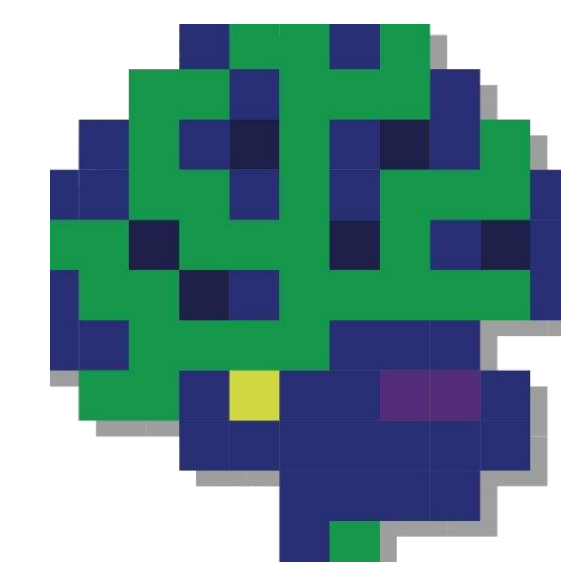


Strategic Control of Episodic Memory Through Post-Gating

Cody Dong, Qihong Lu, Kenneth Norman

PRINCETON UNIVERSITY



Main Points

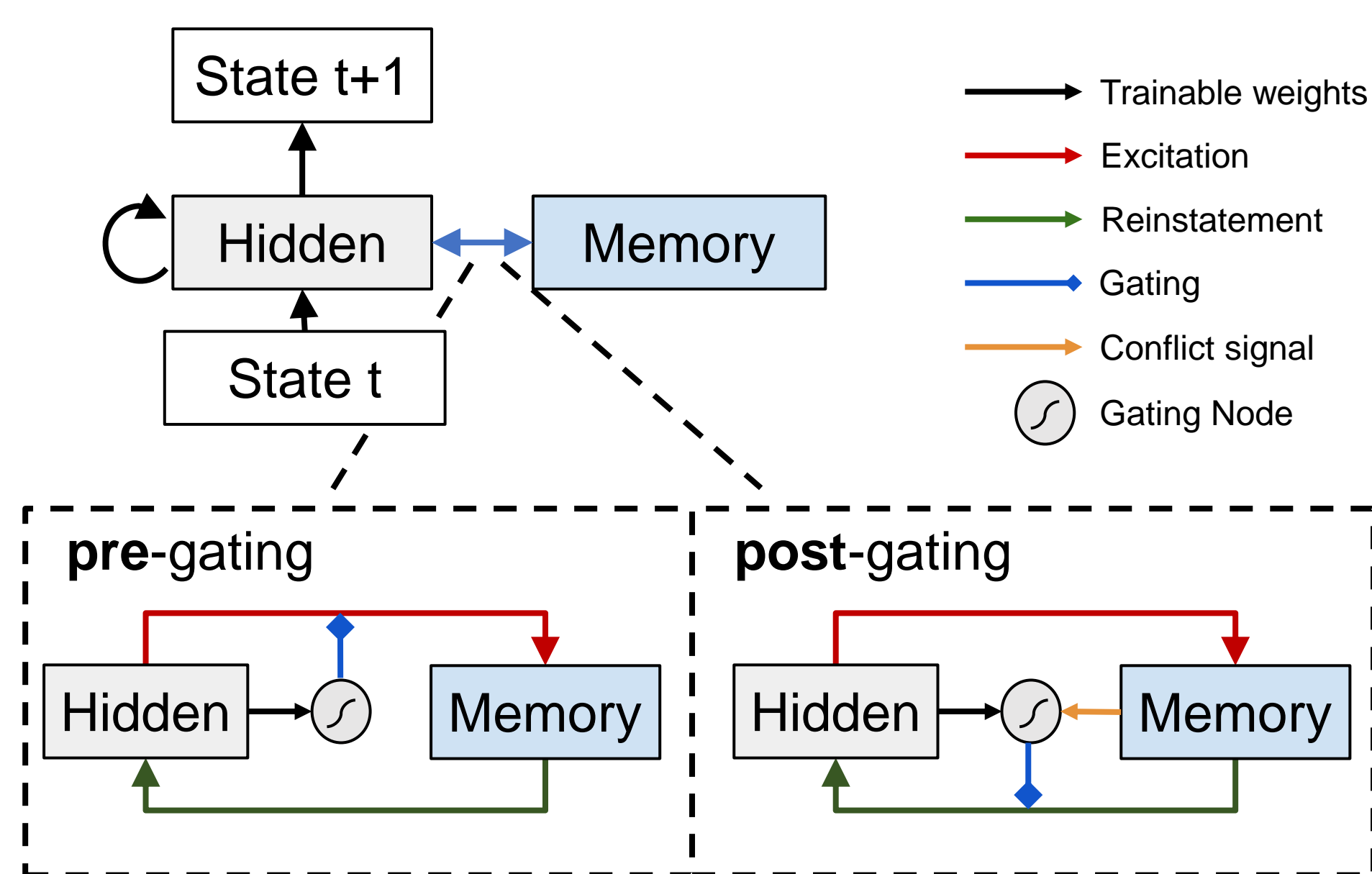
Can people learn to strategically control episodic memory?

- Are people able to modulate retrieval of episodic memory based on the diagnosticity of retrieval cues?

Results:

- People refrained from making predictions in situations when it is unclear which memory to retrieve (i.e., non-diagnostic retrieval cues)
- Comparing memory-augmented models that learn optimal retrieval policies: Post-gating models leveraging a memory conflict signal were able to account for this selective behavior while pre-gating models could not

Memory Augmented Model



- Model^{1,2} is a recurrent neural network connected to an episodic memory module
- Memory Retrieval:
 - stored memories are **activated** based on similarity to current hidden state
 - activated memories compete to be retrieved using a leaky competing accumulator process (LCA)
 - Averaged memory pattern (weighted by final activation from LCA) is added back into the hidden state
- Pre-gating models can **control** initial level of memory activation while post-gating models control the magnitude of the final retrieved memory pattern
- Post-gating models have access to a binary **conflict signal**³ that can potentially influence the degree of retrieval
- Model was trained using reinforcement learning (A2C) on behavioral task analogous to our experiment, with the ability to respond "don't know" to queries

Future Directions

Issue: "Don't Know" responses may not track retrieval gating

- Participants may retrieve candidate memories beforehand but decide to withhold their prediction

Need alternate ways of measuring when retrieval happens:

- Priming: if retrieval of a specific memory occurs, then its subsequent memory is enhanced⁵
- Neural measures of episodic recall⁶

Want tighter coupling between prediction and retrieval:

- Avoid cued recall because it artificially introduces a deliberative step between retrieval and prediction
- Switch to different tasks incorporating retrieval-based predictions that occur naturally (e.g., during story reading)

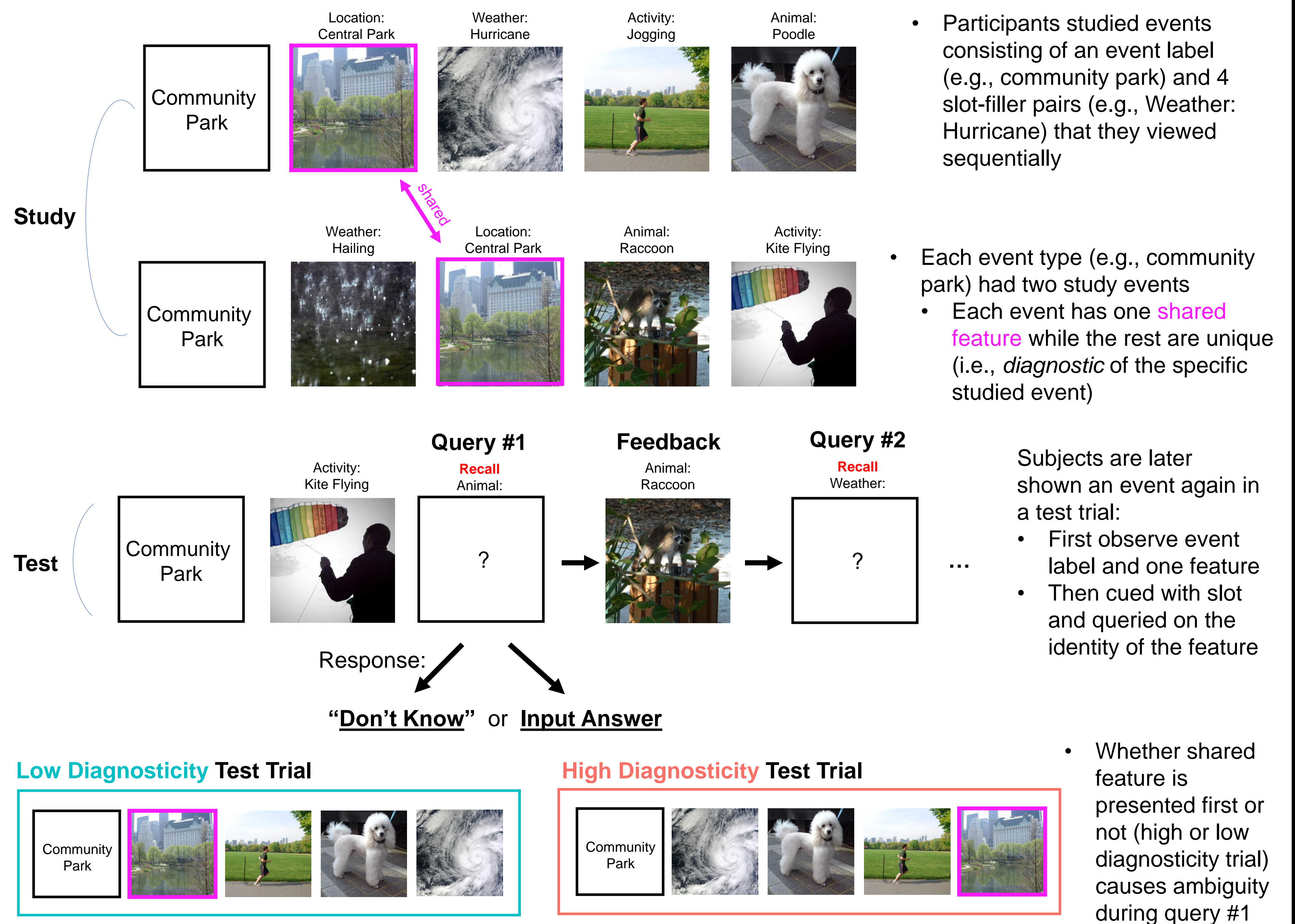
References

- Lu, Q., Hasson, U., Norman, K. A. eLife 2022
- Ritter, S., et al. arXiv 2018
- Botvinick, M., Cohen, J. D., Carter, C.S. Trends Cogn Sci. 2004
- Johnson, M. K., et al. Psychological Bulletin 1993
- Smith, T. A., Hasinski A. E., Sederberg P.B. J Exp Psychol Gen. 2013
- Chen, J., et al. Cerebral Cortex 2016

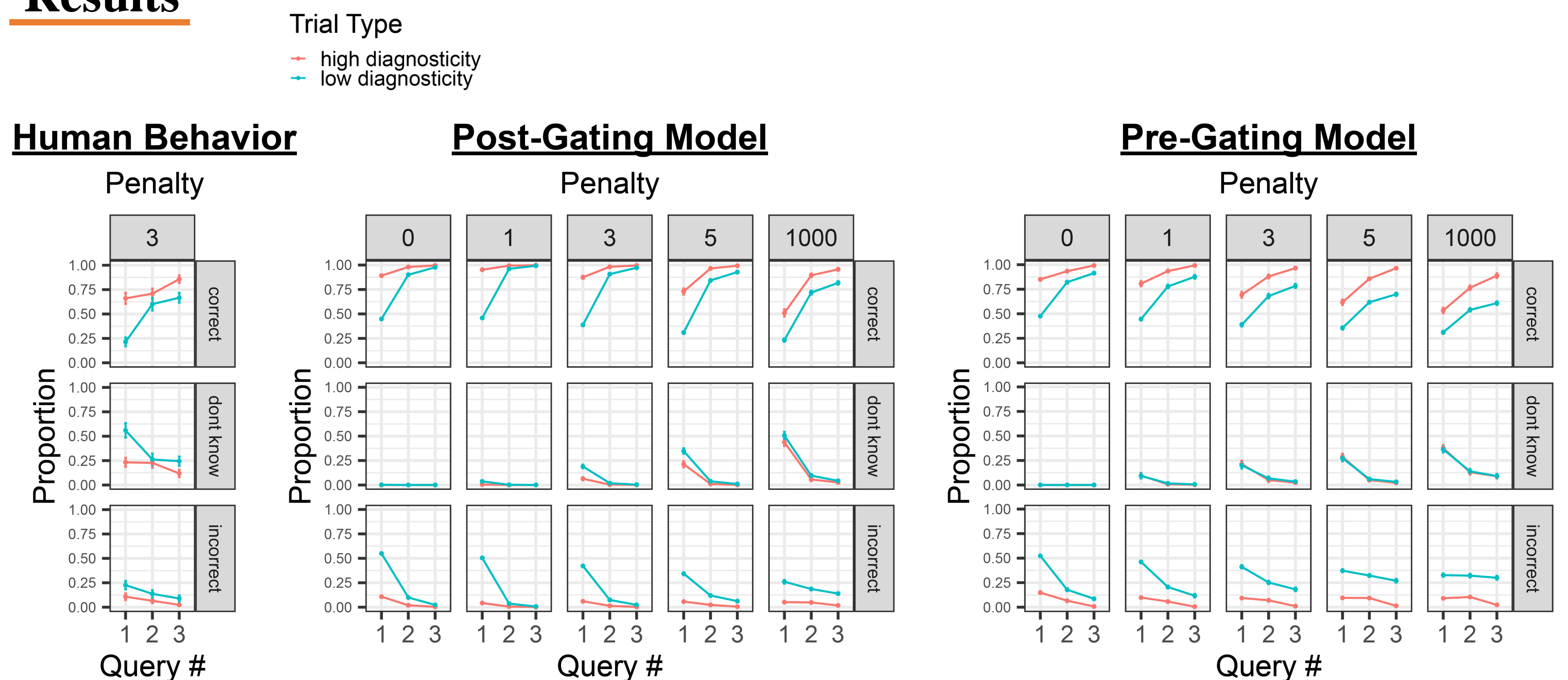
Acknowledgement:

This work was supported by a Multi-University Research Initiative (MURI) grant awarded to KAN (ONR/DoD N00014-17-1-2961).

Experimental Paradigm

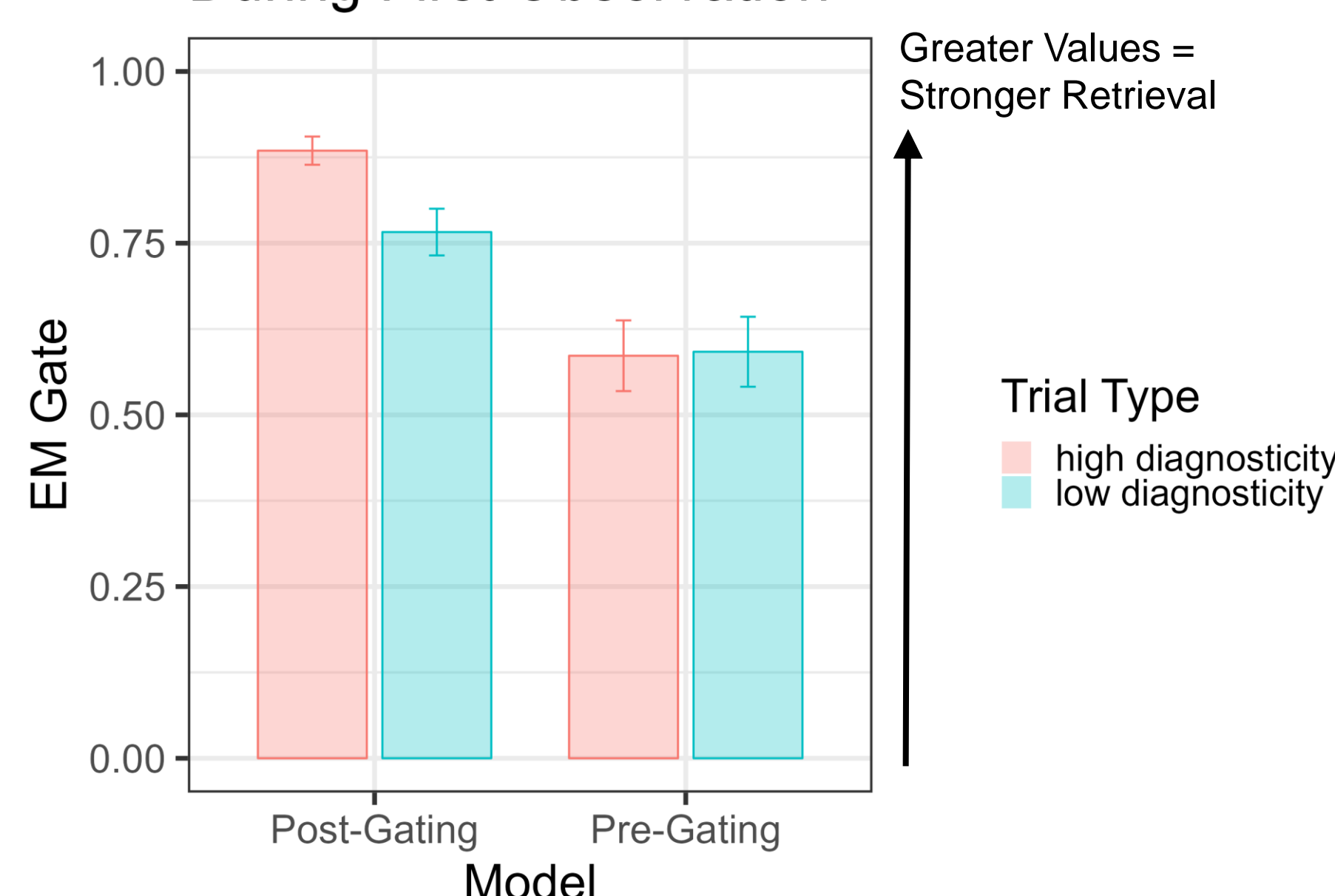


Results



- Participants were able to accurately discern whether the first feature was ambiguous (i.e., during a low diagnosticity trial) and respond don't know at Query #1
 - Decrease in "don't know" responses from Query #1 to Query #2 indicates that, once participants were shown a diagnostic feature, they were able to retrieve the correct memory to respond
- Only the post-gating model qualitatively replicates this human pattern of don't know responses at medium penalties
 - Increasing penalty causes both models to become more cautious, leading to greater proportion of don't know responses

Gating Magnitude During First Observation



- By leveraging a conflict signal between activated memories, the post-gating model can limit retrieval when it is unclear which memory corresponds to the features presented thus far
- In contrast, the pre-gating model is unable to detect when multiple similar events are equally plausible
- Post-gating may generally support selective retrieval based on the characteristics of the activated memories⁴ (e.g., conflict)