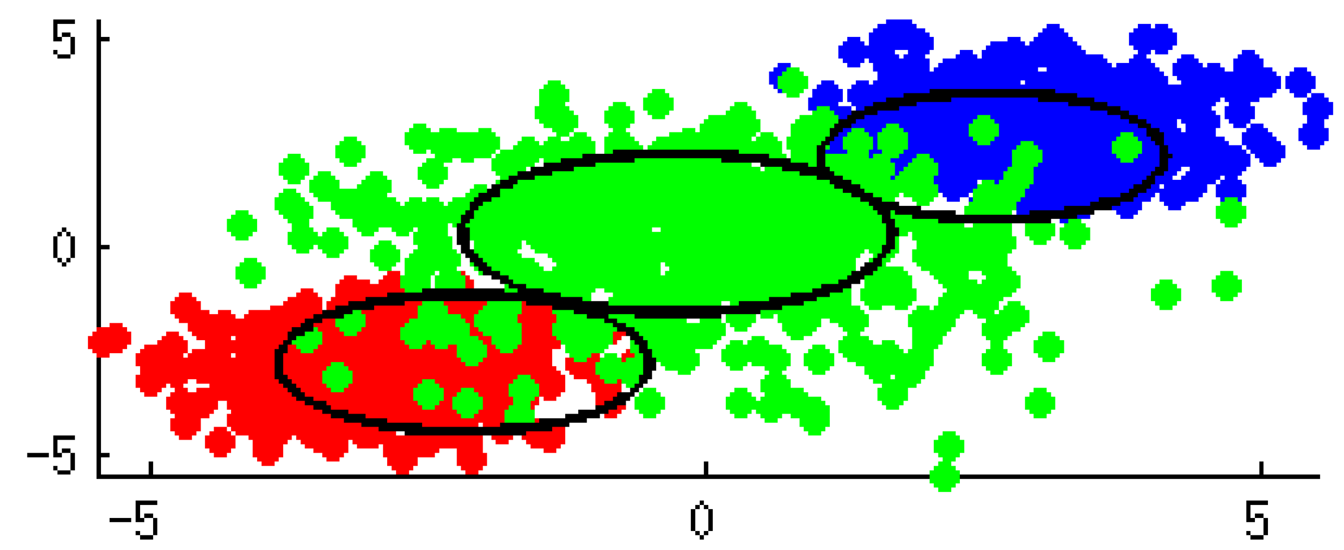
The agent estimates the probability of reward delivery in a reversal task (upper-left). We altered a standard SARSA learning algorithm (blue line) such that when the agent detects an unexpected event beyond the stochasticity of the environment, the ensuing surprise signal (bottom-left) temporarily accelerates learning leading to **more accumulated reward** for surprise-based reinforcement learner (red line). In the dynamic decision making task (upper-right), the learner observes samples from a Gaussian distribution with varying mean (black line). Modulating the learning rate by surprise signals (bottom-right) leads to **faster detection of change points** (red line) than that in the SARSA model (blue line).

References

1. Guo et al. "Neural Correlates of the Learning Rate in a Changing Environment", Cosyne abstract, 2012.
2. Preuschoff et al. "Human insula activation reflects risk prediction errors as well as risk.", J. Neuroscience 28.11 (2008): 2745-2752.

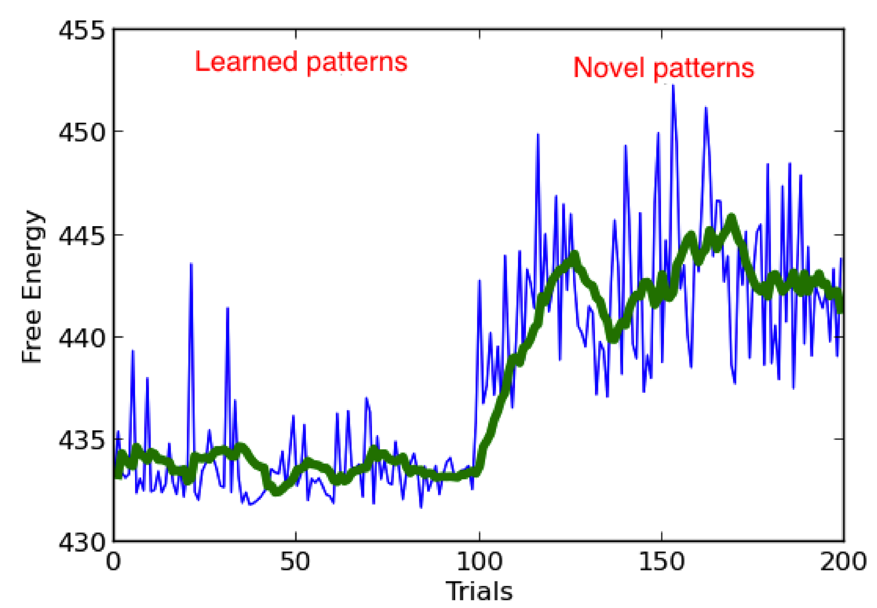
Research was supported by the ERC (grant no. 268 689, H.S.)

Surprise triggers new clusters



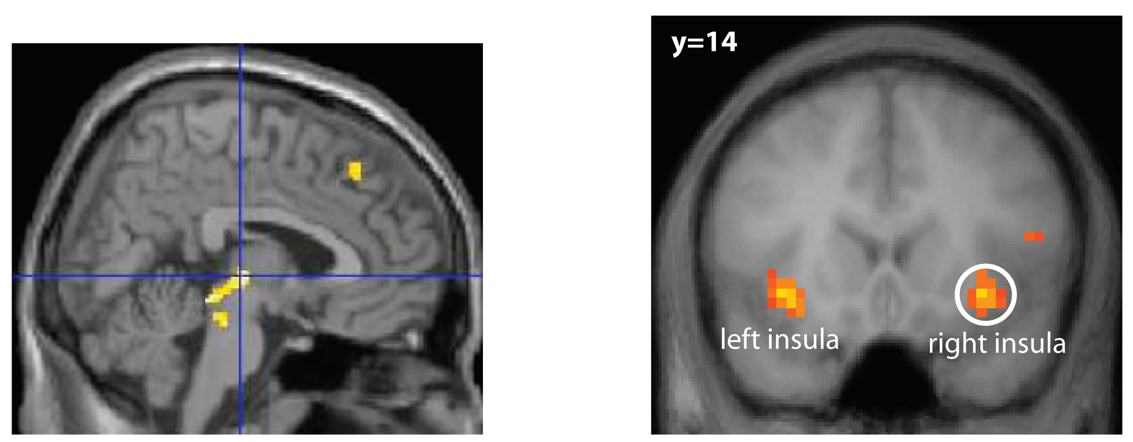
A classic K-means clustering algorithm is modified such that if the total number of clusters is initially unknown, the agent (classifier) equipped with surprise is able to add more clusters (black circles) whenever it judges a pattern (colored data point) to be surprising, i.e., a pattern which may belong to none of the existing clusters. It represents an agent who is able to generate (trigger) new states, **an essential feature for learning new environments**.

Free energy in neural networks



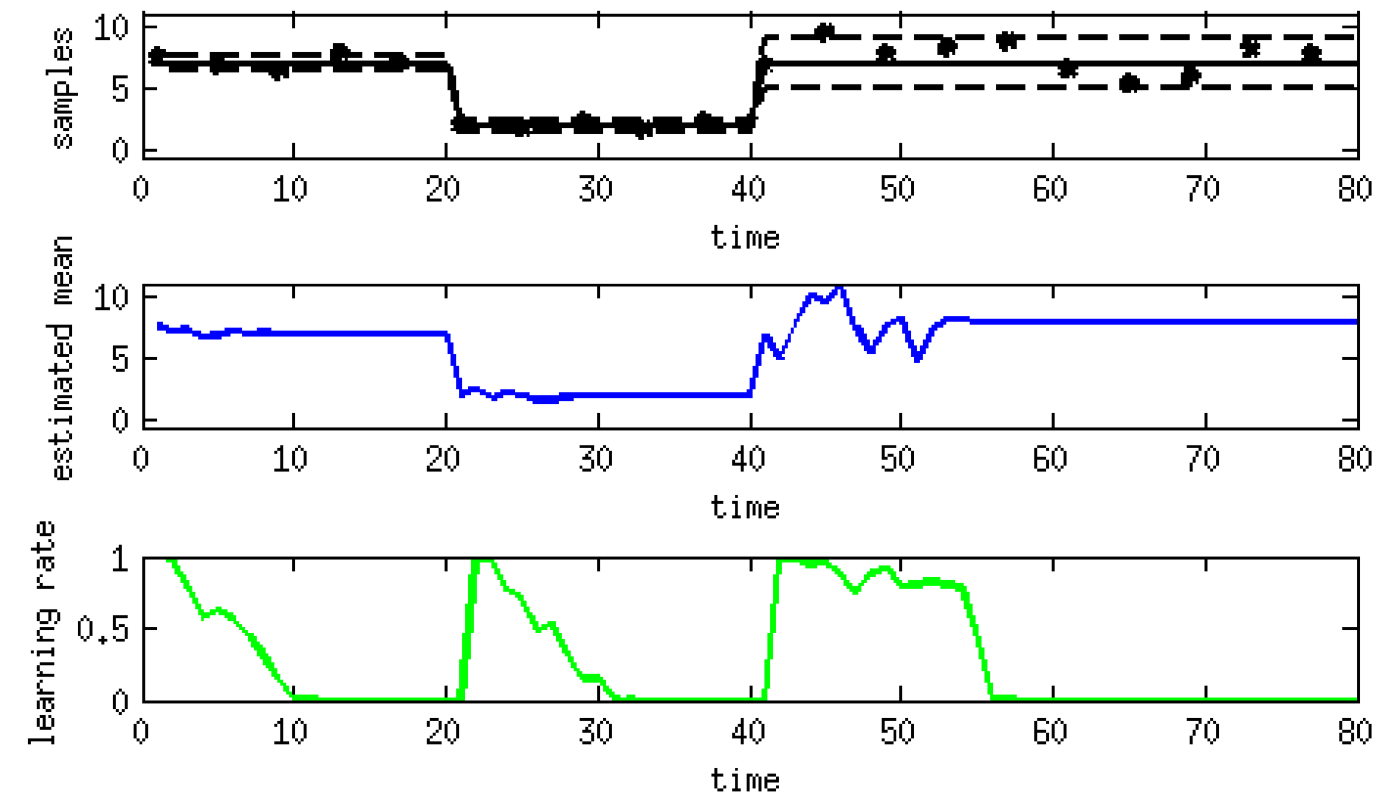
The variational free energy (blue line), used for estimating the likelihood of the input patterns (digits from the MNIST dataset) in a Boltzmann machine can be used as a surprise measure.

Neural signatures of surprise

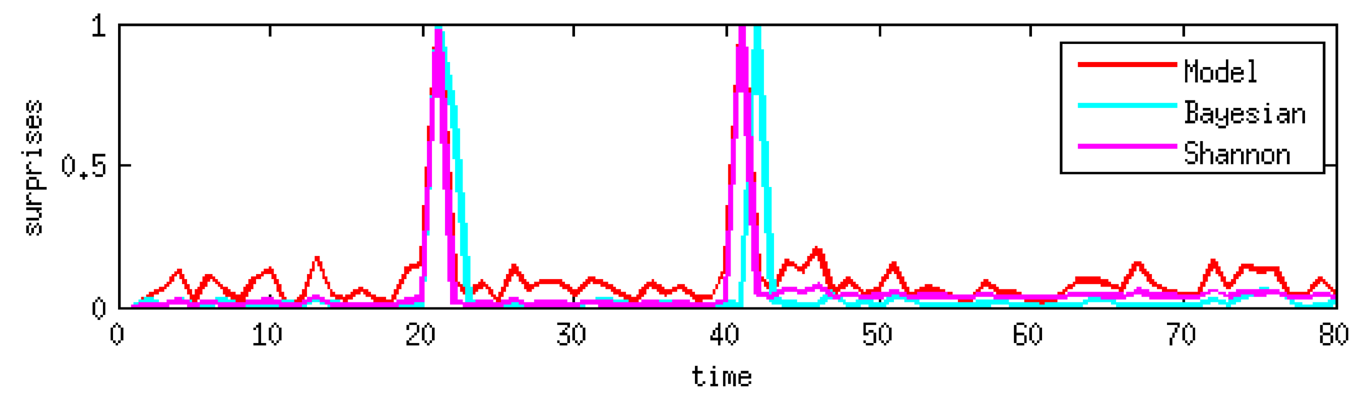


BOLD response to adaptive learning rates during a dynamic decision making task (left) [1]. BOLD response to surprise when measured as risk prediction error in a gambling task (right) [2].

Computational model for learning rate



Reward prediction error $\delta_n = r_n - \hat{\mu}_{n-1}$ and the estimated risk $\hat{\sigma}_n$ of the environment is used to measure surprise $S_n = f(|\delta_n|/\hat{\sigma}_{n-1})$. Dynamics of the learning rate α is then controlled by S_n and the level of estimation uncertainty $\hat{u}_n = std(\hat{\mu})$ which determines variation of the estimated mean reward $\hat{\mu}$. In other words, $\dot{\alpha} = -\alpha/(k\hat{u}_{n-1}) + S_n$ where k is a constant.



Theses results hold equally well for different surprise measures: Shannon surprise $-\log P(r_i|\hat{\mu}_{n-1}, \hat{\sigma}_{n-1})$, Bayesian approach $D_{KL}[P_{n+1}(\hat{\mu}|r_n)||P_n(\hat{\mu})]$, and model-free $|r_n - \hat{\mu}_{n-1}|/\hat{\sigma}_{n-1}$.

Conclusion

Surprise-driven modulations can enhance the learning performance at both the behavioral and neural network level. In two decision making tasks, surprise-based SARSA accelerated learning. A surprise-based clustering algorithm can trigger new clusters if it judges a pattern to be novel. Further, we simulated a classic Boltzmann machine to use the network activity itself to measure how much a new pattern is surprising. Since the surprise signal is generated by the network itself, it can be used as a biologically plausible third factor in multi-factor learning rules.