The City College of New York

Real-Time Estimation of Overt Attention from Dynamic Features of the Face using Deep Learning

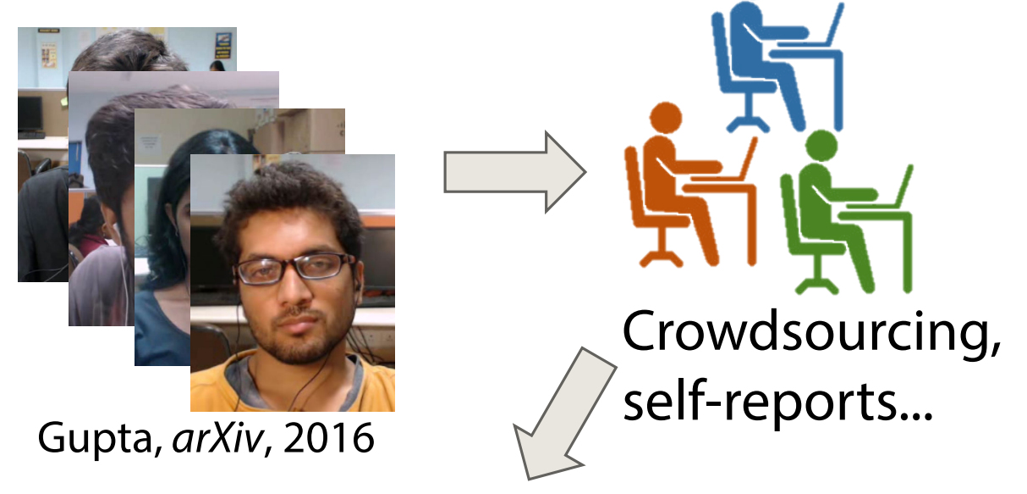
THE CITY UNIVERSITY OF NEW YORK

Aimar Silvan, Lucas C. Parra, Jens Madsen

- Department of Biomedical Engineering, City College of New York -

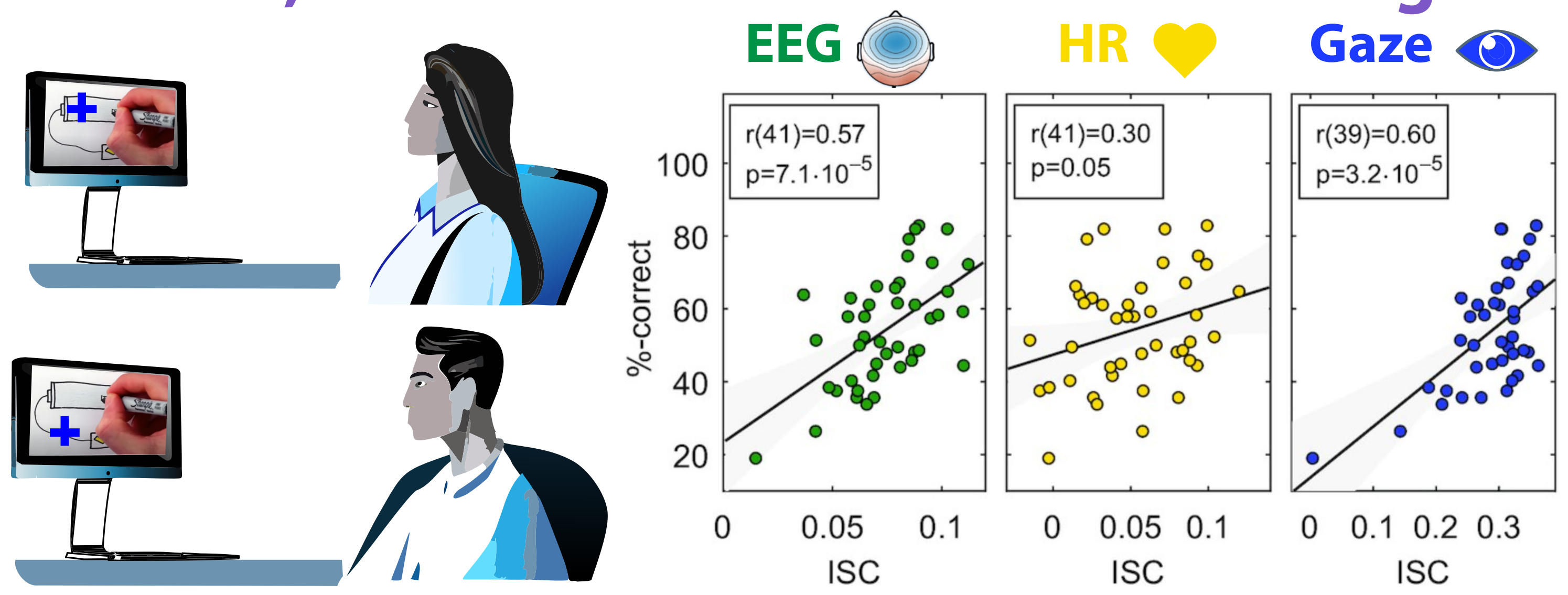
There is a need for measuring attention objectively

- **Challenge**: Measuring task engagement is crucial across neuroscience, education, and psychology.
- **Current AI Limitations:** Existing methods for attention estimation often rely on self-reports, subjective ratings, and distributing face videos, raising privacy concerns.


Gupta, *arXiv*, 2016

Engaged / Unengaged

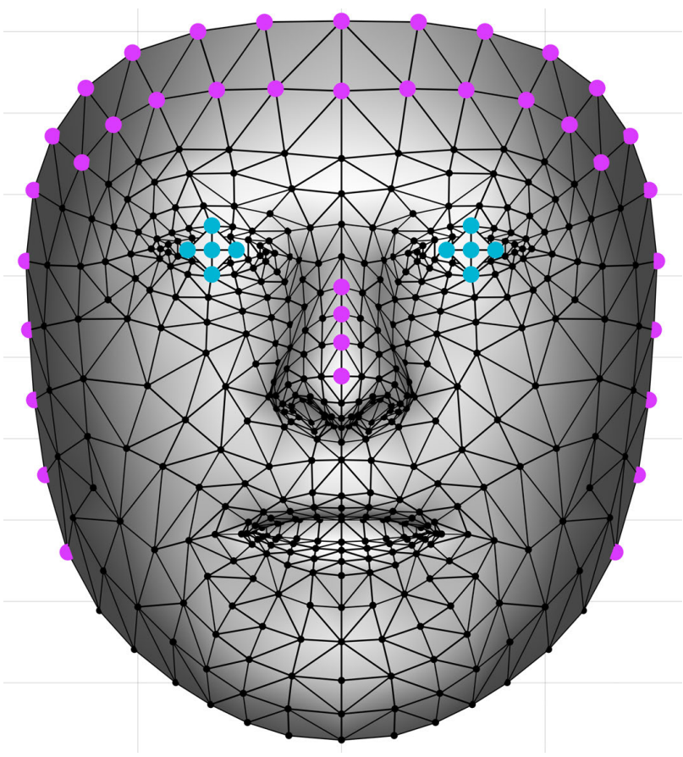
Brains, hearts, and eyes synchronize when attending to videos, but are hard to measure on a large scale



Madsen, *PNAS Nexus*, 2022 & Madsen, *arXiv*, 2025

- **Attentively watching videos synchronizes EEG, HR, gaze, and pupil**, and this Inter-Subject Correlation (**ISC**) strongly **correlates with their performance** when tested on the contents.
- But measuring these synchronized signals typically requires **complex sensors** that **are inaccessible or impractical for large-scale experiments.**

Modern AI methods allow real-time, on-device and privacy-preserving face tracking



- Google MediaPipe enables **real-time** facial landmark and movement extraction from **standard webcams.**
- The model can **run on the user's device**, in-browser, on our experiment platform **Elicit**.
- **Privacy is preserved** by only trasmitting face landmarks and blendshapes, no webcam video.

MediaPipe Elicit https://elicit-experiment.com/

The Data:

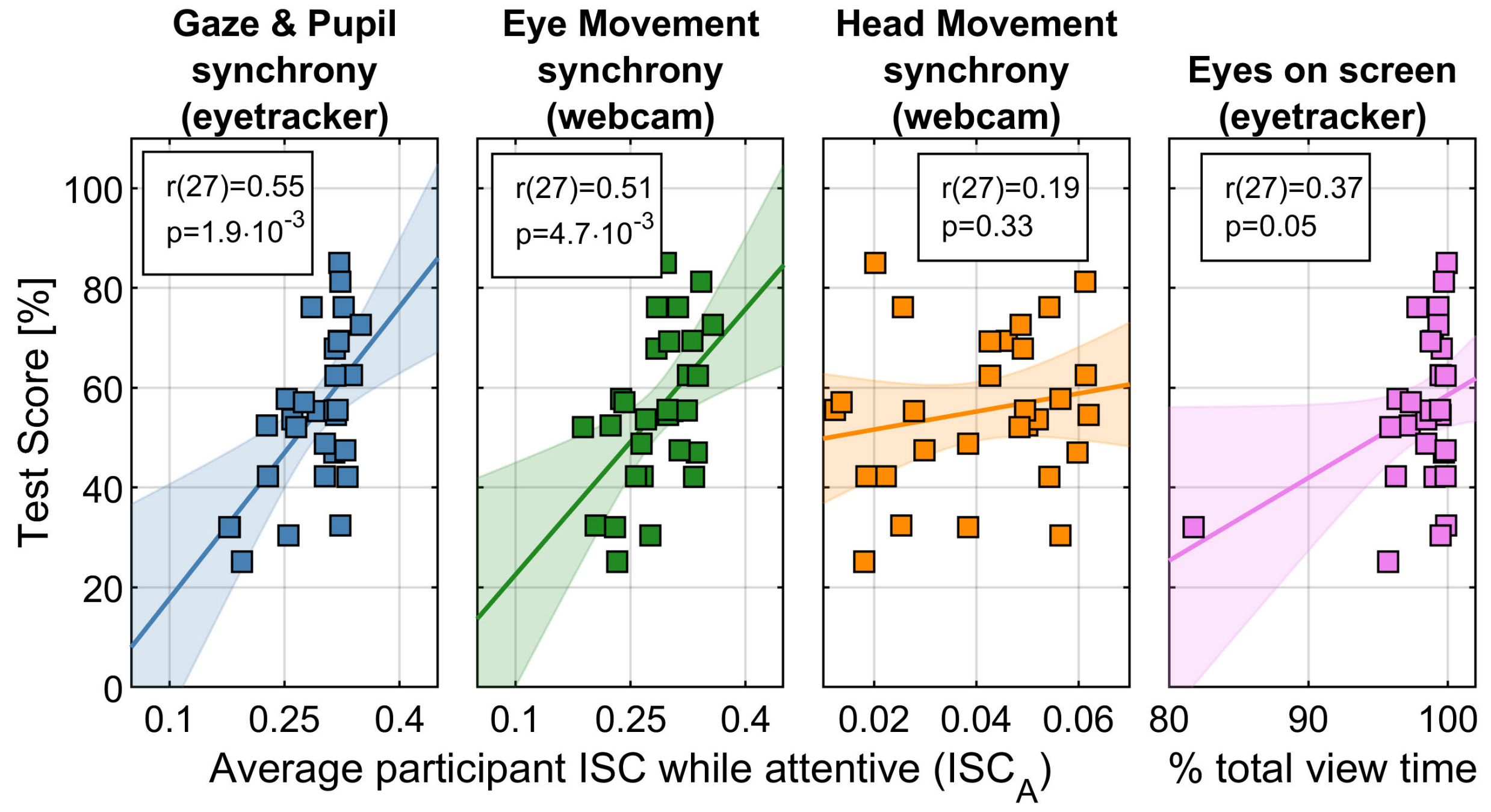
Dataset	Subjects	Stimuli		Webcam Data
	N	*N*	*Set*	*Duration (hours)*
Experiment 1	26 (10M 16F)	5	A	9.18
Experiment 2	29 (10M 19F)			11.14
Experiment 3	28 (8M 20F)	6	B	15.42

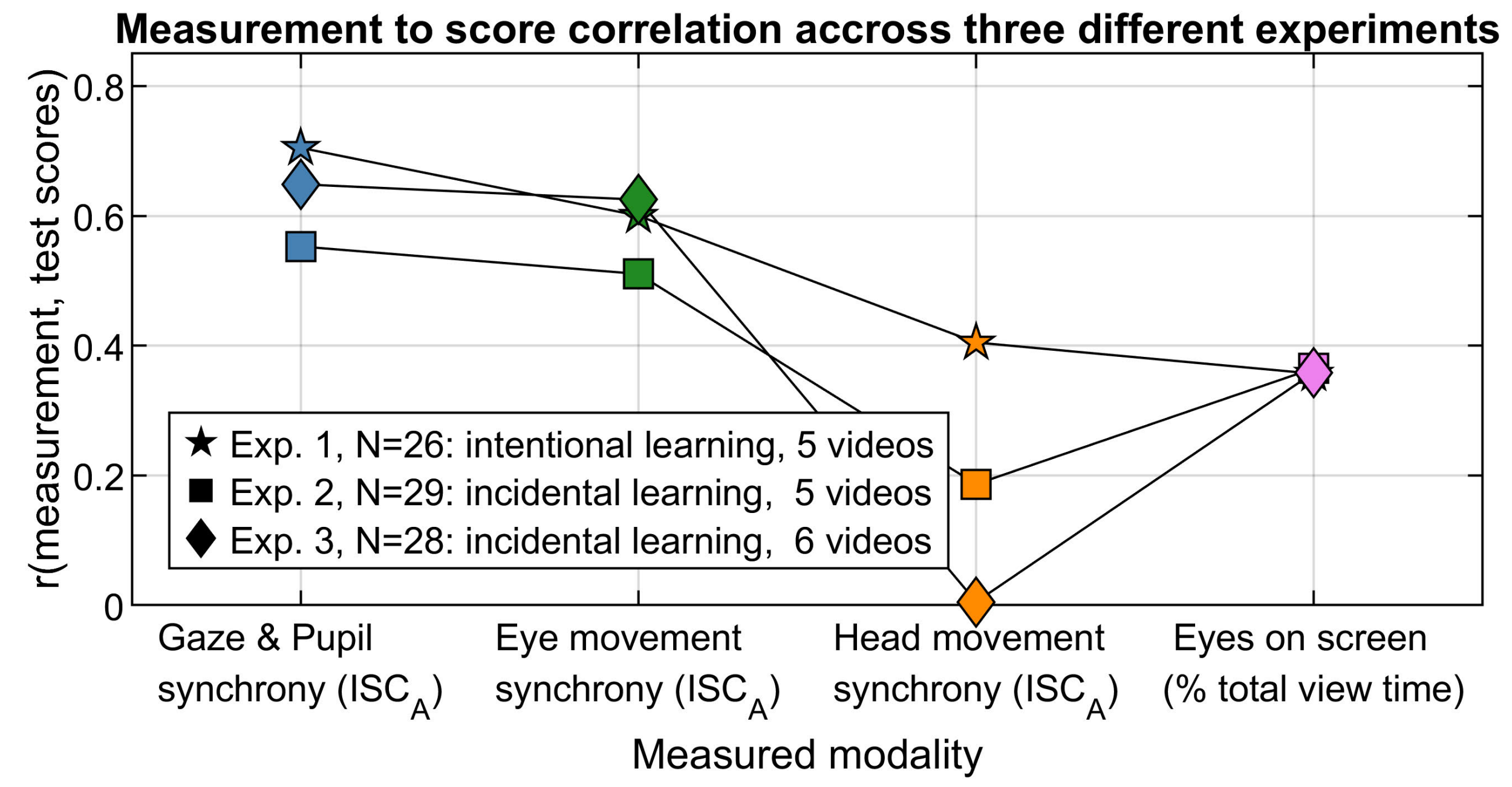
Highlights

- Engagement tracking from a standard webcam is feasible.
- Attention can be measured privately and remotely.
- The prediction generalizes to unseen participants and stimuli.

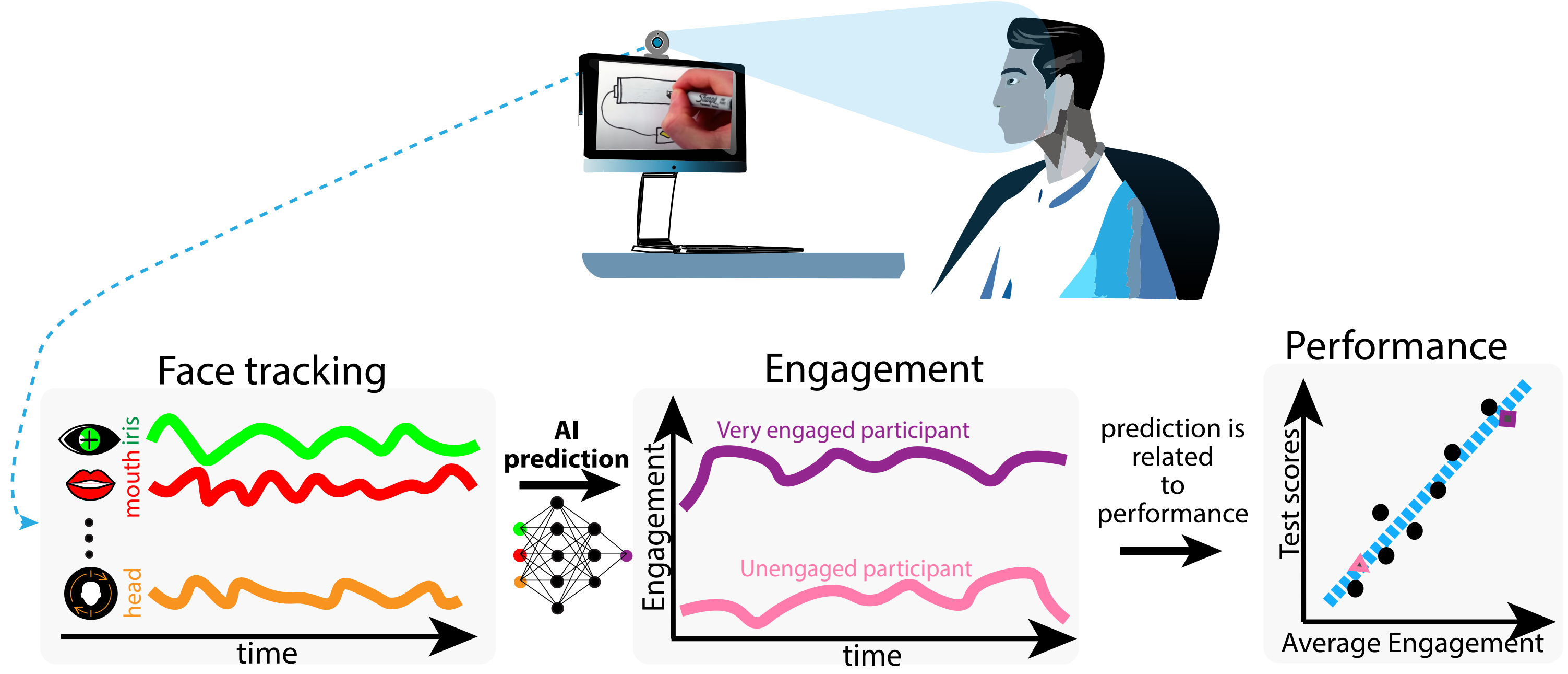
Engagement tracking from the face is comparable to eyetracking

- This is **consistent accross participants, experimental conditions, and tested videos**.
- But we **still need a reference group to compute ISC.**

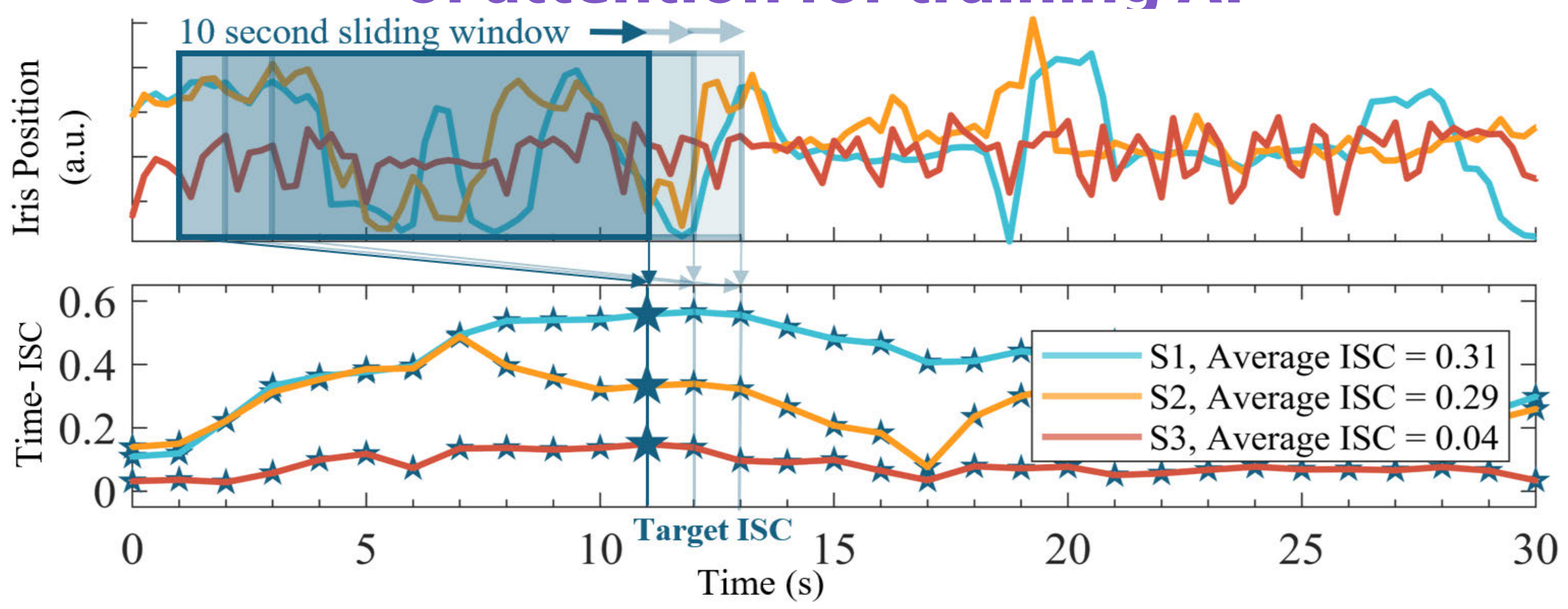


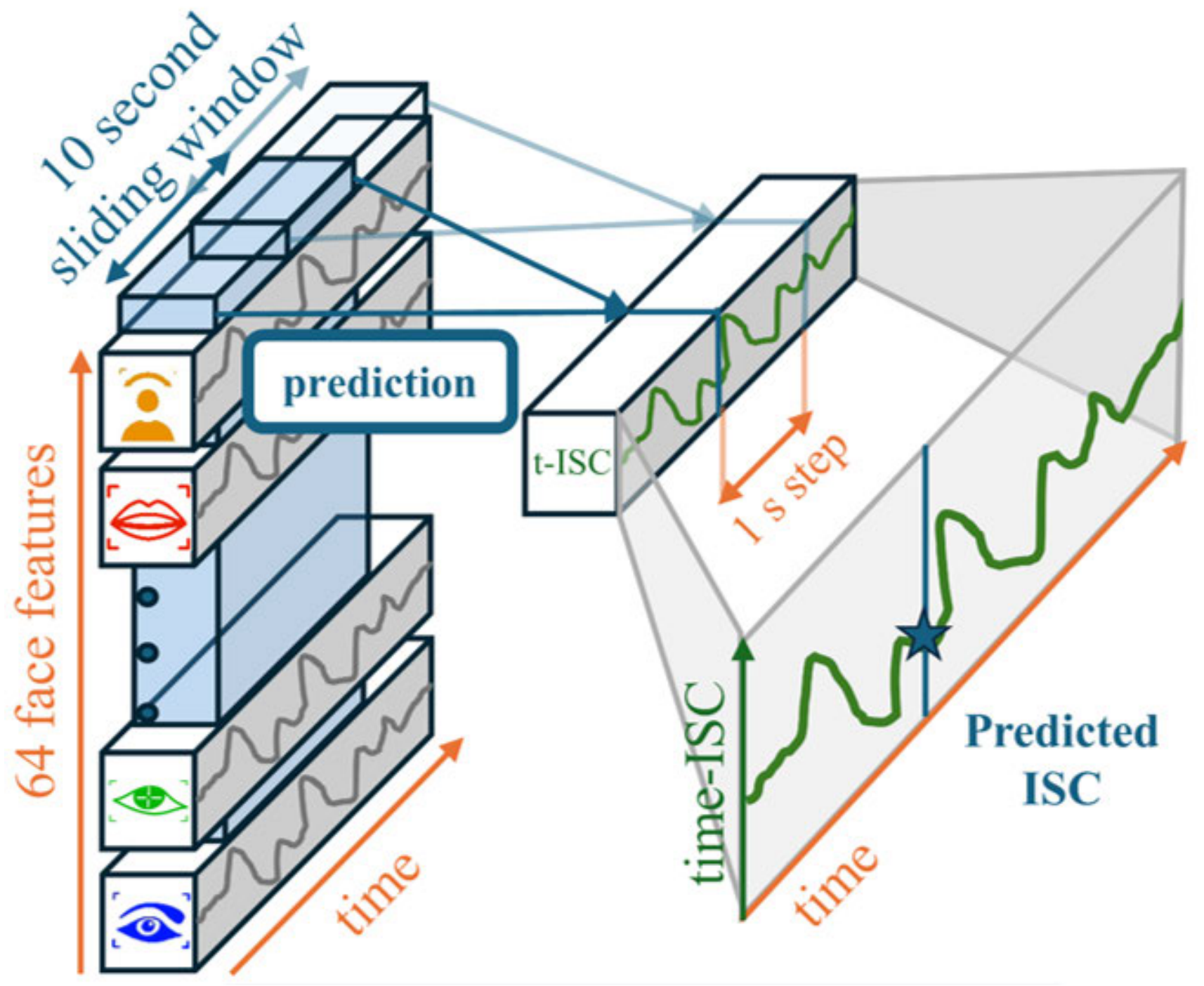


Using facial movements to predict engagement on a single participant



Using time-resolved ISC as an objective index of attention for training AI





Target

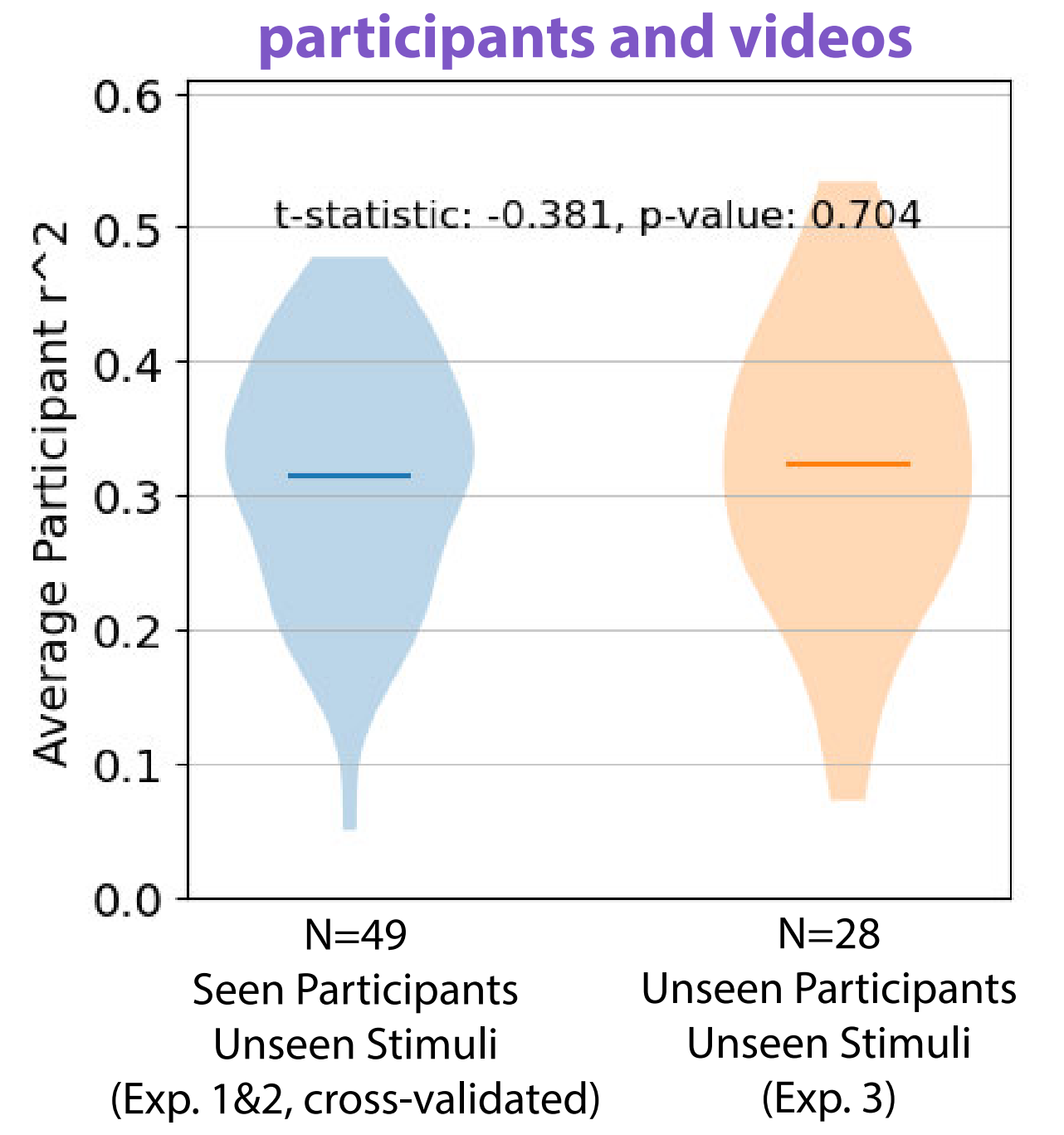
ISC of eye movements measured in 10 second windows.

Features

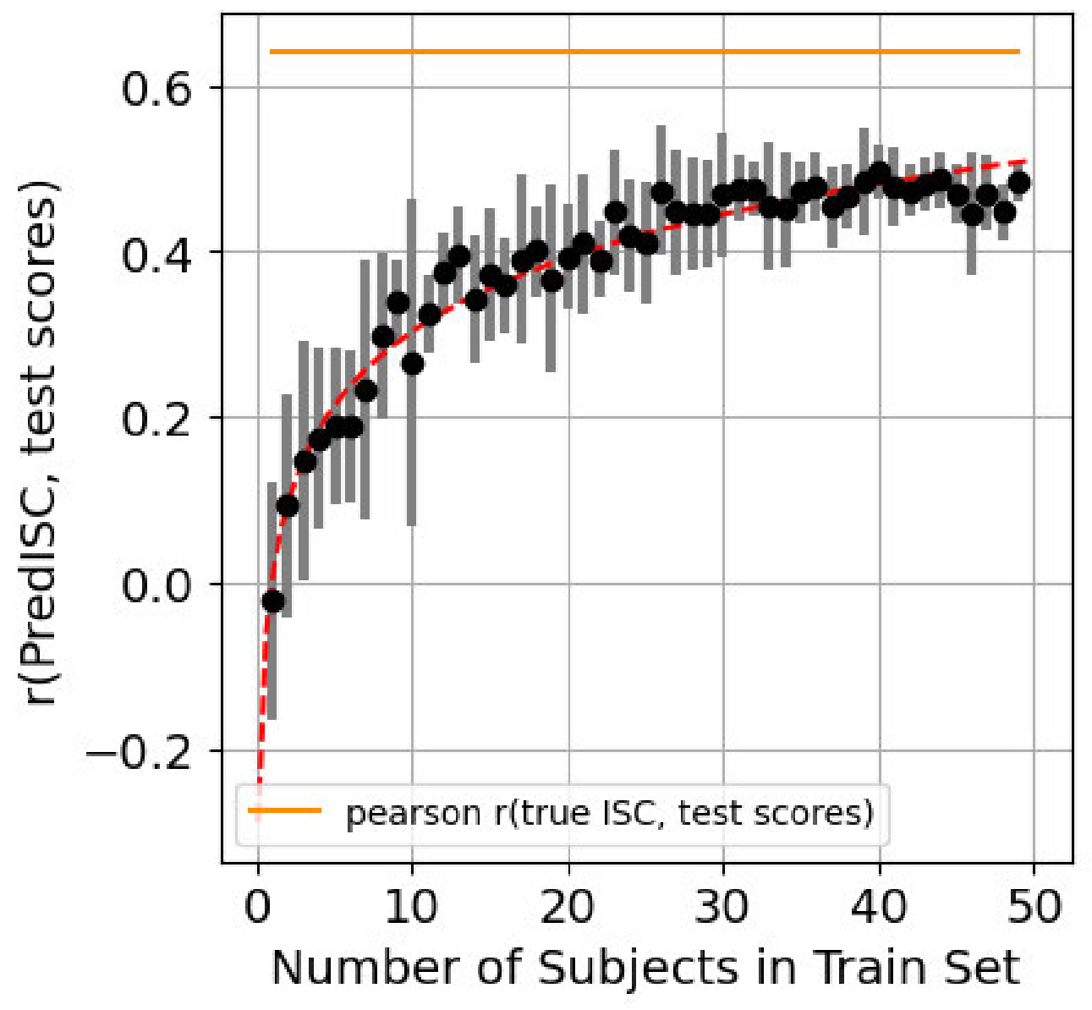
Mediapipe Blendshapes (facial movements) and Head Rotation features as predictors, over the preceeding 10 seconds.

AI engagement predictions translate to unseen participants and videos, and correlate with scores

Results generalize equally well to unseen participants and videos



How many participants are needed to capture performance-relevant engagement?



Contact

asilvanortubay@ccny.cuny.edu

aimarsilvan asortubay

Check it out!

Article Preprint

Live Face Tracking demo

Time Resolved ISC

Code github.com/asortubay/timeISC

References

Madsen, arXiv, 2025
Silvan, arXiv, 2024
Madsen, PNAS Nexus, 2022
Gupta et al., arXiv, 2016

Aknowledgment

Work supported by the National Science Foundation (DRL-2201835)