

# Testing Integrated Information Theory predictions by assessing representational similarity in brain activity

Pablo Oyarzo<sup>1,2</sup>, Zvi Roth<sup>3</sup>, Oscar Ferrante<sup>4</sup>, Aya Khalaf<sup>5</sup>, Ling Liu<sup>6</sup>, David Richter<sup>7,8</sup>, Yamil Vidal<sup>7</sup>, Ole Jensen<sup>4,9</sup>, Huan Luo<sup>6</sup>, Floris P de Lange<sup>7</sup>, Hal Blumenfeld<sup>5</sup>, Lucia Melloni<sup>10,11</sup>, Michael Pitts<sup>12</sup>, Liad Mudrik<sup>13</sup>, Cogitate Consortium

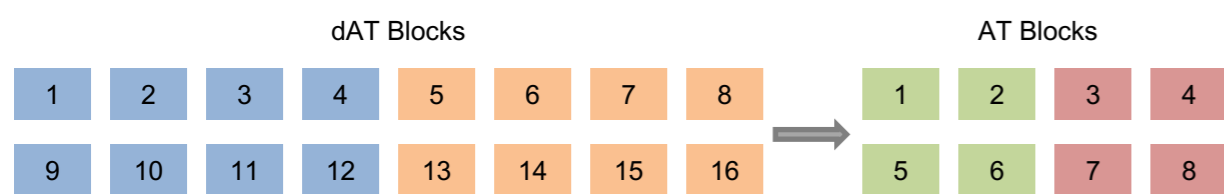
<sup>1</sup>Freie Universität Berlin, Germany; <sup>2</sup>Aarhus University, Denmark; <sup>3</sup>Bar-Ilan University, Israel; <sup>4</sup>University of Birmingham, UK; <sup>5</sup>Yale School of Medicine, USA; <sup>6</sup>Peking University, China; <sup>7</sup>Radboud University Nijmegen, The Netherlands; <sup>8</sup>Vrije Universiteit Amsterdam, The Netherlands; <sup>9</sup>University of Oxford, UK; <sup>10</sup>New York University, USA; <sup>11</sup>Max Planck Institute for Empirical Aesthetics, Germany; <sup>12</sup>Reed College, USA; <sup>13</sup>Tel Aviv University, Israel

## Introduction

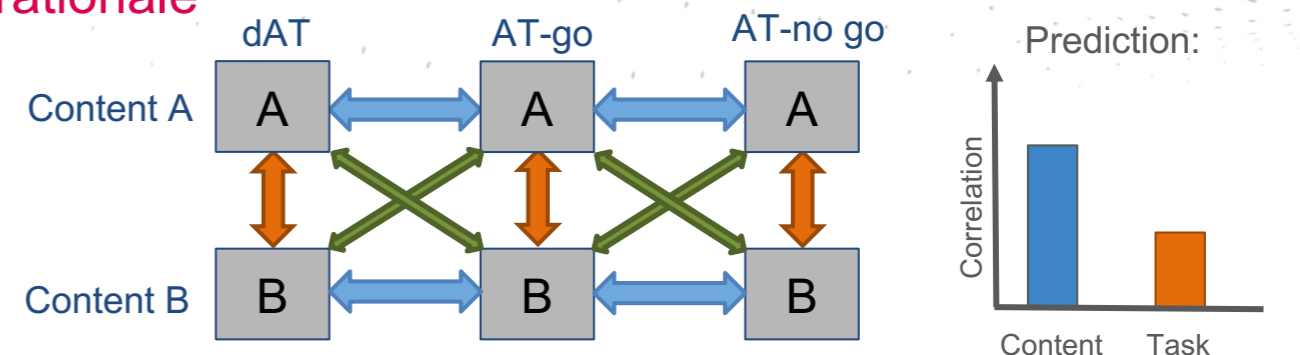
- In this analysis, we tested the critical prediction of the Integrated Information Theory (IIT) that conscious content is encoded in stable, content-specific patterns within posterior cortex, irrespective of task or attention [1].
- Here we present fMRI and MEG results for this prediction in Experiment 2 of the Cogitate Consortium [2].
- We tested whether neural similarity in these regions reflects shared conscious content (i.e., category and location) more strongly than task context.
- Participants viewed identical stimuli either while playing a video game (distracted attention, dAT) or performing a go/no-go detection task (focused attention, AT).

## Methods

### Structure of the tasks



### Analysis rationale



$$\Delta Sim_{content > task} = avg_{i,a \neq b} (corr(C_{i,a}, C_{i,b})) - avg_{i \neq j, a} (corr(C_{i,a}, C_{j,a}))$$

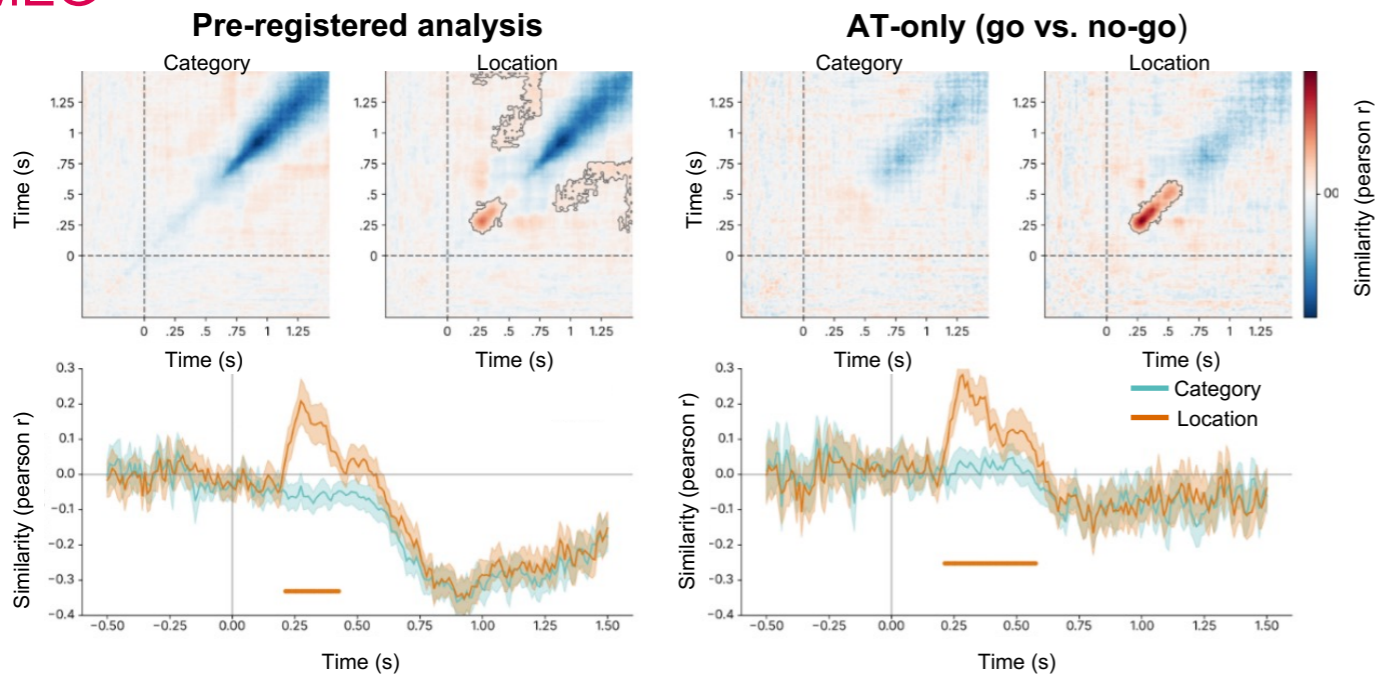
- Distracted Attention (dAT):** adaptive video game with stimuli appearing in the background.
- Focused Attended (AT):** replay of the video game with a go/no-go task for target stimuli.
- Category and location** analysed separately and tested at group level.
- MEG:** source space in posterior ROIs; **fMRI:** whole-brain searchlight.

- Pre-registered contrast:** same content across tasks vs. different content within task (i.e. if content > task).
- Control analyses:** (1) same vs. different content across tasks to confirm content decodability; (2) repetition of pre-registered analysis comparing early vs. late task partitions to rule out condition-order effects.

## Results

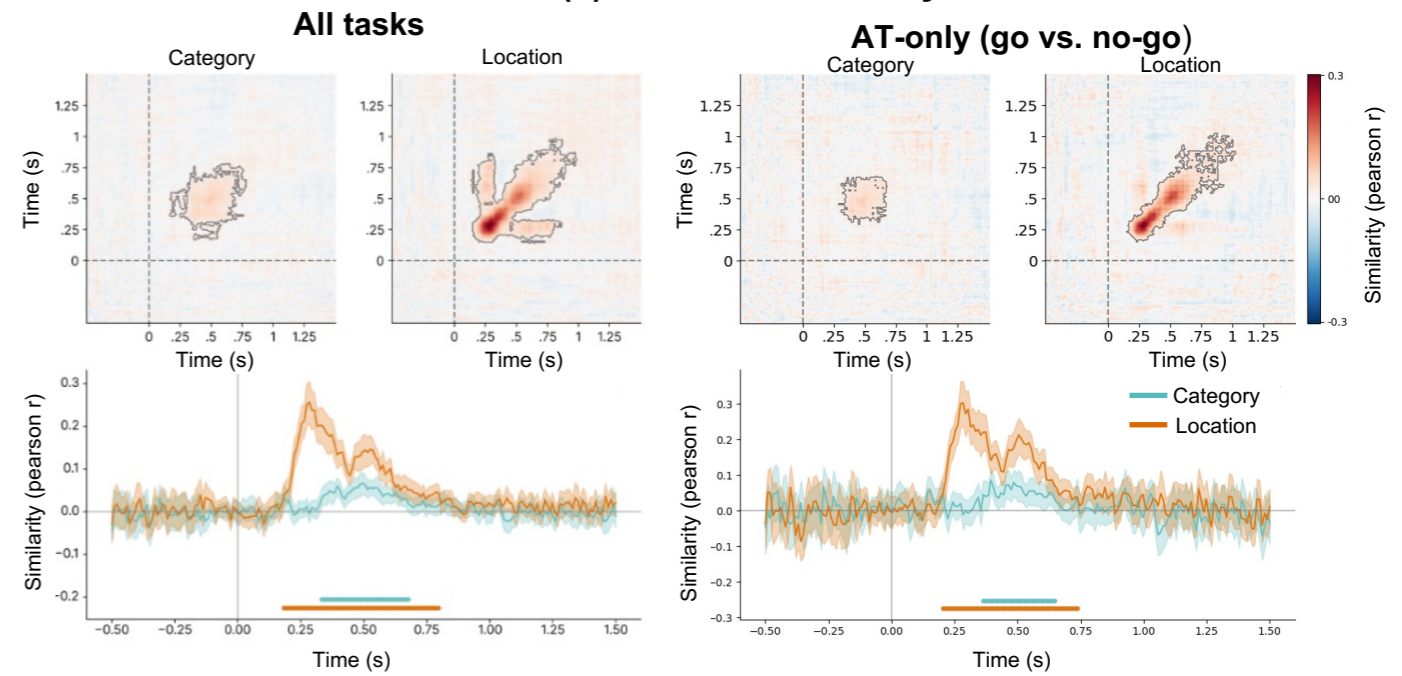
### MEG

#### Content-over-task similarity contrast



Content > task only for **location** (~210–410 ms) Effect replicated for **location** only.

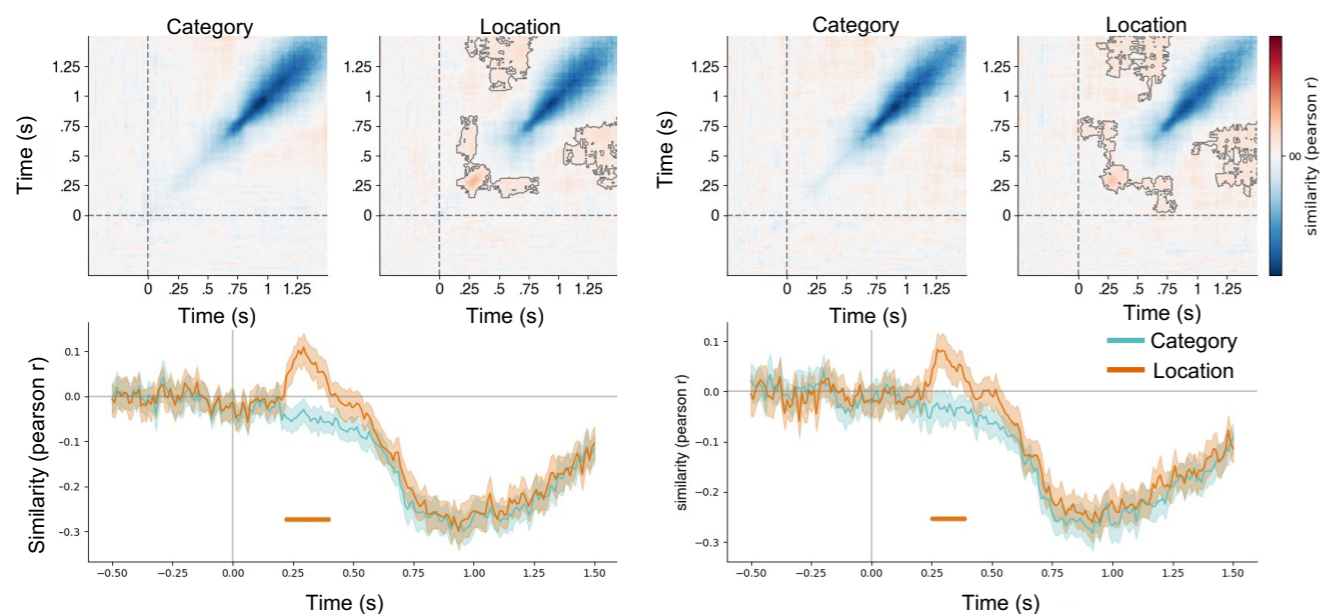
#### Control (1): Content similarity across tasks



Category representations robust across all tasks; confirms content detectability

#### Control (2): Comparison of temporal proximity

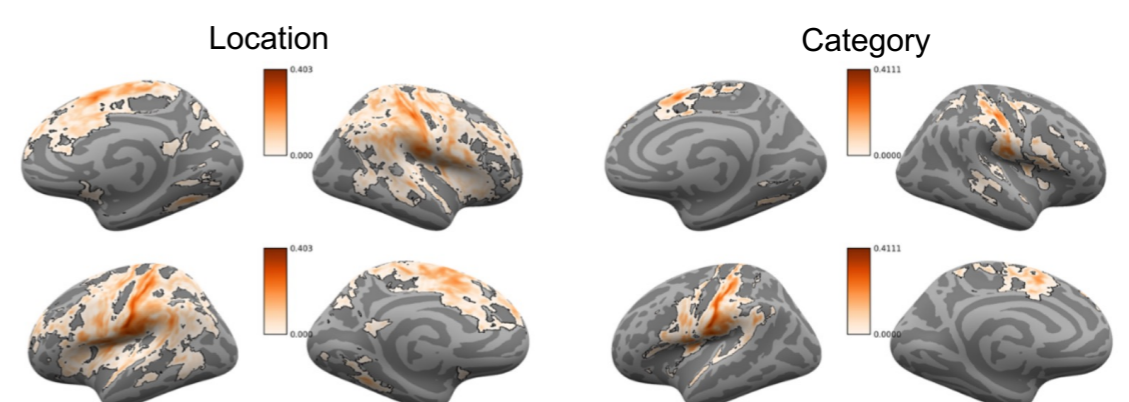
Near pair (dAT late ↔ AT early) Far pair (dAT early ↔ AT late)



Stable effect for **location** across near and far pairs; category was absent throughout

### fMRI

#### Significant Clusters in AT only: Task > Content



No fMRI cluster showed content > task, contrary to IIT prediction.

## Conclusions

- IIT's prediction that content similarity exceeds task similarity was partially supported: **confirmed for location, but not for category**.
- Follow-up analyses ruled out confounds (task proximity, detection sensitivity), strengthening the interpretation.
- The failure to confirm the full prediction, even under favorable conditions, poses a substantive challenge to IIT's assumption of content-dominant representations in posterior regions.

## Contact

Pablo Oyarzo  
pablo@cfin.au.dk

## Acknowledgements

COGITATE is supported by the Templeton World Charity Foundation Inc (TWCF0389).



## References

- Tononi, G., M. Boly, M. Massimini, and C. Koch, *Integrated information theory: from consciousness to its physical substrate*. Nature Reviews Neuroscience, 2016. 17(7): p. 450–461
- Melloni, L., L. Mudrik, M. Pitts, K. Bendtz, O. Ferrante, U. Gorska, et al., *An adversarial collaboration protocol for testing contrasting predictions of global neuronal workspace and integrated information theory*. Plos one, 2023. 18(2): p. e0268577



Pre-registration



arc-cogitate.com