
# Brain-Inspired Multimodal Deep Learning

21/11/2023

Funded by
the European Union

European Research Council
Established by the European Commission

**Rufin VANRULLEN**

*CerCo, CNRS*

ANITI

Université de Toulouse

# PLAN: overview of past and future work

**2019-2023: Chair "Deep Learning with semantic, cognitive and biological constraints"**

With co-chairs C. Braud (IRIT), L. Reddy (CerCo), F. Filbet & G. Faye (IMT)

1. brain decoding with deep generative networks

2. brain-inspired recurrent networks with Predictive Coding
   - more efficient, more robust
   - more aligned with human vision (illusions)
   - detailed mathematical characterization (convergence, oscillations…)

3. brain-inspired multimodal systems based on Global Workspace (ERC GLOW 2023-2028)

**2024-2028? Synergy Chair C3-PO: "Cobots with Conversation, Cognition and Perception"**

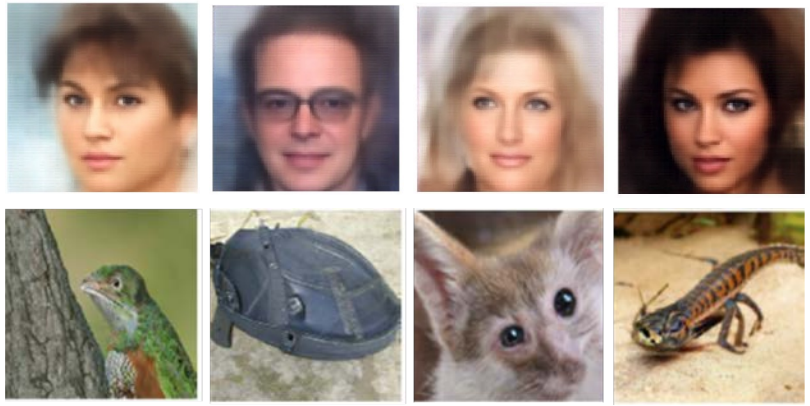
Chairs: N. Asher (IRIT), T. Serre (Brown U.), O. Stasse (LAAS), R. VanRullen (CerCo)

# 1. Brain decoding with deep generative networks

VanRullen & Reddy (2019) *Communications Biology*
Mozafari, Reddy & VanRullen (2020) *IJCNN'20*
Ozcelik et al (2022) *IJCNN'22*
Ozcelik & VanRullen (2023) *Scientific Reports*

images seen by subject in MRI

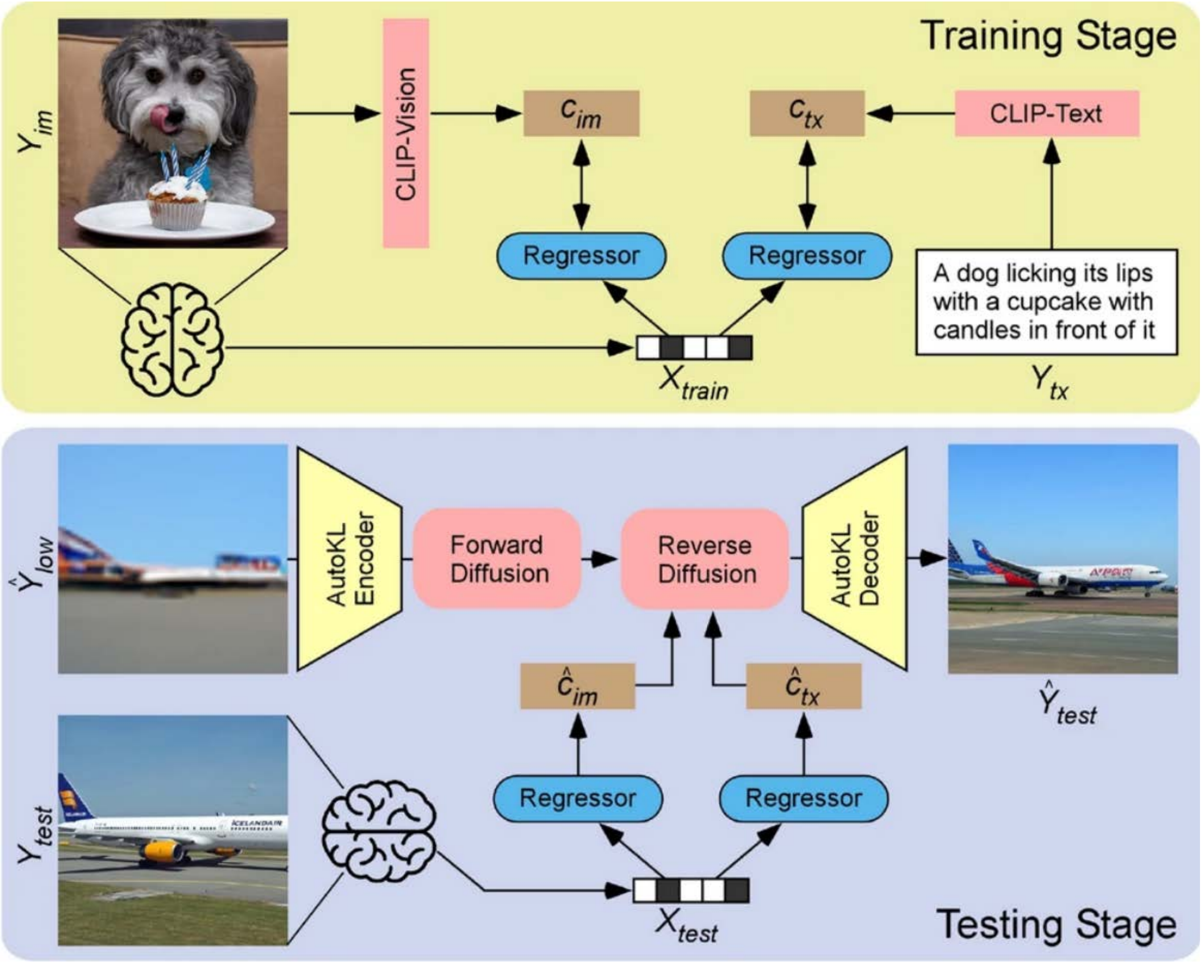
reconstructed images from fMRI signals

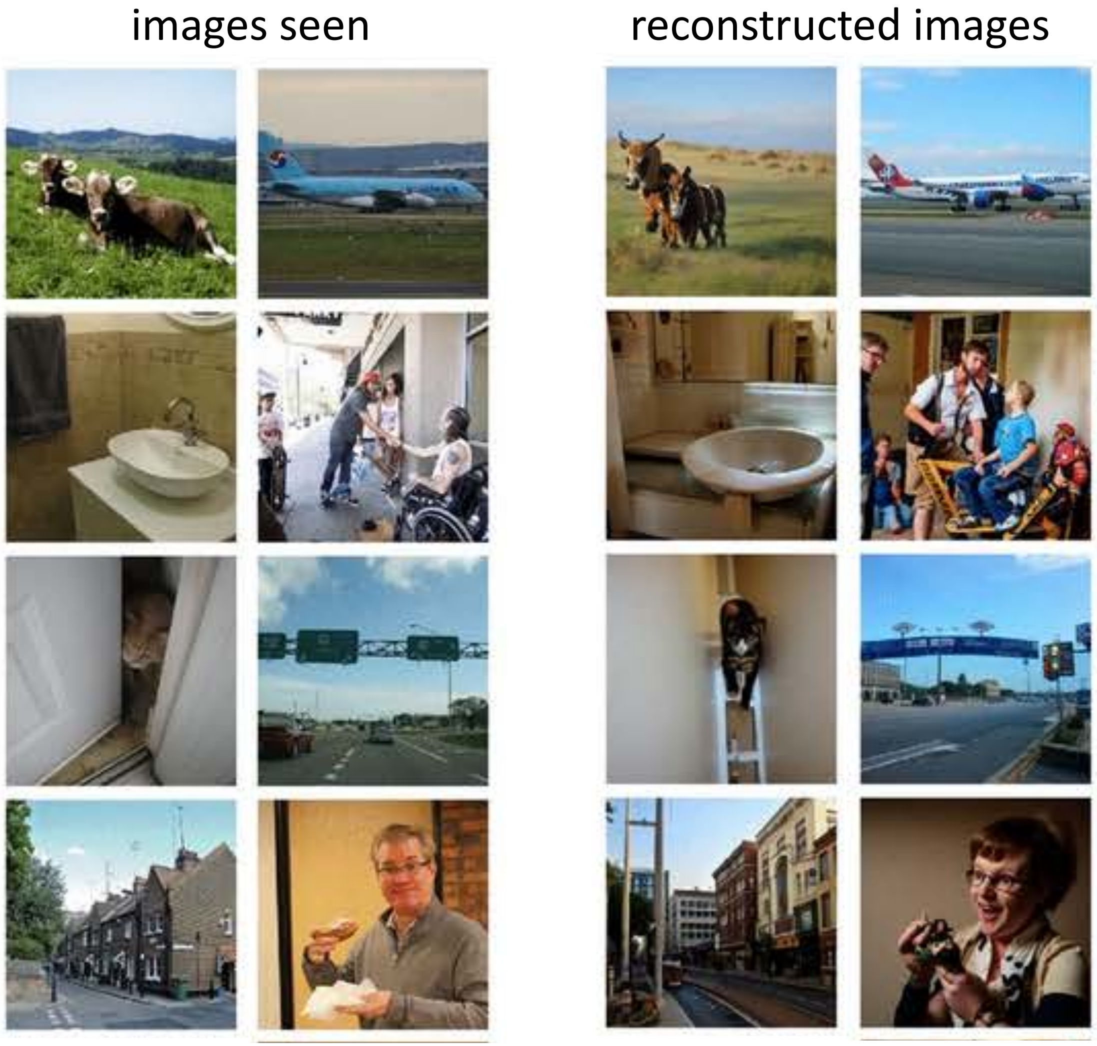


# 1. Brain decoding with deep generative networks

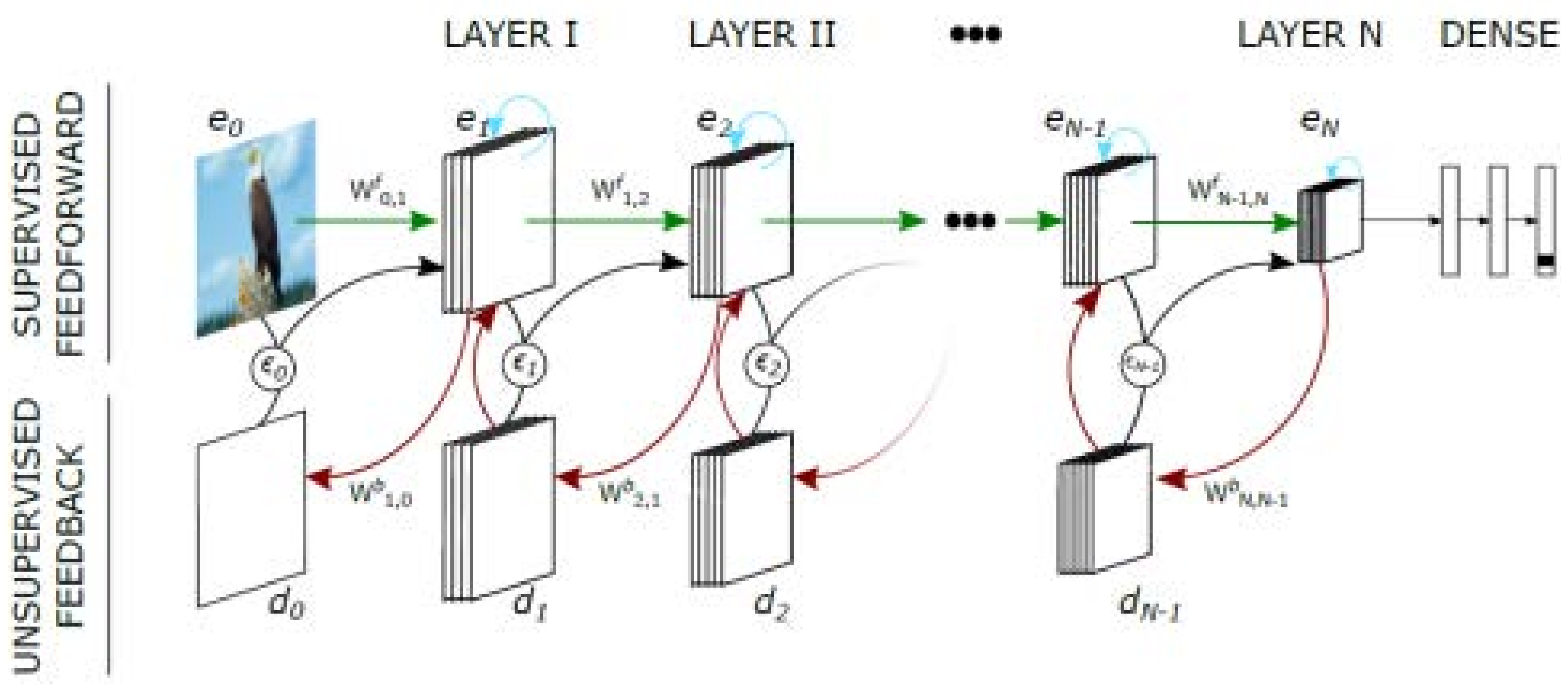
## Brain-Diffuser

Ozcelik & VanRullen (2023) *Scientific Reports*



# 2. Predictive coding




$$\underbrace{}_{}$$

input drive | top-down error correction | memory | bottom-up error correction

$$e_n(t+1) = \overbrace{\beta\left[W^f_{n-1,n}e_{n-1}(t+1)\right]_+}^{\text{input drive}} + \overbrace{\lambda d_n(t)}^{\text{top-down error correction}} + \overbrace{(1-\beta-\lambda)e_n(t)}^{\text{memory}} - \overbrace{\alpha\nabla\epsilon_{n-1}(t)}^{\text{bottom-up error correction}}$$

$$d_n(t) = \left[W^b_{n+1,n}e_{n+1}(t)\right]_+$$

$$\epsilon_{n-1}(t) = \|e_{n-1}(t) - d_{n-1}(t)\|_2^2$$

# 2. Predictive coding

- ***Predify*** software (for PyTorch):

  https://github.com/miladmozafari/predify

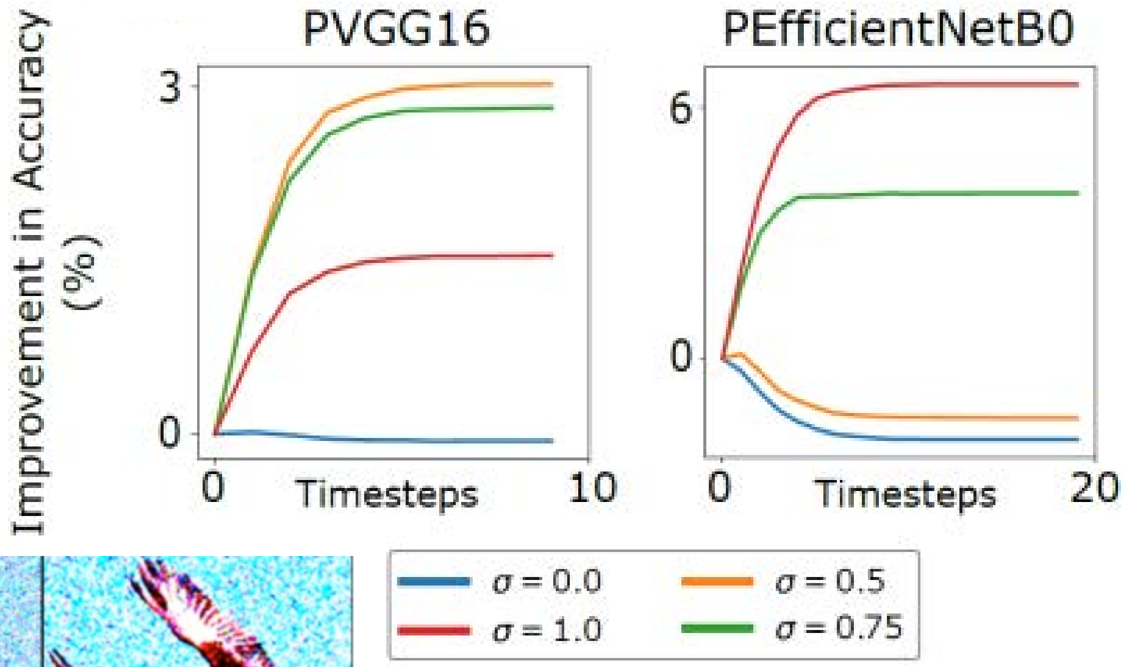
- We ***predified***: VGG16, EfficientNetB0, ResNet18...

$$
\begin{aligned}
e_n(t+1) &= \underbrace{\beta\left[W^f_{n-1,n} e_{n-1}(t+1)\right]_+}_{\text{input drive}} + \underbrace{\lambda d_n(t)}_{\text{top-down error correction}} + \underbrace{(1-\beta-\lambda)e_n(t)}_{\text{memory}} - \underbrace{\alpha \nabla \epsilon_{n-1}(t)}_{\text{bottom-up error correction}} \\
d_n(t) &= \left[W^b_{n+1,n} e_{n+1}(t)\right]_+ \\
\epsilon_{n-1}(t) &= \| e_{n-1}(t) - d_{n-1}(t) \|_2^2
\end{aligned}
$$

# ImageNet accuracy:

Choksi, B., Mozafari, M., Biggs O'May, C., Ador, B., Alamia, A., & VanRullen, R. (2021). *Advances in Neural Information Processing Systems*, *34*, 14069-14083.

Alamia, A., Mozafari, M., Choksi, B., & VanRullen, R. (2023). *Neural Networks*, *157*, 280-287.



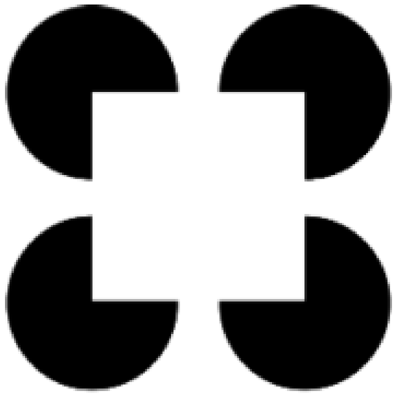

input t=10

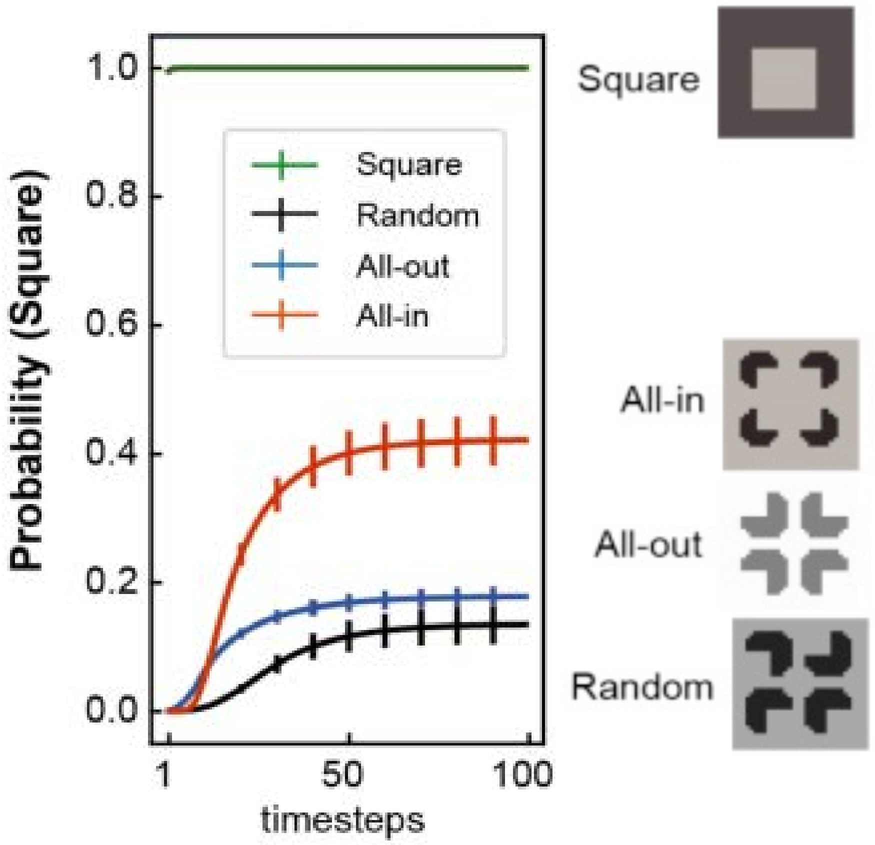
higher is better

- Robustness to Gaussian noise and other corruptions (ImageNet-C)
- Robustness to adversarial attacks

# 2. Predictive coding

- **“Perception” of illusory contours, just like humans**
  - Pang, Z., O’May, C. B., Choksi, B., & VanRullen, R. (2021). Predictive coding feedback results in perceived illusory contours in a recurrent neural network. *Neural Networks*, *144*, 164-175.

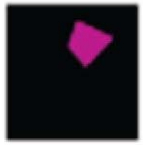




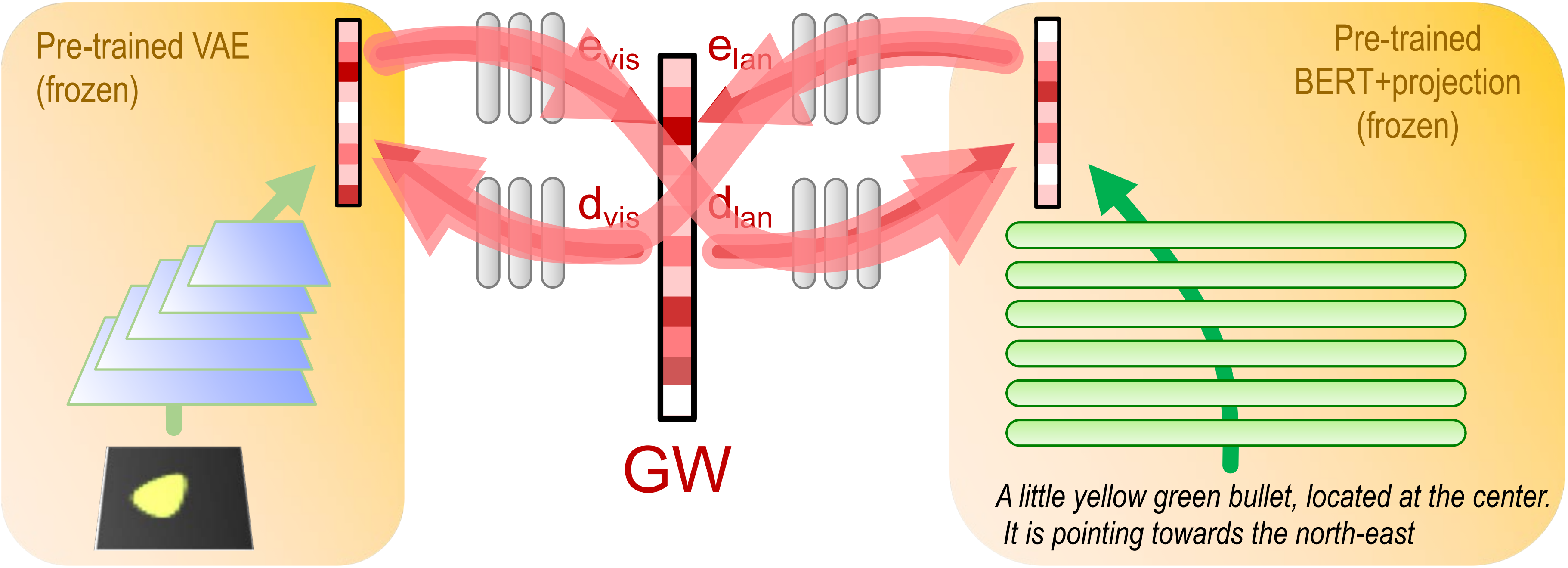

- **Mathematical analysis of convergence:**
  - Faye, G., Fouilhé, G., & VanRullen, R. (2023). Mathematical derivation of wave propagation properties in hierarchical neural networks with predictive coding feedback dynamics. *Bulletin of Mathematical Biology*.

**Multi-modal systems are getting very good: CLIP, DALL-E, GATO, Flamingo, PALM-E, etc.**

- ➔ Vision + Language + Action + …
- ➔ Require enormous models, gigantic supervised training sets…
- ➔ How does the brain do it ?
- ➔ **Global Workspace Theory**

| vision | language |
|---|---|
|  | A little yellow green bullet, located at the center. It is pointing towards the north-east. |
|  | It is a quite small trapezoidal shape in medium violet red color. It is located in the top center and pointing towards the right. |
|  | A medium-sized red triangle, in the bottom-left corner and rotated east. |

## Global Workspace (simplified) architecture

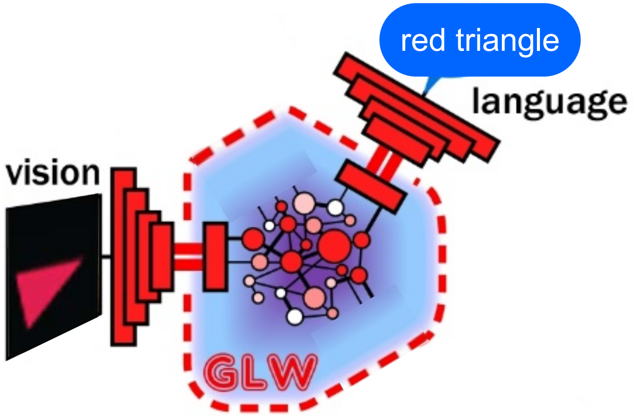


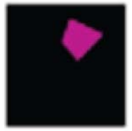

- **Training objectives:**
  - **Supervised: N matching pairs {vis$_i$, lan$_i$} — we vary N to assess the need for labels**
  - **Unsupervised: cycle-consistency (a.k.a. « broadcast »)**

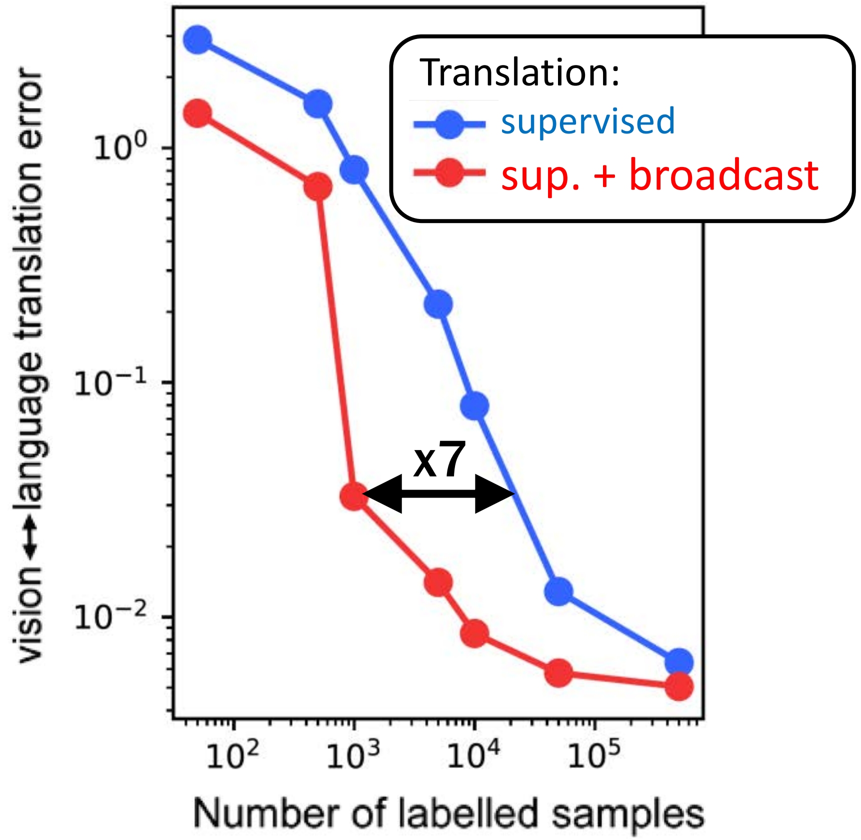
    **➔ only needs unpaired data!**

## Global Workspace (simplified) architecture

Devillers, Maytie & VanRullen (2023). *Semi-supervised multimodal representation learning through a Global Workspace*, arXiv



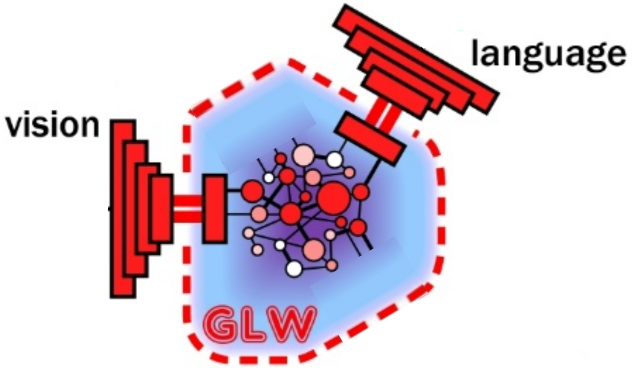

| vision | language |
|---|---|
|  | A little yellow green bullet, located at the center. It is pointing towards the north-east. |
|  | It is a quite small trapezoidal shape in medium violet red color. It is located in the top center and pointing towards the right. |
|  | A medium-sized red triangle, in the bottom-left corner and rotated east. |



# 3. Multimodal systems

## Global Workspace: where we are now…

B. Devillers




A bright red diamond, pointing to the bottom right
A green triangle at the top, towards left
An egg-shaped oval, dark-blue, pointing up, on the bottom-left of the image

Attributes = [1,0,0,-1,-0.5,-60,2,1,0,0]

shape position rotation size color

L. Maytie

| attributes | values |
|---|---|
| position x | 1.767 |
| position y | -0.853 |
| rotation z | -2.042 |
| color (R,G,B) | (1, 0.343, 0) |

| attributes | values |
|---|---|
| position x robot | -3.333 |
| position y robot | 3.101 |
| rotation z robot | -0.286 |
| position x table | 1.933 |
| position y table | 2.220 |
| rotation z table | -0.575 |
| color (R,G,B) | (1, 0, 0.107 |

sim2real (with LAAS)

| attributes | values |
|---|---|
| position x | -1.674 |
| position y | -0.099 |
| rotation z | 2.651 |
| pitch | -1.24 |
| roll | -0.232 |

# 3. Multimodal systems

## Global Workspace: where we are going…

Funded by the European Union

European Research Council
Established by the European Commission

### ERC Advanced Project GLOW (2023-2028)

- Develop brain-inspired multimodal deep learning systems
- Evaluate their use and relevance for machine learning
- Advance our knowledge of the brain

### ANITI Synergy Chair C3-PO (2024-2028)??

- Cobots with Conversation, Cognition & Perception
- Chairs: R.VanRullen (CerCo), N. Asher (IRIT),
  T. Serre (Brown), O. Stasse (LAAS)
- Frugal multimodal robotic systems with
  grounded perception, language and action
- Main industrial partners: Airbus, Linagora