The Role of Context Gating in Predictive Sentence Processing

Yasemin Gokcen, David Noelle, and Rachel Ryskin
Department of Cognitive and Information Sciences
University of California, Merced

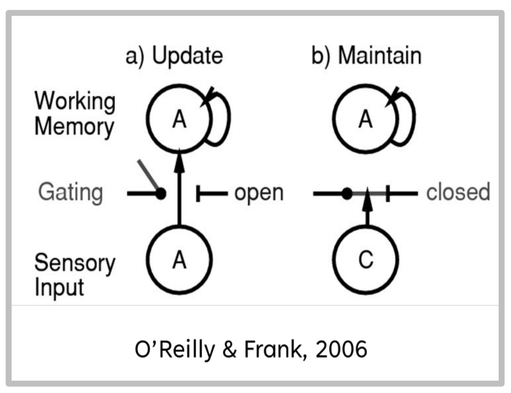
BACKGROUND

Prediction from context during language processing

- When listening to or reading a sentence, the brain is constantly predicting what word will come next.
- ***Surprisal*** measures how unexpected a word is given the context before it. (Levy, 2008)
  - Higher score number = Word is more unexpected
- The brain's memory of past words is imperfect, it needs a mechanism to decide what to hold onto.

Working memory & gating

- The brain's prefrontal cortex (PFC) uses a gating mechanism to manage working memory: selectively storing relevant information and filtering out what isn't needed.



- Memory is *updated* when gate *opens* to relevant information.
- Memory is *maintained* when gate *closes* to irrelevant information.
- The anatomy and physiology of the PFC are thought to support this mechanism (Noelle, 2012).

QUESTION

Are the gating mechanisms of prefrontal cortex used to maintain context for prediction during language processing?

GENERAL APPROACH AND METHODOLOGY

- Compared two language models that differ in whether they use a gating mechanism.
- Measured how well each model predicts real human brain responses to spoken words.

Language Models:

- Recurrent Neural Network (RNN): predicts words from context, but has no gating mechanism.
- **Long Short-Term Memory (LSTM): same architecture, but includes PFC-like gating.**
- Both models were trained on the same large text dataset and tested on the same spoken stories.

Neural activity:

- Participants listened to naturalistic stories while we recorded brain activity using EEG (electroencephalography).
  - 10 stories (~5 min each), designed to feature varied and complex language.
  - 32 participants (19 female, mean age = 21.65 years old)

Analysis

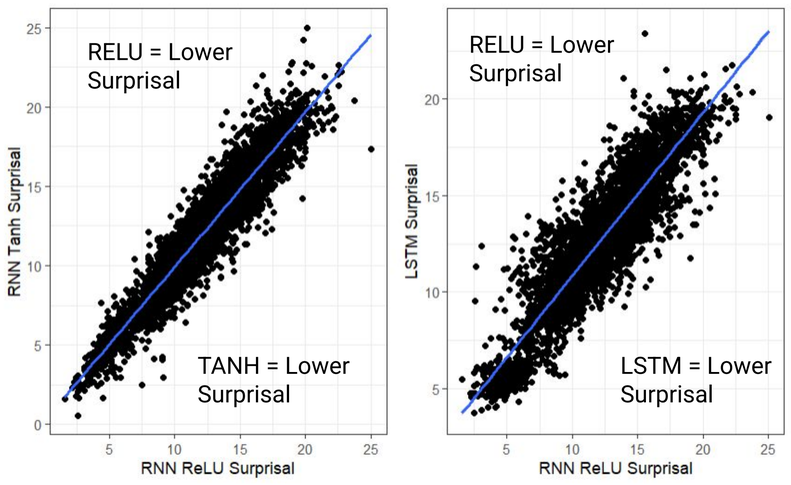
- Statistical models to predict brain response to each word
- Control for low-level factors
  - Word frequency, position
- Isolate the effect of prediction and gating specifically
- Whether how unexpected a word is, and whether gating helped predict it, affects how strongly the brain responds

```
MeanAmp ~ Frequency + PreviousFrequency +
 CosineSimilarity + PreviousCosineSimilarity +
GPT Surprisal + PreviousSurprisal + WordDuration
+ SentencePosition + StoryPosition +(1|Story)+ 1
 + CosineSimilarity + GPT Surprisal ||Subject)
```

A METRIC OF GATING UTILITY FOR PREDICTION

- Both models agree on most word predictions + correlate highly (r = .89 for LSTM vs. RNN)

- **Gating utility score: For each word, we measure how much better the gated model (LSTM) predicts it compared to the ungated model (RNN). A higher score means gating helped more for that word.**



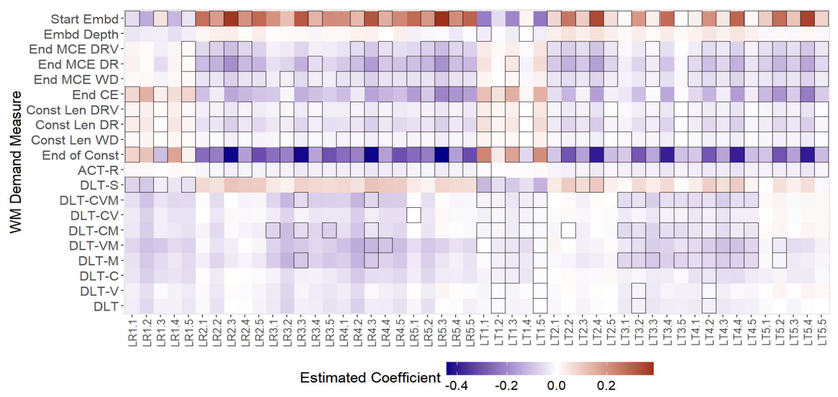
One dot = one word appearance

Words where the LSTM outperforms the RNN (shown below the diagonal) are words where gating may be doing important work.

GATING UTILITY IS CORRELATED WITH WM DEMANDS

We tested whether gating is most useful when working memory is under the greatest strain, using three established measures of memory demand in sentences:

- **DLT**: Dependency Locality Theory (Gibson, 2000)
  - Memory cost increases when words that depend on each other are far apart in a sentence.
- **ACT-R**: Adaptive Character of Thought—Rational (Anderson et al., 2004)
  - Measures the effort required to retrieve earlier words from memory.
- **LCP**: Left-Corner Parsing (Lewis & Vasishth, 2005)
  - Memory demand increases with complex nested sentence structures.



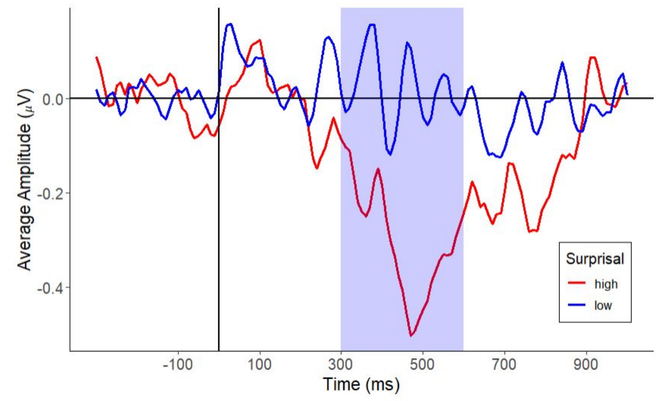
- Each cell shows the relationship between a memory demand measure and the gating score.
- Bold border = statistically reliable result.

Across all three frameworks, higher memory demand is associated with greater gating utility.

When sentences are harder to hold in memory, the gated model has the biggest advantage over the ungated one — suggesting that gating is specifically recruited when context memory matters most.

SURPRISAL IS RELATED TO N400 ERP EFFECT

- We measured participants' brain activity (EEG) in response to each word and used statistical models to test what predicted the strength of the brain response.
- Surprisal values from OpenAI's GPT-Davinci

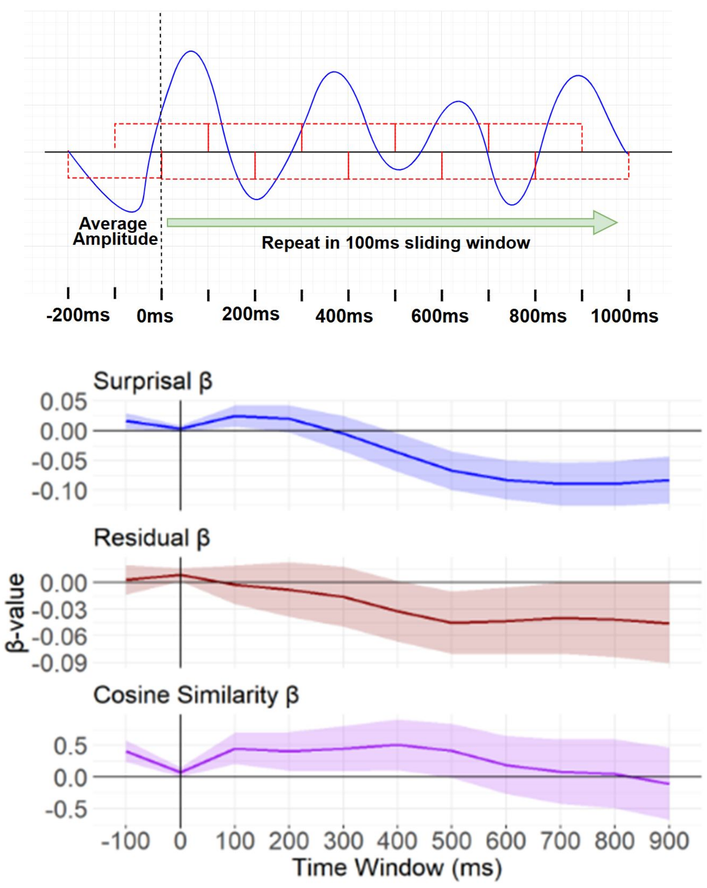


Word surprisal was the key predictor of the N400 brain response.

Words with higher surprisal produced a stronger N400 response, replicating a well-established finding in a naturalistic listening setting (Frank et al., 2015).

TIME COURSE OF EFFECT OF GATING ON NEURAL RESPONSE

- To find out when gating influences brain activity, we tracked the gating score's effect on the brain response over time after each word.



- Word surprisal produces a reliable brain response in the 300–500 ms window
- The gating score produces a separate, later effect starting around 400–600 ms, suggesting it reflects a distinct cognitive process.

DISCUSSION & FUTURE DIRECTIONS

- The brain's prefrontal gating system appears to play a key role in supporting language prediction during natural listening.

- Our brain recordings show this gating process has a distinct neural signature
  - Emerging slightly later than basic prediction difficulty.

Poster Number: 156

Scan for more.

ARCS FOUNDATION INC NORTHERN CALIFORNIA CHAPTER