Quantifying information gain in narrative stimuli using open-source large language models

Aidan Goeschel[1,2*] **Rohin A. Palsule**[2*] **Janice Chen**[3] **Aaron M. Bornstein**[2+] **Ari Khoudary**[2+]

[1] Department of Computer Science, University of California, Irvine
[2] Department of Cognitive Sciences, University of California, Irvine
[3] Department of Psychological & Brain Sciences, Johns Hopkins University
[*,+] These authors contributed equally.

agoesche@uci.edu, rapalsul@uci.edu, aaron.bornstein@uci.edu, ari.khoudary@uci.edu

Abstract

We present an open-source, ensemble-based implementation of the *Sequentiality* metric (Sap et al., 2022) and demonstrate the feasibility of using it to approximate an *eligibility trace* that guides human credit assignment in causally complex environments. To do this, we demonstrate that the original *Sequentiality* metric—designed to quantify narrative "flow" in transcriptions of episodic recall—can equivalently be shown to quantify contextual information gain. We test this interpretation by applying the metric to stimuli and human behavior from the FilmFest dataset (Lee & Chen, 2022), and show that *Sequentiality* significantly correlates with the importance and causal centrality of events, but *not* their semantic centrality. This work expands the range of questions *Sequentiality* can be used to investigate, and takes an important first step toward understanding the neural and computational processes facilitating credit assignment in humans in naturalistic, multi-causal environments.

Introduction

All learning systems—humans, animals, and neural networks—are tasked with solving the *credit assignment* problem: determining which previous actions or events caused a current outcome. Accurately attributing credit to past events is a challenging computational problem whose complexity scales nonlinearly with the space of possible actions and the time elapsed between action and outcome. Recently, Chen and Bornstein (2024) proposed that *narrative schema* stored in long-term memory might be an essential component of humans' ability to efficiently solve complex credit assignment problems by providing complex prior knowledge about the plausibility of and causal relationships among events in a context-dependent, multi-timescale manner.

Chen and Bornstein (2024)'s theory was developed to account for mounting evidence that human memory systems are uniquely sensitive to causal information in narrative stimuli (Antony et al., 2024; Cohn-Sheehy et al., 2021; Lee & Chen, 2022; Lee et al., 2020; Zadbood et al., 2022). Notably, neural regions associated with value (e.g., vmPFC) activate during narrative viewing (Baldassano et al., 2018; Chang et al., 2021; Zadbood et al., 2022), suggesting that encoding of narrative stimuli involves dynamic, real-time valuation of incoming information. Toward this end, we present an open-source, ensemble-based implementation of the previously-developed *Sequentiality* metric (Sap et al., 2022), and show that it correlates with human judgments of event importance and causal centrality.

Sequentiality as a metric of contextual information gain

Sequentiality quantifies the time-evolving log-likelihood ratio of consecutive sentences in a piece of written text as

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

where s_i is sentence i, τ is the topic that acts as the metric's prior belief for the story, and h is the history length, dictating the number of previous sentences $s_{i-h:i-1}$ that are used to establish a local context; h is a free parameter ($h \in \mathbb{Z}^+$). Accordingly, the likelihood of a particular sentence s_i is defined by the log odds ratio of two generative models: one based on a user-defined topic τ only, and one based jointly on the topic and preceding h sentences. Although *Sequentiality* was originally developed to formalize a notion of narrative "flow", it can be equivalently re-expressed such that $c(s_i, h)$ corresponds to a metric interpretable as contextual information gain:

$$c(s_i, h) = -\frac{1}{|s_i|} log\left(\frac{p(h|\tau)}{p(h|\tau, s_i)}\right) \qquad (2)$$

which represents how much information sentence s_i tells us about the context established in history h that is not already contained in topic τ. This quantity is a promising metric for capturing the dynamic value of information as conveyed in complex, narratively-structured stimuli, and thus could serve as a powerful tool for future research.

Extended abstract presented at the *9th Conference on Cognitive Computational Neuroscience*, New York, NY, USA, 2026.

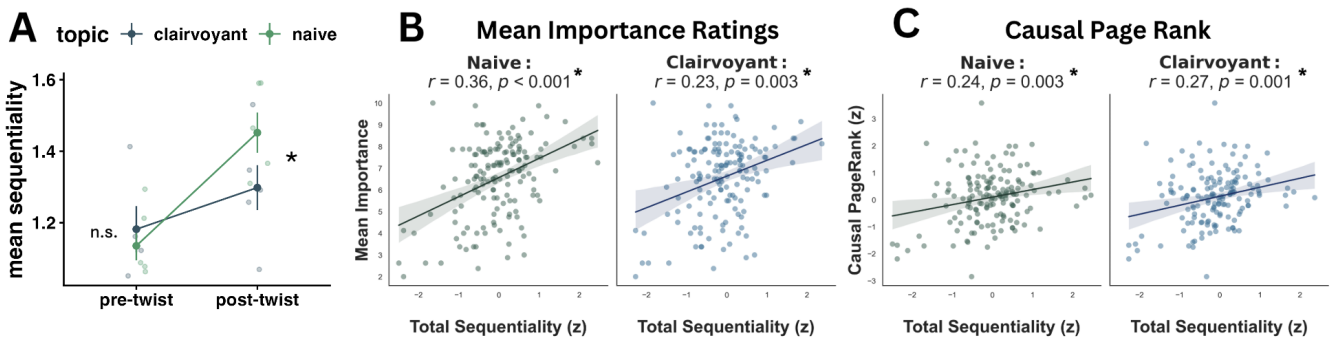

Figure 1: (A) *Sequentiality* is higher for events that happen after an unexpected "twist" in a narrative. (B,C) *Sequentiality* is significantly correlated with human ratings of event importance and causal centrality in the FilmFest dataset. "Clairvoyant" = topic with full information about story outcome; "naive" = topic set as title of the movie.

Methods and Results

We develop an open-source version of of the *Sequentiality* metric and test it using the FilmFest dataset (Lee and Chen, 2022). Our method uses an ensemble of LLMs (Sakana TinySwallow 1.5B, GPT2-XL, OLMo-2-13B, Llama-3.1-8B, and Qwen2.5-7B) to obtain likelihoods, thus minimizing potential biases deriving from training data and architectural differences. Further, any text generation model available via the HuggingFace API can be substituted or added to the ensemble. To test the feasibility of using *Sequentiality* to study credit assignment, we apply it to annotations of narrative stimuli, as well as human ratings about events in those narratives; we used a maximum history length of the entire story for each analysis. Data and code used to generate these results, as well as a link to our ensemble *Sequentiality* metric, can be found on GitHub.

Sensitivity to narrative "twists"

Previous work has shown that narratives with a "twist" or surprise ending induce marked changes in both the free recall and neural activity of humans engaging with the movie as they re-interpret the previously-observed events in light of the new information (Sava-Segal et al., 2026; Zadbood et al., 2022), a process central to credit assignment in complex environments. We computed *Sequentiality* values for annotations of *The Boyfriend*—a FilmFest movie with a "twist" halfway through—using topics that correspond to initially watching the surprise ending ("naive") and re-watching the same movie with knowledge of the twist ("clairvoyant"). A regression on *Sequentiality* values generated by each model in the ensemble returned a significant interaction: post-twist events had higher average *Sequentiality* values for naive relative to clairvoyant topics (β = 0.2, t_{12} = 4.858, $p < .001$; Figure 1A), consistent with an information gain interpretation of the metric.

Associations with human behavior

Next, we investigate whether *Sequentiality* is associated with human ratings of event importance and causal centrality, two constructs implicated in the psychology of credit assignment (Lee & Chen, 2022). As a control, we investigate its association with semantic centrality as quantified by the narrative networks developed by Lee and Chen (2022). In direct support of our information gain interpretation, Figures 1B and C show that event importance is significantly correlated with both "naive" and "clairvoyant" *Sequentiality* values (all $r > 0.22$, $p < .005$). However, there was no correlation between "naive" or "clairvoyant" *Sequentiality* values and the semantic centrality of events (both $r < .03$, $p > .5$). This dissociation accords with empirical findings of the independence of causal and semantic centrality in human credit assignment (Antony et al., 2024; Chen & Bornstein, 2024; Lee & Chen, 2022).

Discussion

We developed an open-source, ensemble-based implementation of the *Sequentiality* metric and showed that it is significantly associated with factors that are empirically and theoretically linked to credit assignment in humans (Chen & Bornstein, 2024). These findings serve as proof-of-concept that the metric can be used to link time-evolving contents of narrative stimuli with neural activity associated with perception of latent causal structure. Our work also expands the range of questions that *Sequentiality* can be used to study by demonstrating its formal equivalence with contextual information gain and by making the metric openly-accessible for anyone to use. Future work will use *Sequentiality* to design experimental stimuli that permit directly testing how contextual information gain relates to behavioral and neural signatures of credit assignment.

Acknowledgements

We thank Drs. Maarten Sap and Anna Jafarpour for their guidance in troubleshooting our open-source implementation and validating it with the HIPPOCORPUS dataset, respectively. For helpful feedback on this project, we thank members of the Cognitive Computational Neuroscience and Reflexion Labs at UCI, the Chen and Honey labs at Johns Hopkins University, Lio Wong, and Judith Fan. This work was supported by NIMH T32MH119049 (to A.K.), NINDS R01NS119468 (to A.M.B.; PI E.R. Chrastil), and NIA R01AG088306 (to A.M.B.).

References

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

Baldassano, C., Hasson, U., & Norman, K. A. (2018). Representation of real-world event schemas during narrative perception. *J. Neurosci.*, *38*(45), 9689–9699.

Chang, L. J., Jolly, E., Cheong, J. H., Rapuano, K. M., Greenstein, N., Chen, P.-H. A., & Manning, J. R. (2021). Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. *Sci Adv*, *7*(17).

Chen, J., & Bornstein, A. M. (2024). The causal structure and computational value of narratives. *Trends in Cognitive Sciences*, *28*(8), 769–781. https://doi.org/10.1016/j.tics.2024.04.003.

Cohn-Sheehy, B. I., Delarazan, A. I., Crivelli-Decker, J. E., Reagh, Z. M., Mundada, N. S., Yonelinas, A. P., Zacks, J. M., & Ranganath, C. (2021). Narratives bridge the divide between distant events in episodic memory. *Memory & Cognition*, *50*(3), 478–494. https://doi.org/10.3758/s13421-021-01178-x.

Lee, H., Bellana, B., & Chen, J. (2020). What can narratives tell us about the neural bases of human memory? *Current Opinion in Behavioral Sciences*, *32*, 111–119.

Lee, H., & Chen, J. (2022). Predicting memory from the network structure of naturalistic events. *Nature Communications*, *13*, 4235. https://doi.org/10.1038/s41467-022-31965-2.

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), e2211715119. https://doi.org/10.1073/pnas.2211715119.

Sava-Segal, C., Grall, C., & Finn, E. S. (2026). Narrative "twist" shifts within-individual neural representations of dissociable story features. *Proc. Natl. Acad. Sci. U. S. A.*, *123*(11), e2512071123.

Zadbood, A., Nastase, S., Chen, J., Norman, K. A., & Hasson, U. (2022). Neural representations of naturalistic events are updated as our understanding of the past changes. *Elife*, *11*(e79045).