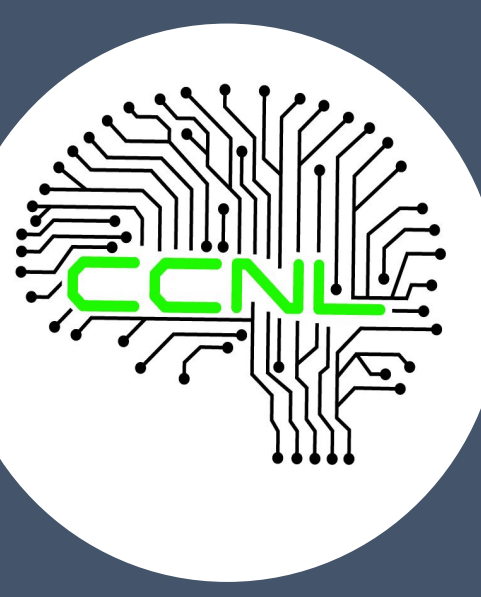




Comparing a large language model-derived metric of contextual information gain with human judgements in narratives



Palsule, R.A.¹, Goeschel, A.², Khoudary, A.¹, Yoo, J.³, Chen, J.⁴, Bornstein, A.M.¹

¹ Department of Cognitive Sciences, University of California, Irvine; ² Department of Computer Science, University of California, Irvine; ³ Princeton Neuroscience Institute, Princeton University;

⁴ Department of Psychological & Brain Sciences, Johns Hopkins University

Background

- Causal structure helps shape memory encoding and retrieval (Lee & Chen, 2022; Antony et al., 2024)
- Narrative stimuli (movies, books, etc.) are powerful tools for probing effects of causal structure on cognition and behavior (Chen & Bornstein, 2024)
- Inferences about causal structure play a key role in assigning credit to causally relevant events in long-term episodic memory (DuBrow et al., 2017; Yoo, 2025)
- We leverage recent advances in large language modeling to develop a metric that quantifies the *contextual information gain* of individual events in text-based narrative stimuli (Goeschel*, Palsule*, et al., 2026)
- Under this reformulation, we evaluate what aspects of a narrative *Sequentiality* captures, and hypothesize it will deviate from its original schema-based interpretation to one more sensitive to the structure of presented information

Research Question

How does our metric of contextual information gain relate to human judgements of narrative structure while viewing naturalistic narrative stimuli?

Open Sourced Replication

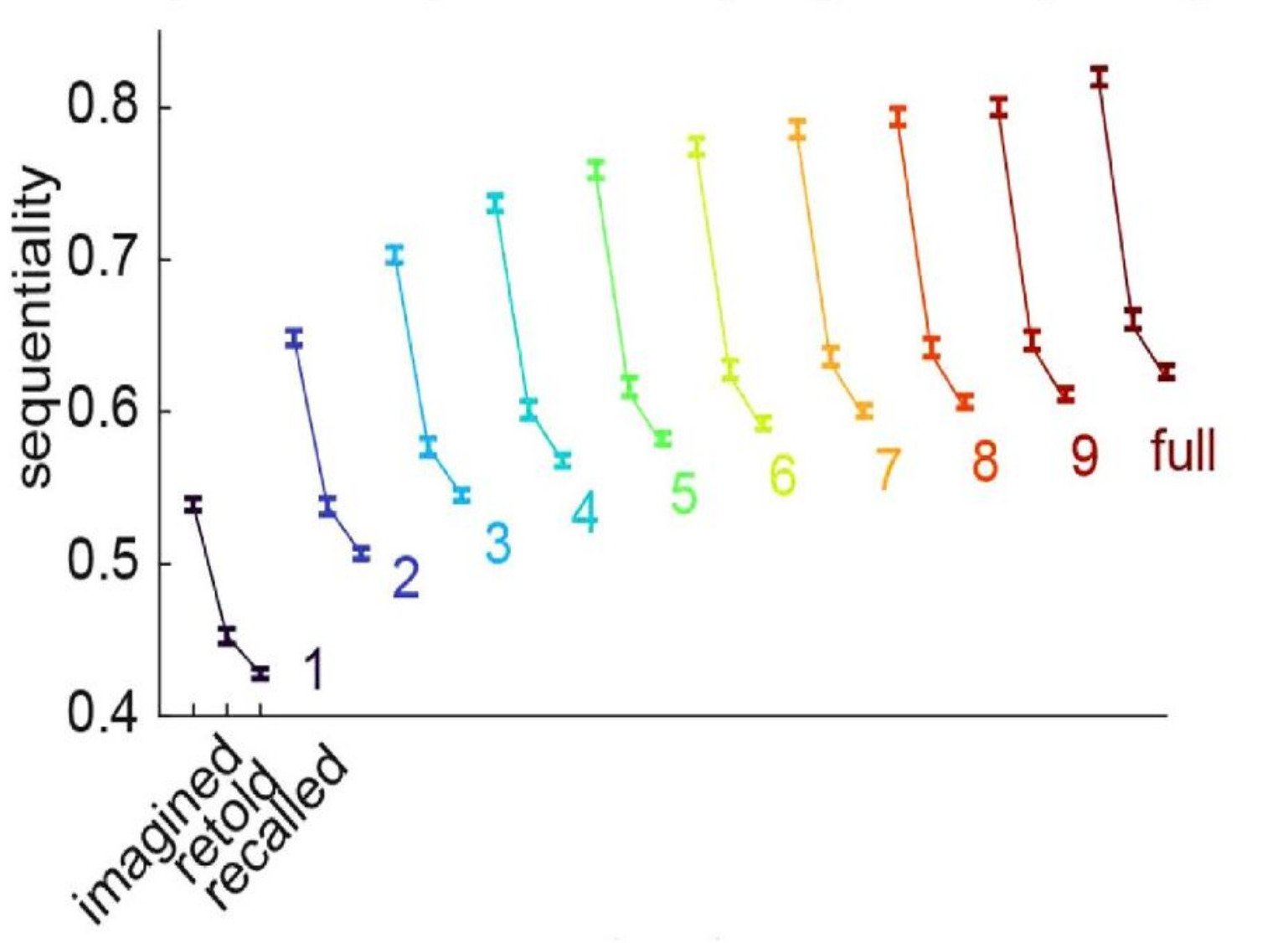
Reinterpretation for *Sequentiality* in Sap et al. (2022):

$$c(s_i, h) = -\frac{1}{|s_i|} (\log p(s_i|\tau) - \log p(s_i|\tau, h)) \equiv -\frac{1}{|s_i|} \log \left(\frac{p(h|\tau)}{p(h|\tau, s_i)} \right)$$

Sap et al. (2022)

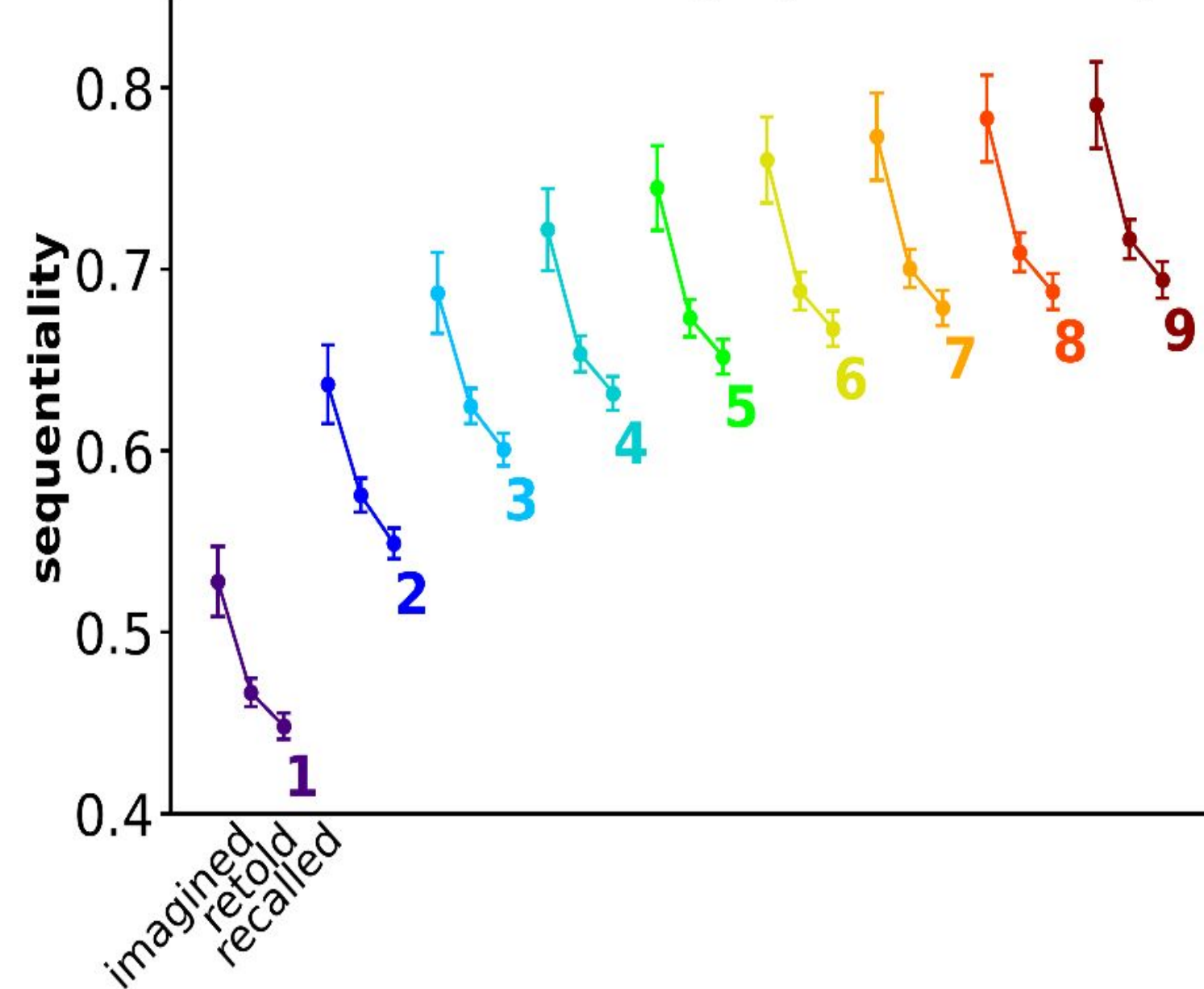
Original (Sap et al., 2022):

sequentiality with varying history length



Replication:

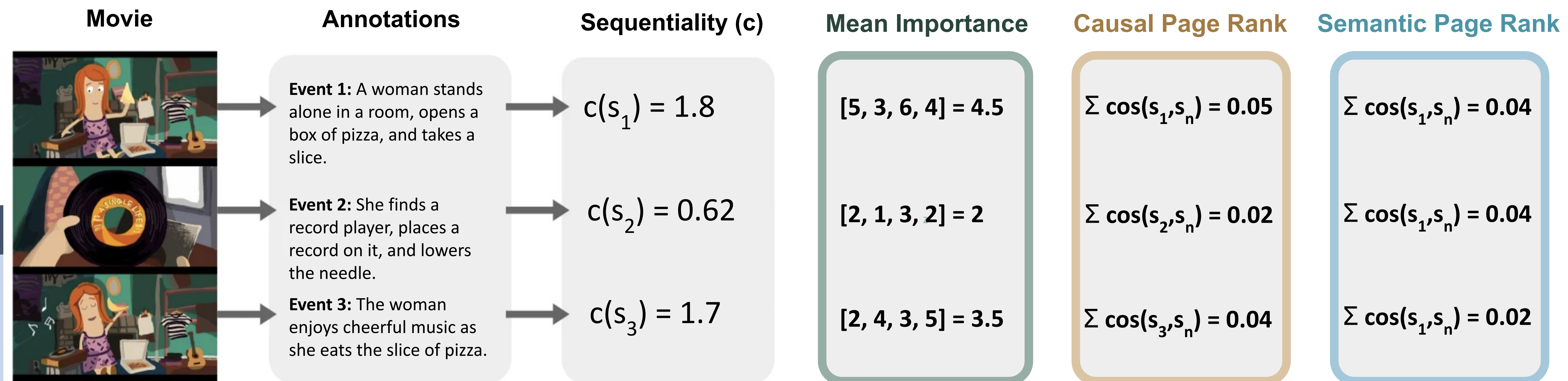
Ensemble Average (n=5 models)



Methods: Using the FilmFest Dataset (Lee & Chen, 2022)

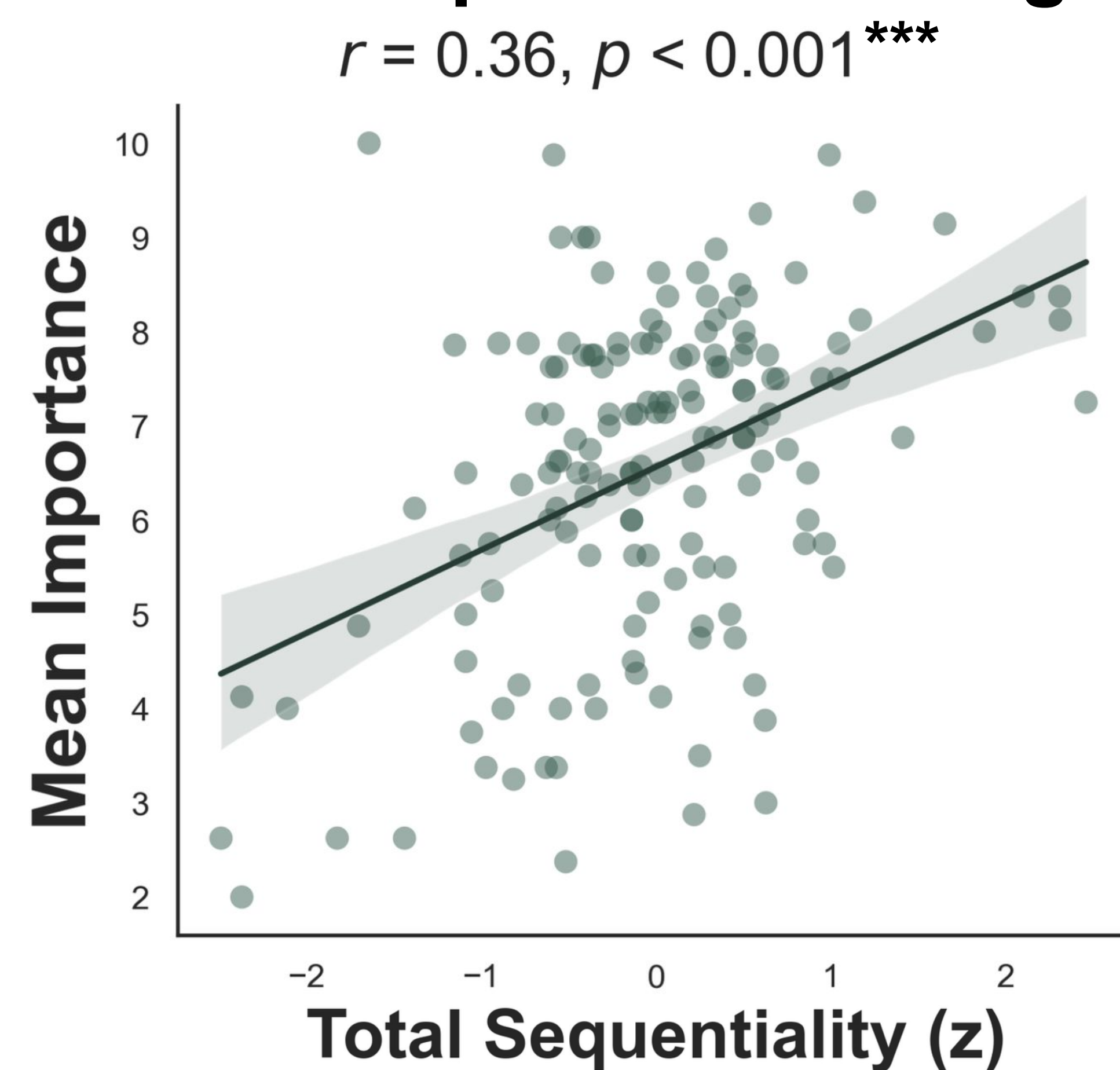
FilmFest Stimuli

- 10 short movies watched in succession during fMRI scanning (n=21)
- Human raters segmented each movie into event-level annotations (Lee & Chen, 2022)
- Annotations passed into model to generate Sequentiality values for each event in each story (n=216 annotations across all stories)

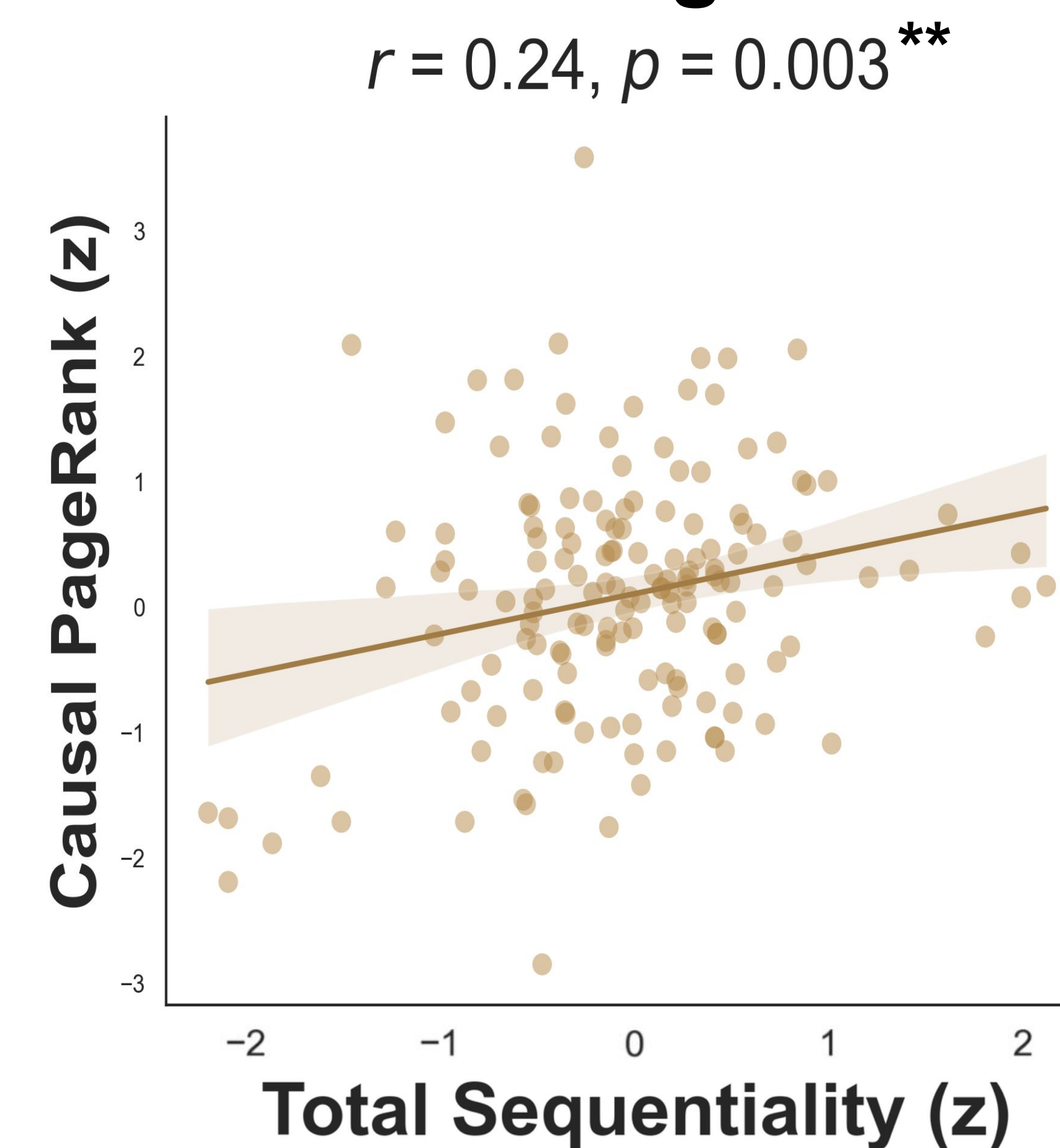


Results: Sequentiality tracks Human Judgments of Importance and Causality

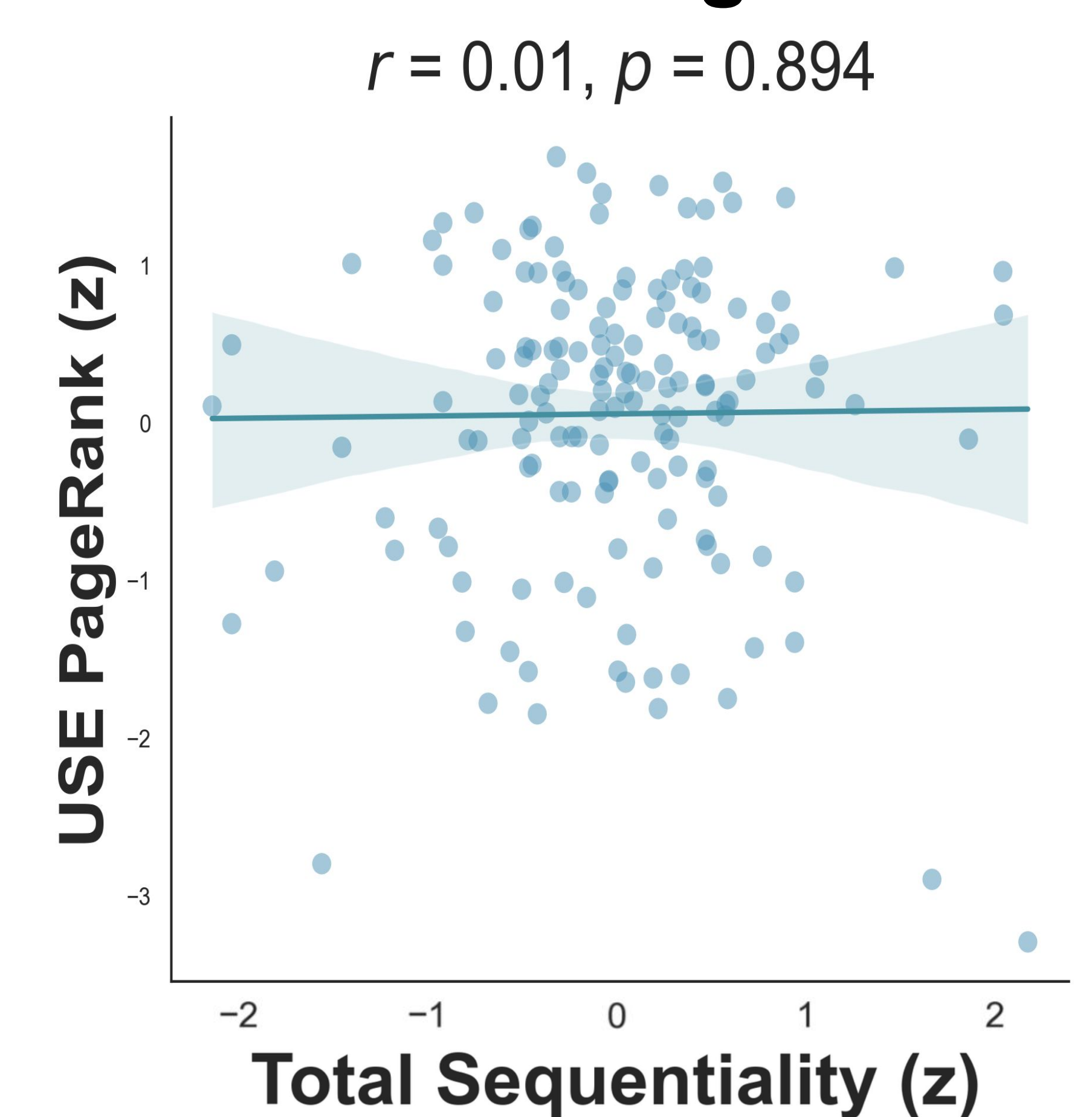
Mean Importance Ratings



Causal Page Rank



Semantic Page Rank



Preprint and GitHub



Contact me at rapalsul@uci.edu if you have any questions or insights!

Conclusion

Sequentiality, a metric believed to represent *contextual information gain*, is significantly correlated to the importance and causal structure of narratives while being unrelated to its semantic structure.

Future Directions

- Apply to long timescale narratives where credit assignment is complex and nonlinear
- Investigate how Sequentiality relates to neural activity during encoding, consolidation, and retrieval
- Use Sequentiality to guide generation of narrative stimuli that can test effects of causal structure on memory and behavior

References

Antony, J., Lozano, A., Dhoat, P., Chen, J., & Bannion, K. (2024). Causal and chronological relationships predict memory organization for nonlinear narratives. *J. Cogn. Neurosci.*, 36(11), 2368–2385.

Chen, J., & Bornstein, A. M. (2024). The causal structure and computational value of narratives. *Trends in Cognitive Sciences*, 28(8), 769–781.

DuBrow, S., Rouhani, N., Niv, Y., & Norman, K. A. (2017). Does mental context drift or shift? *Curr Opin Behav Sci.*, 141–146.

Goeschel, A., Palsule, R., Chen, J., Bornstein, A.M., Khoudary, A. (2026) Quantifying information gain in narrative stimuli using open-source large language models; preprint.

Lee, H., & Chen, J. (2022). Predicting memory from the network structure of naturalistic events. *Nature Communications*, 13, 4235.

Panela, R. A., Barnett, A. J., Barense, M. D., & Herrmann, B. (2025). Event segmentation applications in large language model enabled automated recall assessments. *Commun. Psychol.*, 3(1), 184.

Sap, M., Jafarpour, A., Choi, Y., Smith, N. A., & Horvitz, E. (2022). Quantifying the narrative flow of imagined versus autobiographical stories. *Proceedings of the National Academy of Sciences of the United States of America*, 119(45).

Yoo, J. (2025). Contingency as a fundamental principle for learning cognitive graphs.

Acknowledgements

Thank you to the Chen Lab at JHU and the UCI Computational Cognitive Neuroscience Lab for their assistance. Additionally, special thanks to Dr. Maarten Sap and Dr. Anna Jafarpour for introducing this metric and assisting with the model reformulation.