Modeling absolute and relative familiarity signals through different learning rules

William Warnier[1], John Read[1,2], Emma Delhaye[1,2,3] & Jacques Sougné[2,4]

[1]GIGA Research, GIGA-CRC Human Imaging, University of Liège
[2]Psychology & Neuroscience of Cognition Research Unit, University of Liège
[3]CICPSI, Faculty of Psychology, University of Lisbon
[4]UDI FPLSE, University of Liège

william.warnier@student.uliege.be

Introduction

Familiarity is an automatic signal that supports recognition without recalling contextual details. Two types of familiarity can be distinguished[1]:

- **Absolute** : long-term memory trace from accumulated prior exposure.
- **Relative** : signal generated through recent exposure.

However, their respective contribution to the conscious experience of familiarity remains unclear.

In this study, we implemented two Hebbian-inspired learning rules using artificial neural networks modeling absolute and relative familiarity, respectively. We compared models' behavior using three datasets with varying familiarity level before learning.

→ We hypothesize that these rules will lead to different types of changes in models' familiarity signal distributions depending on baseline familiarity levels.

Methods

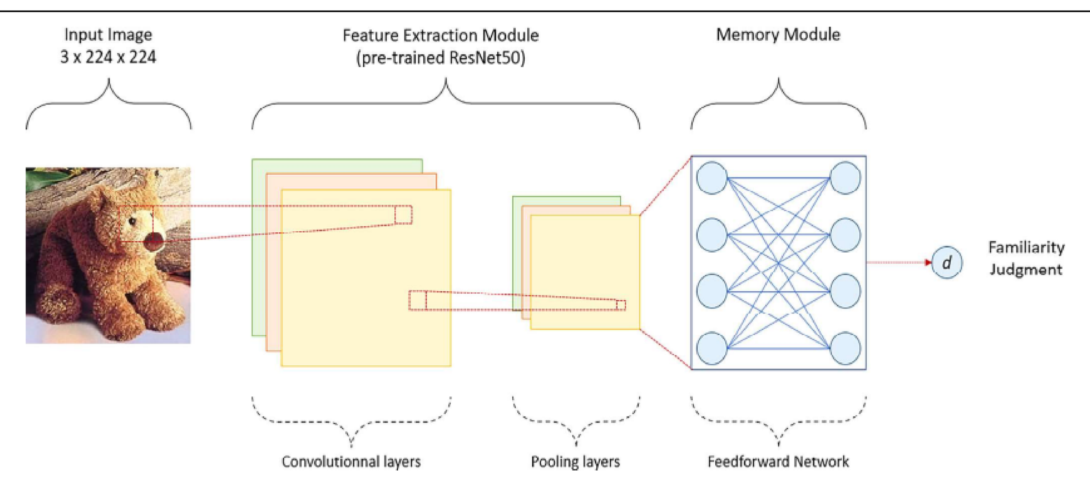


Figure 1: Full architecture of the models. ResNet50[2] as feature extraction module followed by two fully connected layers as memory module.

We implemented two learning rules for the memory module :

- Hebbian learning (strenghtening connections for active neurons)
- Anti-Hebbian learning (weakening connections for active neurons)

We manipulated baseline familiarity during simulations using three datasets:

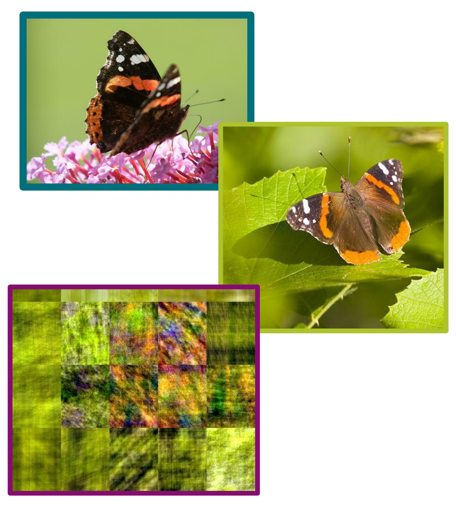


Figure 2: Example of images from the three datasets.

- High Familiarity: 997 images from ImageNet's training set (seen by ResNet50 during its pre-training);
- Low Familiarity: 965 images from ImageNet's validation set (not seen by ResNet50);
- No Familiarity: 965 phase-scrambled images devoid of familiarity ;

One run of a simulation for a model using a specific dataset took place as follows:

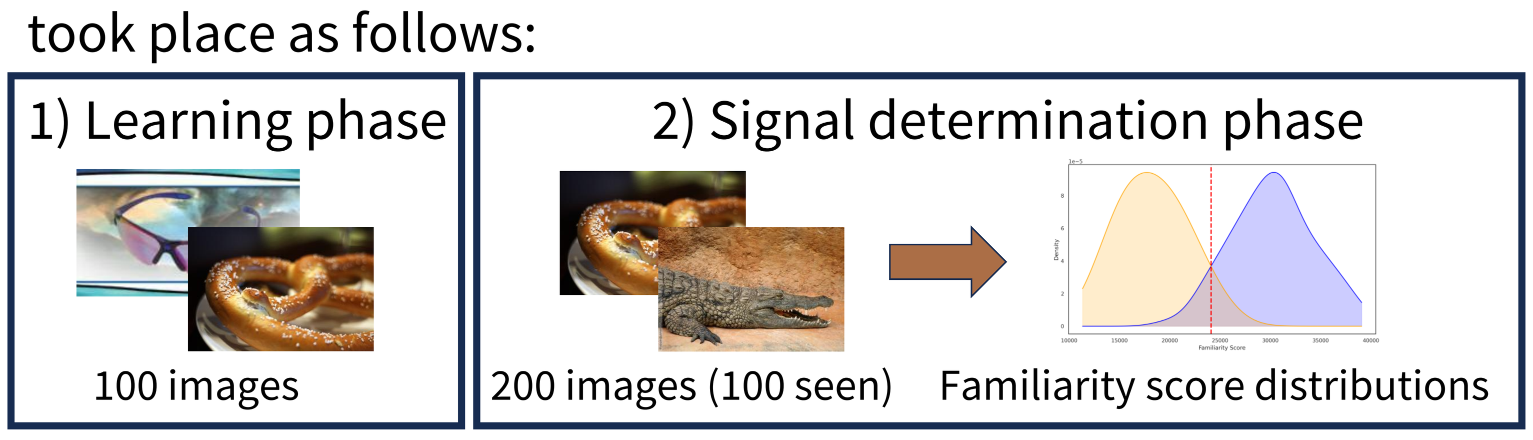


Figure 3: Visual representation of all the phases taking place during a simulation.

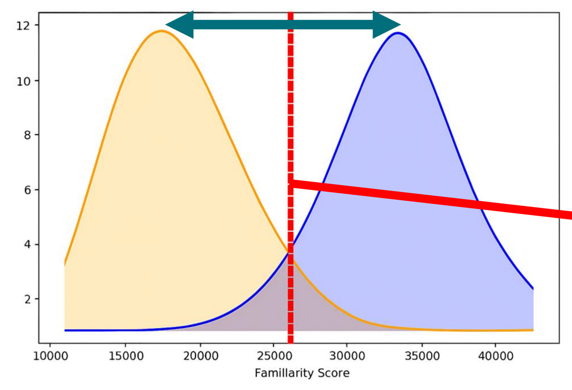
Familarity scores (*fam*) for learned (X) and novel (Z) stimuli are computed using the following formula:

$$fam(X,Z) = \sum_{j=1}^{m} x_j h_j$$

where m is the number of neurons, x_j and h_j are respectively the input and output of the fully-connected layer.

Results

Two metrics were extracted based on familiarity scores distributions.



Separation Index = (mean_dx - mean_dz) / pooled_std

Decision Threshold set at curves intersection

Figure 4: Graphical representation of the two main extracted metrics.

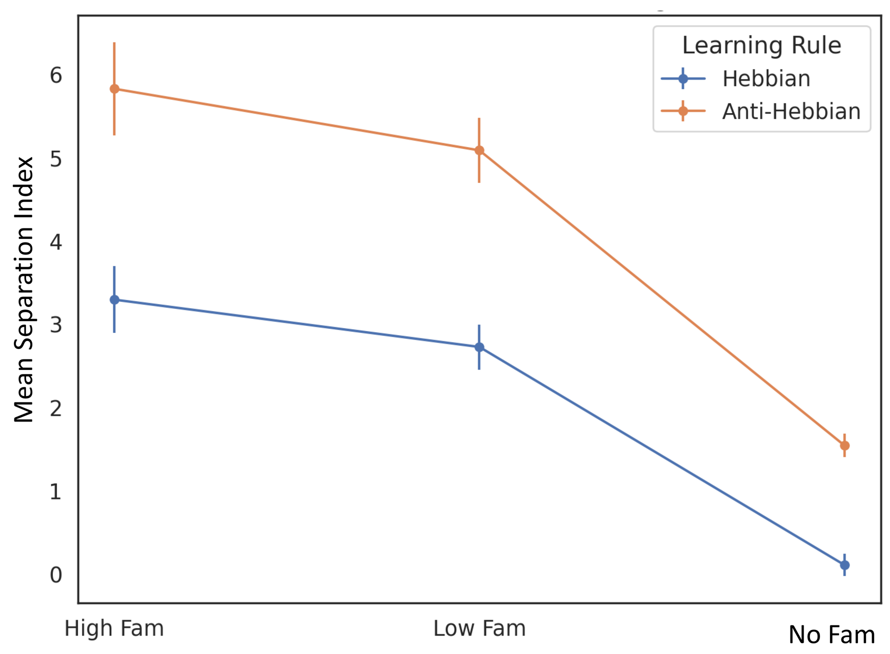


Figure 5: Mean Separation Index accross datasets for both models. Bars show standard error.

Scores were averaged across 100 runs. Both models showed similar pattern of results, that is a decrease in the mean distribution Separation Index accross the three levels of familiarity.

The Hebbian model exhibited a higher Decision Threshold (needing less signal strength for recognition decision) in the high compared to low familiarity condition while the Anti-Hebbian threshold seemed stable.

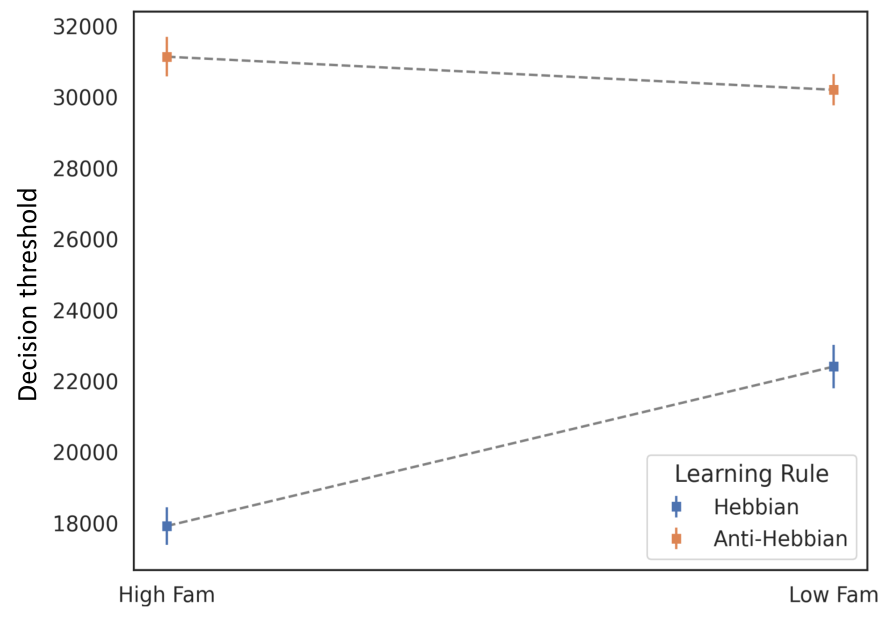


Figure 6: Mean Decision Threshold (distributions intersection) for high and low familiarity datasets.

Discussion

Although they were computed based on familiarity scores distributions, both extracted indexes shared similarites with Signal Detection Theory (SDT) indexes: *d'* and Criterion, respectively.

Our results showed that both Hebbian and Anti-Hebbian models exhibit lower perceptual sensitivity (*d'*) in low familiarity conditions. This suggests that both learning rules are influenced by the absolute familiarity level of input images. However, only the Hebbian model showed a clear modulation of its decision threshold towards a "more liberal Criterion" in the high familiarity condition whereas the Anti-Hebbian model maintained stable thresholds.

This analysis further aligns with the theoretical framework proposed by Read et al.[3] in which Hebbian learning is more sensitive to long-term exposure, and thus more likely to adjust its internal decision boundary based on the amount of absolute familiarity. In contrast, Anti-Hebbian learning is thought to support familiarity signals derived from recent stimuli exposure without the need to adjust its Criterion.

→ More in-depth analyses are needed as both models' seem nevertheless influenced by baseline familiarity.

References

1. Mandler, G. (1980). Recognizing : The judgment of previous occurrence. *Psychological Review*, *87*(3), 252-271. https://doi.org/10.1037/0033-295X.87.3.252
2. He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep Residual Learning for Image Recognition*. 770-778. https://doi.org/10.1109/CVPR.2016.90
3. Read, J., Delhaye, E., & Sougné, J. (2024). Computational models can distinguish the contribution from different mechanisms to familiarity recognition. *Hippocampus*, *34*(1), 36-50. https://doi.org/10.1002/hipo.23588

LIÈGE université
GIGA
CRC Human Imaging