



HAL
open science

Gain modulation of actions selection without synaptic relearning

Elif Köksal-Ersöz, Pascal Chossat, Frédéric Lavigne

► **To cite this version:**

Elif Köksal-Ersöz, Pascal Chossat, Frédéric Lavigne. Gain modulation of actions selection without synaptic relearning. 2024. hal-04418804

HAL Id: hal-04418804

<https://hal.science/hal-04418804>

Preprint submitted on 26 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Gain modulation of actions selection without synaptic relearning

Elif Köksal-Ersöz ^{*1, 2}, Pascal Chossat^{2, 3}, and Frédéric Lavigne⁴

¹Univ Rennes, INSERM, LTSI – UMR 1099, Rennes, France.

²Project Team MathNeuro, INRIA-CNRS-UNS, Sophia Antipolis, France.

³Université Côte d’Azur, Laboratoire Jean-Alexandre Dieudonné, Nice, France.

⁴Université Côte d’Azur, CNRS-BCL, Nice, France.

Abstract

Adaptation of behavior to a changing environment requires the brain to change the probability of selecting actions based on their reward or punishment. The present study proposes a simple neural mechanism of gain modulation that makes possible rapid changes in the probability of selecting actions after different levels of negative feedback. The effects of gain modulation on action selection are studied in a network of neuronal populations of excitatory neurons regulated by inhibition. Results show that neuronal gain can embed memories of past experiences without modification of synaptic efficacies, in complement to synaptic plasticity.

Keywords: neuronal gain, short-term depression, local inhibition, latching dynamics, slow-fast dynamics

1 Introduction

Adaptation of behavior requires the brain to select actions that bring benefits and to avoid those that bring costs. The optimal strategy for selecting the most probably rewarded action is to update the relation between actions and the probability of rewards and punishments from past experience [25, 28, 5]. Statistical learning of the probability of actions and feedbacks requires repetition of trials and errors to fit the new probabilistic structure of the action-feedback relation [20, 26, 21]. In a stable environment the probability of the outcomes is best estimated by experiences going back a long way to ensure exploitation of the rewarded actions. If the probabilistic structure of rewards in the environment changes, the selection of actions becomes prone to errors due to uncertainty on the type of feedback that will come up. Exploration of different actions and relearning must then update the action-outcome relations [8, 11]. This raises the question of the degree of recency of the experiences and of the intensity of the outcomes to be taken into account.

In case of rapid changes, old experiences are no longer useful for current selection of actions. Animal studies report that ancient and recent rewards are both memorized [9, 14, 6]. Changes in the rate of rewards and punishments can be adjusted by changes in a single learning rate to update synaptic efficacies [5, 14, 24, 23]. For example, synapses connecting a given context to different possible actions are updated at a rate that depends on the rate of rewards and punishments of the different actions. Computational models have shown that the learning rate can vary with the magnitude of the error signal to optimize the weighting of old and recent experiences by learning over multiple timescales [15, 16].

Some changes in the environment are transient, lasting only as long as it returns to its initial state. This is the case when different structures of action-feedback are associated to different alternating contexts. In this case, it is beneficial to adapt behaviors to the transient state without forgetting the previous state and hence without the need to relearn it. Synaptic relearning leads to forgetting of the previous probabilistic structure of action-outcome relations. The previously learned and forgotten environment has then to be learned again. More dramatically, in the extreme case of severe and dangerous punishment, the action that led to it must not be repeated. Some errors must not be made twice. This does not give time for statistical relearning. Then how to change actions without synaptic relearning, and in the extreme case without the need for any further learning trial?

Various cortical functions are reported to rely on the modulation of the input-output gain at the level of neurons, defined as the slope of the neurons transfer function [27, 7, 12, 13]. In cortical networks models, gain modulation changes neurons’ correlations in output activity [10] and gives the network the computational ability to change the level of neuronal activity by a context for fixed values of synaptic efficacies [19]. Further, a recent computational model has reported that the network can switch activation between sequences of neurons, which encode memory items, as a function of modulation of the gain of these neurons [18]. The modulation of neuronal gain is efficient in changing the probability of selecting a sequence of items or another in the network phase space for a fixed synaptic matrix. Here we investigate the conditions under which neuronal gain alone enables switching from one action to another without the need for synaptic relearning and possibly with immediate effect. To this aim, we study if and how gain modulation can change the sequence of neurons activated during the punished trial and on the immediately following trial.

*Corresponding author: elif.koksal-ersoz@inserm.fr

2 Methods

The model has been directly inspired by [18], where the retrieving of multiple sequences in a collection of P learned states ξ^1, \dots, ξ^P has been investigated using the framework introduced in [2, 17]. In this model each learned state is a dynamically stable ‘pattern’ made of two active units coding for populations of neurons in the neural network, the others units being inactive. Moreover, these patterns can be destabilized under the effect of short term synaptic depression (STD), hence allowing for dynamics of activation of patterns in the network state space.

In [2, 17] any two consecutive patterns shared one active unit, so that the patterns formed a chain of overlapping states $\xi^1 - \xi^2 - \dots - \xi^P$. It was shown that, under the effect of noise, the overlapping condition allowed to produce a sequential stochastic dynamics, one state ‘jumping’ to the next with high probability. This behavior is called ‘latching’ dynamics.

In [18] we analyzed the case when, as the system starting from ξ^1 reaches a given state ξ^m in the chain (the branching ‘node’), there is a choice among several continuing branches. In the simplest case, which is considered in the present work, the chain splits at the node into two chains $\xi^m, \xi^{m+1}, \dots, \xi^q$ (branch 1, Br-1) and $\xi^m, \xi^{q+1}, \dots, \xi^P$ (branch 2, Br-2).

As shown in [18], as long as the connectivity matrix is symmetric, the probabilities that the dynamics starting from ξ^1 continues on either branch 1 or branch 2 are equal, but whenever the weights of connections from ξ^m to ξ^{m+1} are greater than those from ξ^m to ξ^{q+1} , the probability to continue on branch 1 is larger than the probability to continue on branch 2. The branching network which we numerically investigated is pictured in Figure 1A. Its parameters are $N = 10$, $P = 9$, $m = 4$ and $q = 7$.

The equations for the units are derived from [3], in which we have replaced the membrane potential u_i of each unit i by the activity $x_i = S(u_i) = 1/(1 + e^{-\gamma u_i})$. The variables are now the activities which take values in the interval $[0, 1]$ after $S^{-1}(x_i)$ has been replaced by its polynomial expansion (we chose the simplest, linear approximation). The inhibition effects within the network are modeled by a term proportional to the averaged activity (see [22]).

The equation for unit i reads

$$\dot{x}_i = x_i(1 - x_i) \left(-\frac{4}{\gamma} x_i + \sum_{j=1}^N J_{i,j} x_j - \lambda \sum_{j=1}^N x_j - \lambda \nu_i x_i \right) + \eta \quad (1)$$

where γ is the gain and λ is the inhibitory coefficient. The coefficient ν_i accounts for the possibility of short-range inhibitory loops between excitatory neurons and inhibitory interneurons. This allows for selective self-inhibition of excitatory neurons, which depends on the number of afferent connections to excited units. In our case we take $\nu_i = 0$ for $i \neq m+1$ and $\nu_{m+1} = 1$ (unit $m+1$ receives inputs from units $m, m+1$ and $q+1$). Finally η is a noise term. We refer the reader to [18] for details. Note that $x_i = 0$ or 1 are always solutions of (1). Hence, any state such that $x_i = 0$ or 1 , $i = 1 \dots, N$ is a steady-state of the system. The learned states are stable solutions of this type.

The STD is expressed as follows. Let $J_{ij}(t)$ be the strength of the connection from units j to i at time t . We write $J_{ij}(t) = J_{ij}^{max} \cdot s_j(t)$, where J_{ij}^{max} is the connectivity matrix resulting from the learning process and $s_j(t)$ follow the STD law given in [29], which is equivalent to

$$\tau_r \dot{s}_i = 1 - s_i - \rho s_i x_i. \quad (2)$$

The weights J_{ij}^{max} are computed from the simple Hebbian rule $J_{ij}^{max} = \sum_k \xi_i^k \xi_j^k$.

We chose here the simplest, yet bio-inspired formulas for the model equations. We expect that more complicated or refined formulas would not significantly change the main outcome.

The parameter values of (1)-(2) are optimized from [18] as $\lambda = 0.6, I = 0, \nu_4 = 1, \rho = 300, \tau_r = 1.2, \gamma = 10$ without punishment, $\gamma = \{9, 5, 3.3, 2.5\}$ with punishment (from low (10%) to high (75%) rates).

3 Results and Discussion

The model presented here allows mathematical analysis and simulation of the selection of actions as a function of gain modulation and for a fixed synaptic matrix. We investigate a system of $N = 10$ units (x_i , for $i \in \{1, \dots, 10\}$) encoding 9 patterns (Fig 1A), which we write (A,B, ..., I) for convenience (Fig 1B). The units are placed in a 3-node graph where units 1-3 are along branch 0, units 5-7 along branch 1, units 8-10 along branch 2 and unit x_4 is the branching node connected to x_3, x_5 and x_8 (Fig 1A). The synaptic coupling coefficient between units 4 and 5 is 10% stronger than between units 3 and 4, and units 4 and 8. For simulations, the system was initialized at pattern A and punishment was applied to units 5 and 6 (coding for pattern E) by reducing their gain when they became activated during the trial N. The feedback is assumed here to depend on error signaling [30] with effects not on synaptic learning but on neuronal gain.

Results show that feedback-dependent modulation of neuronal gain of punished populations of neurons changes the probabilities of activating the branches of states coding for actions. Three main cases are observed depending on the level of punishment:

- In the absence of punishment the gain is the same in all units in the two branches. In that case, the stronger connection between units 4 and 5 drives the network behavior. It induces more frequent activation of branch 1 by activating pattern D then pattern E after the initial sequence A-B-C (Figure 1B). The model classically selects more frequently the branch with stronger synaptic efficacy with the branching node, assumed to encode the most rewarded branch. This correspond to an exploitation strategy to increase the probability of reward by selecting the most rewarded action [8, 11].
- After punishment of medium intensity, the gain is decreased in the units active at time of punishment (here units 5 and 6 coding for pattern E). The assignment of reward or punishment affects neurons active in the branch selected by the network at the time of feedback [4], here by changing their gain.

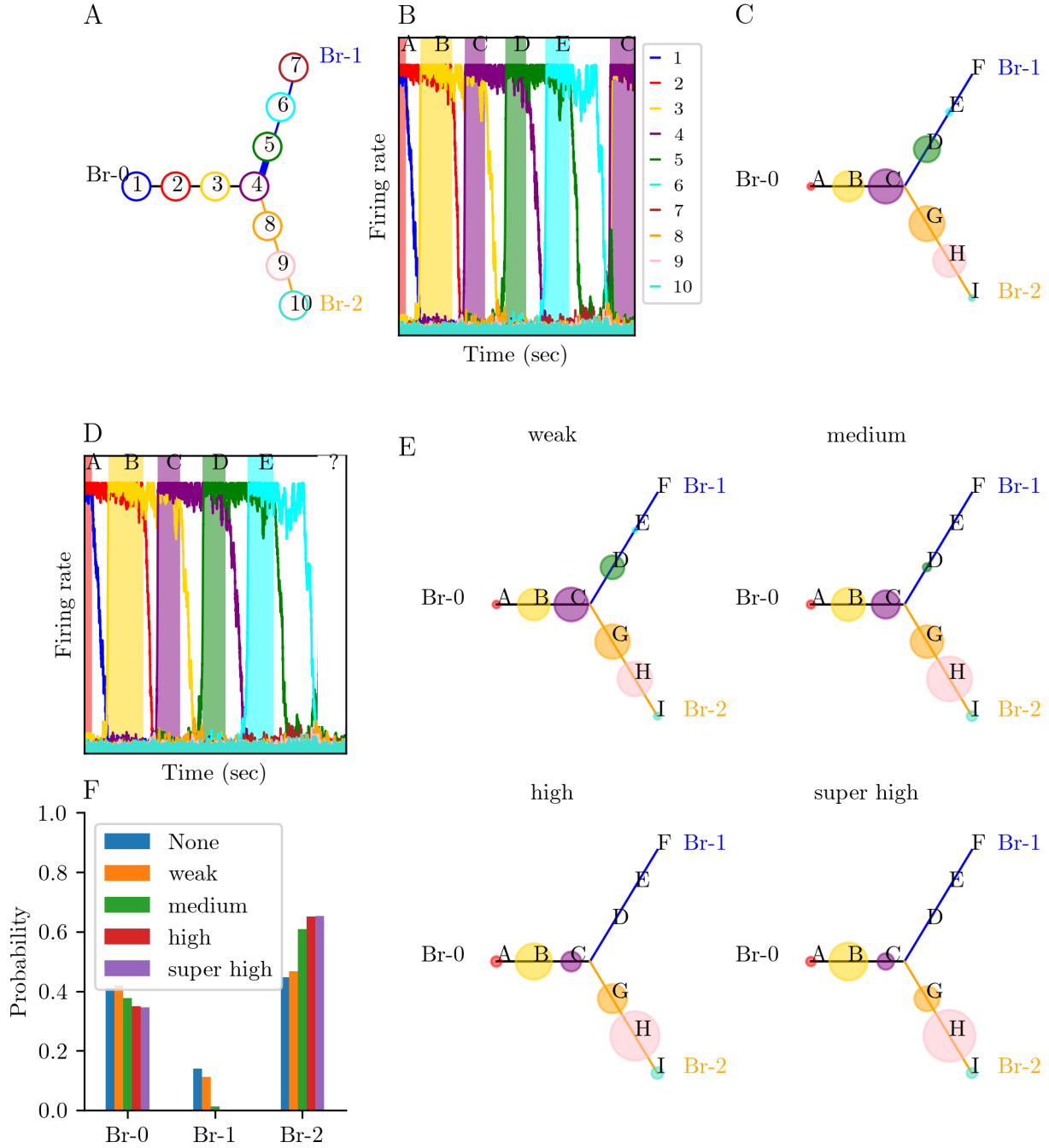


Figure 1: (A) Regular behavior of a system of $N=10$ units encodes 9 patterns (A, B, C . . . I) with a branching node at unit 4. Each numbered circle represents a unit. Two consecutive units encode a pattern. Unit 4 at the branching point encodes patterns C, D, and G. The synaptic efficacy between units 4 and 5 is 10% stronger than between units 4 and 8. When the system starts from pattern A on the initial branch (Br-0), it follows the sequence A-B-C and usually continues with DE on branch 1 (Br-1), thanks to the stronger connection between units 4 and 5 (B). After such a regular sequence, it randomly activates a learned pattern (C), which is pattern C in panel (B). The size of the circles around the patterns on the graph is proportional to their probability of activation. If pattern E is punished (D), the random activation after this pattern depends on the punishment rate (E). When the punishment rate is weak (10% of decrease in neuronal gain), the system can still activate patterns D or E but in only in 11% of the trials. For a medium punishment rate (50% decrease of neuronal gain), this ratio decreases to 1%. Whereas for a high punishment rate (66% decrease of neuronal gain), the system does not activate patterns along branch 1 anymore. The system instead activates patterns along branch 2 with a 40% increase from weak to strong punishment. The size of the circles around the pattern codes is proportional to the activation probability. The bar plot (F) combines the probability of activated patterns in panels (C) and (E), where ‘none’ stands for no punishment case in panel (C). 1000 simulations were performed for each parameter combination.

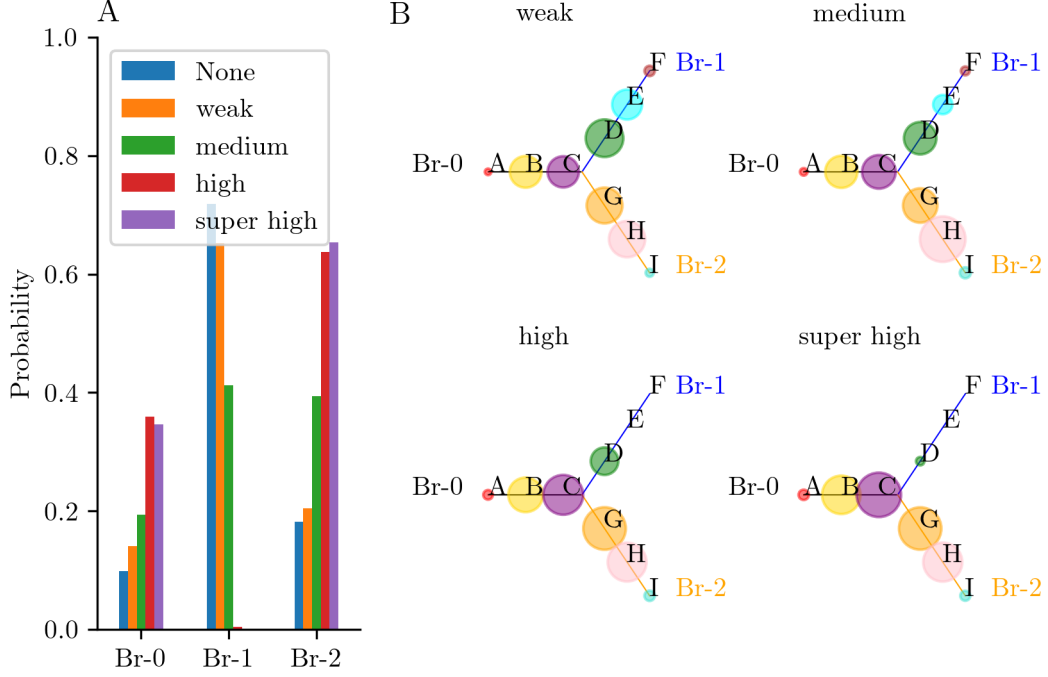


Figure 2: Trial $N+1$. (A) Probability of possible regular behaviors when the system is initiated from pattern A. System either stays on branch 0 (Br-0), takes branch 1 (Br-1) or branch 2 (Br-2). Without any punishment ('None'), the system takes Br-2 in 71% of the trials. For a weak punishment rate, the system activates Br-1 in 65% of trials and Br-1 in 20% of trials. Moderate punishment equalizes the preference (41% for branch 1 and 39% for branch 2). System takes Br-2 for high punishment rates (0% for branch 1 vs 64% for branch 2). (B) Activated patterns after the regular sequences on panel (A). The activation probability is indicated by the circle size around the pattern names (letters A,..., I). Increasing the punishment rate decreases the probability of patterns D and E being activated. 1000 simulations were performed for each parameter combination.

Increasing the intensity of punishment decreases the probability of recalling patterns D or E in the punished branch 1. This occurs immediately during the punished trial N (Fig 1E) and in the following trial $N + 1$ (Fig 2B). After weak punishment, the system still activates patterns D or E in the punished branch 1 since it is more synaptically associated to the branching node (unit 4) (Fig 2A). Increasing punishment immediately switches the choice of the system to the patterns on branch 2 (Figs 1F and on the following trial 2B). This is due to the fact that a low gain in the punished branch makes neurons populations less responsive to input activity coming from the initial branch 0 (patterns A-B-C), despite the stronger synaptic efficacy between the punished branch 1 and the branching node. In that case the punished branch is less activated and the probabilities to select the two branches are re-balanced. This correspond to an exploration strategy to search for rewarded actions when no specific action brings more rewards or punishments [8, 11].

- After strong punishment, the gain is strongly decreased in punished units 5 and 6. After the first regular sequence A-B-C (Fig 2A), the activation of patterns along branch 1 is stopped immediately at trial N (Fig 1F) and also at trial $N + 1$ (Fig 2B), even though it was initially the most frequently activated because the most strongly (synaptically) associated to the branching node.

The model shows that modulation of neuronal gain modifies the network behavior between a synaptic drive of the branch activated (no punishment), a switch of the ratio of activation of the punished vs. unpunished branches to a balanced activation (medium punishment), and an immediate blocking of the punished branch 1 (strong punishment). These effects are very similar during the punished trial N and on the following trial $N + 1$.

Interestingly, changes in branch activation can be generated by different processes of sequential activation in the phase space of the network. When arriving at the punished state D, activation can either

- go back to the starting branch 0 coding for the context (mainly pattern B),
- stay at the pattern preceding the punished pattern (pattern C),
- or jump directly to the unpunished branch 2 (patterns G or H).

We identify here elementary navigation processes within the network's states space (*building blocks* [18]) that depend on modulation of neuronal gain by feedback. Such processes allow to adapt the network behavior to punishment immediately during the punished trial N and also on the following trial $N + 1$. The immediacy of the switch in action selection allows to

avoid statistical trial-and-error relearning that would involve further errors and strong punishments. It then prevents from repeating a harmful error by switching action after only one strongly punished error. The present model identifies different types of navigation processes that depend on neuronal gain after punishment. We hope that it can provide a framework for modeling and experimental approaches investigating the effects of punishment on gain modulation and action without synaptic relearning.

4 Conclusion

From a learning point of view, the fact that gain modulation changes the network behavior depending on feedback indicates that gain embeds knowledge on the relation between actions and feedback. In that sense the value of neuronal gain contributes to store memories of past experiences [31]. This form of gain-based neural learning could act as a complement to synaptic learning and contribute to the alternation between exploitation and exploration strategies. Synaptic learning would allow knowledge to change rapidly and/or slowly depending on the volatility of the environment [16]. This knowledge is stable in the absence of environmental change and induces a gradual forgetting of previous knowledge in a changing environment, and would not be the only way of storing knowledge [1]. Gain-based neural learning, on the other hand, enables knowledge to be transiently shifted in memory without altering knowledge previously stored in the synapses. This capacity enables rapid adaptation to sudden changes in the potentially dangerous environment. It requires no repetition of errors for statistical learning, and allows a return to initial knowledge without relearning. Gain modulation could then allow the system to rapidly change behavior and provide it the necessary time to relearn by minimizing errors.

Acknowledgements

The authors gratefully thank Pascal Benquet and Gianluigi Mongillo for insightful discussions on a previous version of this article.

References

- [1] W. C. Abraham, O. D. Jones, and D. L. Glanzman. Is plasticity of synapses the mechanism of long-term memory storage? *NPJ science of learning*, 4(1):9, 2019.
- [2] C. Aguilar, P. Chossat, M. Krupa, and F. Lavigne. Latching dynamics in neural networks with synaptic depression. *PLOS ONE*, 12(8):e0183710, Aug. 2017.
- [3] S. Amari. Characteristics of random nets of analog neuron-like elements. *IEEE Transactions on systems, man, and cybernetics*, 5:643–57, 1972.
- [4] W. F. Asaad, P. M. Lauro, J. A. Perge, and E. N. Eskandar. Prefrontal neurons encode a solution to the credit-assignment problem. *Journal of Neuroscience*, 37(29):6995–7007, 2017.
- [5] T. E. Behrens, M. W. Woolrich, M. E. Walton, and M. F. Rushworth. Learning the value of information in an uncertain world. *Nature neuroscience*, 10(9):1214–1221, 2007.
- [6] A. Bernacchia, H. Seo, D. Lee, and X.-J. Wang. A reservoir of time constants for memory traces in cortical neurons. *Nature neuroscience*, 14(3):366–372, 2011.
- [7] M. Carandini and D. J. Heeger. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62, 2012.
- [8] J. D. Cohen, S. M. McClure, and A. J. Yu. Should i stay or should i go? how the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481):933–942, 2007.
- [9] G. S. Corrado, L. P. Sugrue, H. S. Seung, and W. T. Newsome. Linear-nonlinear-poisson models of primate choice dynamics. *Journal of the experimental analysis of behavior*, 84(3):581–617, 2005.
- [10] B. Doiron, A. Litwin-Kumar, R. Rosenbaum, G. K. Ocker, and K. Josić. The mechanics of state-dependent neural correlations. *Nature neuroscience*, 19(3):383–393, 2016.
- [11] P. Domenech, S. Rheims, and E. Koehlin. Neural mechanisms resolving exploitation-exploration dilemmas in the medial prefrontal cortex. *Science*, 369(6507):eabb0184, 2020.
- [12] E. Eldar, J. D. Cohen, and Y. Niv. The effects of neural gain on attention and learning. *Nature neuroscience*, 16(8):1146–1153, 2013.
- [13] K. A. Ferguson and J. A. Cardin. Mechanisms underlying gain modulation in the cortex. *Nature Reviews Neuroscience*, 21(2):80–92, 2020.

- [14] S. Fusi, W. F. Asaad, E. K. Miller, and X.-J. Wang. A neural circuit model of flexible sensorimotor mapping: learning and forgetting on multiple timescales. *Neuron*, 54(2):319–333, 2007.
- [15] K. Iigaya. Adaptive learning and decision-making under uncertainty by metaplastic synapses guided by a surprise detection system. *Elife*, 5:e18073, 2016.
- [16] K. Iigaya, Y. Ahmadian, L. P. Sugrue, G. S. Corrado, Y. Loewenstein, W. T. Newsome, and S. Fusi. Deviation from the matching law reflects an optimal strategy involving learning over multiple timescales. *Nature communications*, 10(1):1466, 2019.
- [17] E. Köksal Ersöz, C. Aguilar, P. Chossat, M. Krupa, and F. Lavigne. Neuronal mechanisms for sequential activation of memory items: Dynamics and reliability. *PLOS ONE*, 15(4):e0231165, 2020. Publisher: Public Library of Science.
- [18] E. Köksal Ersöz, P. Chossat, M. Krupa, and F. Lavigne. Dynamic branching in a neural network model for probabilistic prediction of sequences. *Journal of Computational Neuroscience*, Aug. 2022.
- [19] F. Lavigne and N. Darmon. Dopaminergic neuromodulation of semantic priming in a cortical network model. *Neuropsychologia*, 46:3074–87, 2008.
- [20] L. Lazartigues, F. Mathy, and F. Lavigne. Statistical learning of unbalanced exclusive-or temporal sequences in humans. *PLoS ONE*, 16(2):e0246826., 2021.
- [21] L. Lazartigues, F. Mathy, and F. Lavigne. Probability, dependency, and frequency are not all equally involved in statistical learning. *Experimental Psychology*, 2023.
- [22] I. Lerner, S. Bentin, and O. Shriki. Excessive attractor instability accounts for semantic priming in schizophrenia. *PLoS ONE*, 7:e40663, 2012.
- [23] M. R. Nassar, K. M. Rumsey, R. C. Wilson, K. Parikh, B. Heasley, and J. I. Gold. Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature neuroscience*, 15(7):1040–1046, 2012.
- [24] M. R. Nassar, R. C. Wilson, B. Heasley, and J. I. Gold. An approximately bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience*, 30(37):12366–12378, 2010.
- [25] R. A. Rescorla. Variation in the effectiveness of reinforcement and nonreinforcement following prior inhibitory conditioning. *Learning and motivation*, 2(2):113–123, 1971.
- [26] A. Rey, J. Fagot, F. Mathy, L. Lazartigues, L. Tosatto, G. Bonafos, J.-m. Freyermuth, and F. Lavigne. Learning higher-order transitional probabilities in nonhuman primates. *Cognitive Science*, 46(4):e13121, 2022.
- [27] E. Salinas and P. Thier. Gain modulation: a major computational principle of the central nervous system. *Neuron*, 27(1):15–21, 2000.
- [28] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- [29] M. V. Tsodyks and H. Markram. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proceedings of the National Academy of Sciences*, 94:719–23, 1997.
- [30] S. Yagishita. Cellular bases for reward-related dopamine actions. *Neuroscience Research*, 188:1–9, 2023.
- [31] W. Zhang and D. J. Linden. The other side of the engram: experience-driven changes in neuronal intrinsic excitability. *Nature Reviews Neuroscience*, 4(11):885–900, 2003.