Neural representations of structured semantic knowledge mediate variability in episodic memory

Shao-Fang Wang[1], Jiefeng Jiang[2], Tyler Bonnen[1], Chris Iyer[1], Anthony D. Wagner[1,3]
[1]Dept. of Psychology, [3]Wu Tsai Neurosciences Institute, Stanford University; [2]Dept. of Psychological & Brain Sciences, University of Iowa

Introduction

Computational models and behavioral evidence indicate that semantic structure has a powerful effect on episodic memory[1,2]. Emerging neural data suggest that similarity in cortical activity patterns of events reflect, in part, similarity of semantic knowledge and relates to episodic memory decisions[3,4].

- Cortical activity patterns are similarity-based: the amount of shared features across episodes influences the similarity of two events' cortical representations
- Hippocampal activity patterns are often separated: hippocampus orthogonalizes activity patterns for similar episodes[5]

Aim1: Do model-based measures of semantic similarity predict later recognition memory?

- Do semantic similarity measurements derived from a Natural Language Processing (NLP) model predict human memory behavior?

Aim2: How does semantic similarity influence the similarity of cortical and hippocampal encoding patterns?

- How does cortical pattern similarity of events affect hippocampal pattern similarity and subsequent memory?

Stimuli Design

We generated word lists with varying degrees of within-list semantic similarity using word embeddings from NLP model GloVe[6]. Similarity was quantified as cosine similarity (cs).

Common English nouns (1-gram, 14592 words) from Brysbaert et al., 2014: excluded nouns with frequency and mean concreteness ratings <5th percentile, yielding a dataset of 7383 words.

Target word candidates: We selected nouns with >50th percentile of frequency and mean concreteness rating from the noun dataset (262 words). We selected nouns with < 0.4 cs with other nouns in the target word dataset as the target word candidates (119 words).

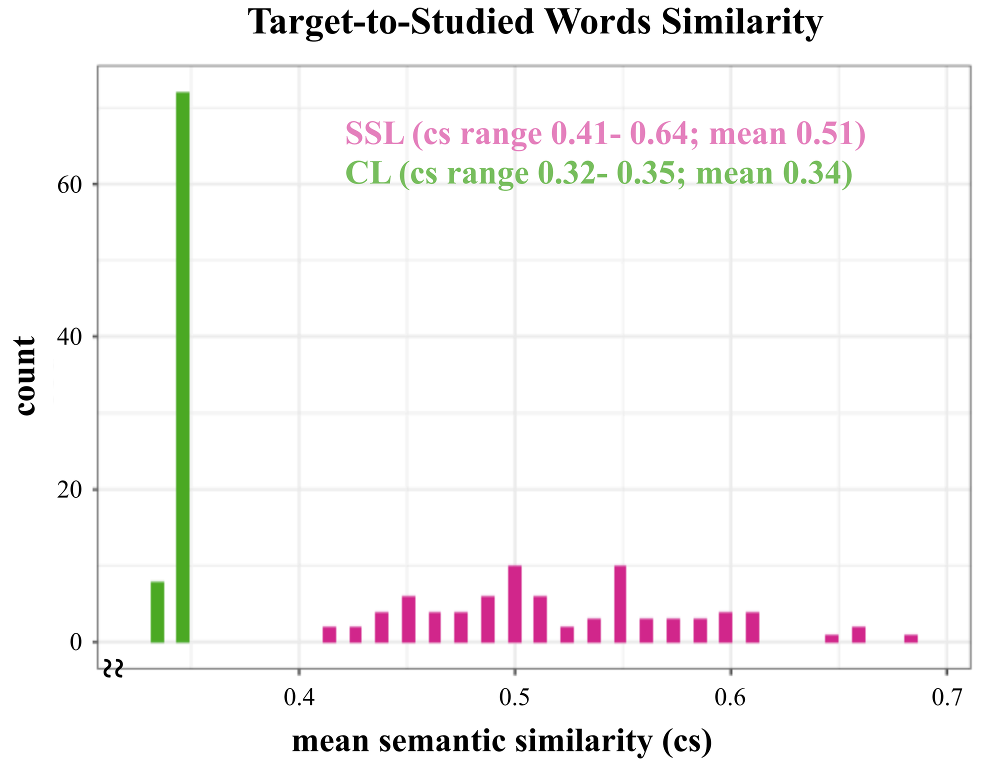
Studied words: We calculated cs between target word candidates and all other words in the noun dataset and selected the closest 5 words to each target word candidate as studied words (excluding studied words with > 0.4 cs with any other target words).

- **Semantically similar list (SSL):** cs between all studied words and the target word > 0.4. This process resulted in 80 target words.
- **Control list (CL):** Used the same 80 target words to generate control lists with 5 studied words with <0.35 cs with a target word.

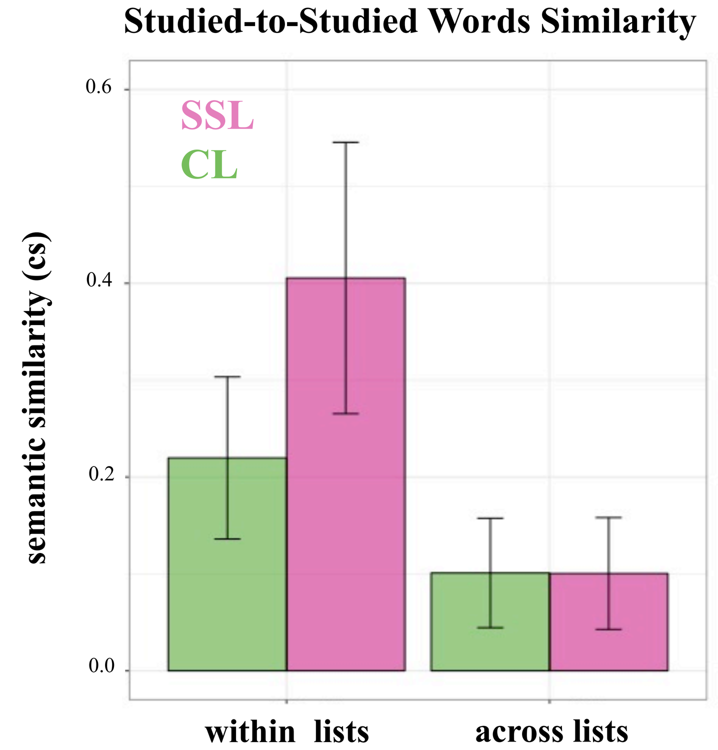
Example SSL
L001: **date**, month, deadline, beginning, year, birth
L002: **king**, throne, lord, reign, monarch, ruler

Example CL
L001: **date**, availability, countdown, periods, billing, booking
L002: **king**, century, fortress, hermit, nation, bible

Target-to-Studied Words Similarity



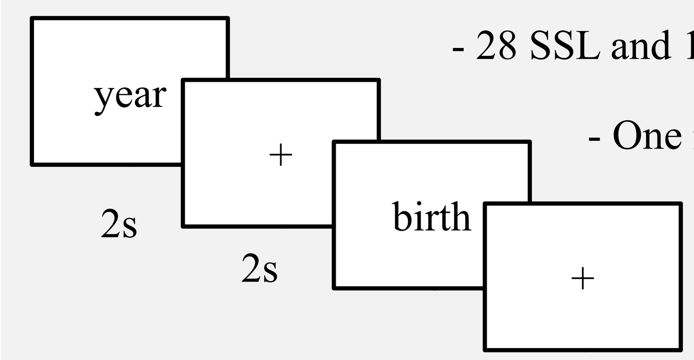
Studied-to-Studied Words Similarity



- M.E. conditions (within lists, across lists) (F(1)=1415.2, p<0.001)
- M.E. list conditions (SSL, CL) (F(1)=270.3, p<0.001)
- Interactions between conditions and list conditions (F(1)=273.5, p<0.001)

Experimental Paradigm & Results

Memory encoding phase



- 28 SSL and 12 CL
- One nonword /list
- Task: can you bring the meaning to mind (yes/no)?

→ 10 min sustained attention task (gradCPT)[7] →

Memory retrieval phase

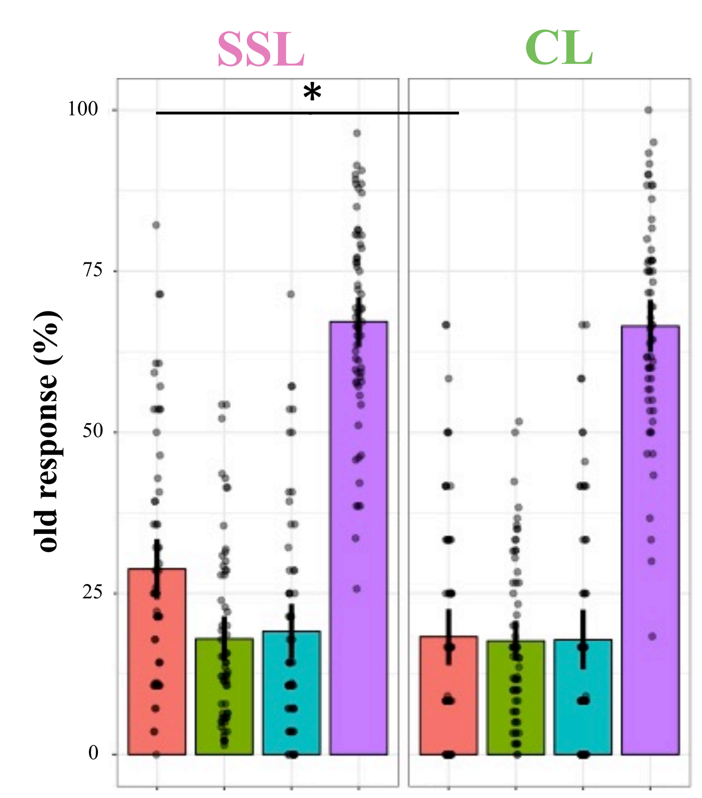
Recognition memory decisions on:
- studied words (old)
- unseen target words of studied lists (critical lures)
- unseen target words of unstudied lists (new lures)
- unseen words of unstudied lists (new)

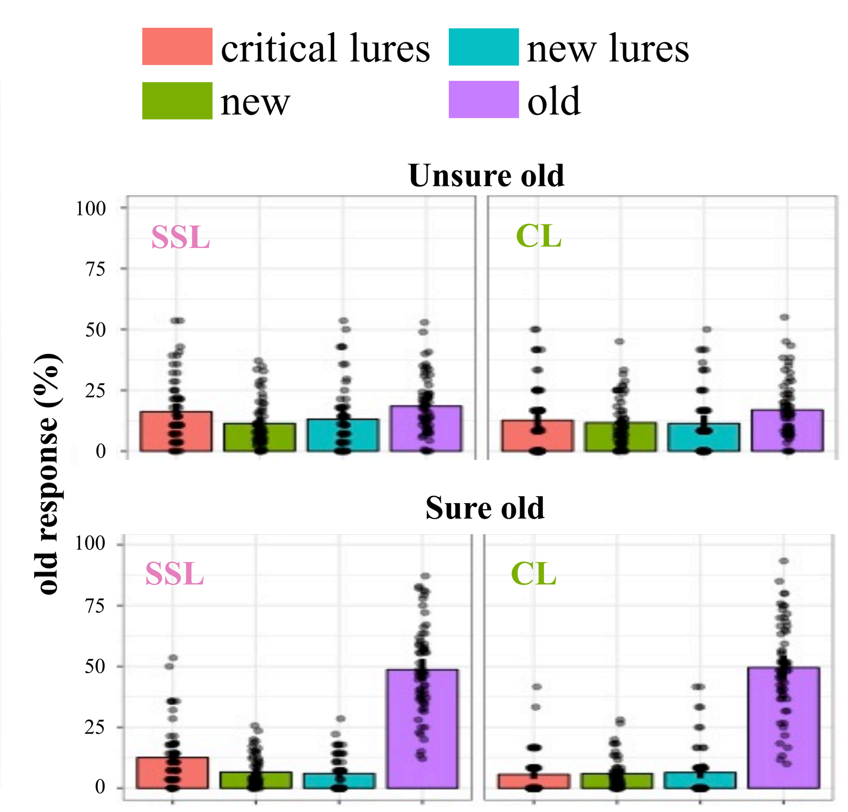
Responses: sure new, unsure new, unsure old, sure old
Word presentation 3s; ISI 1s

Critical lures: unseen target words of lists studied during encoding
L001: date, month, deadline, beginning, year, birth
Old: studied words seen during encoding

New lures: unseen target words of lists not studied during encoding
L003: message, email, reply, sender, reminder, error
New: unstudied words

Falsely recognition of critical lures higher for SSL than CL

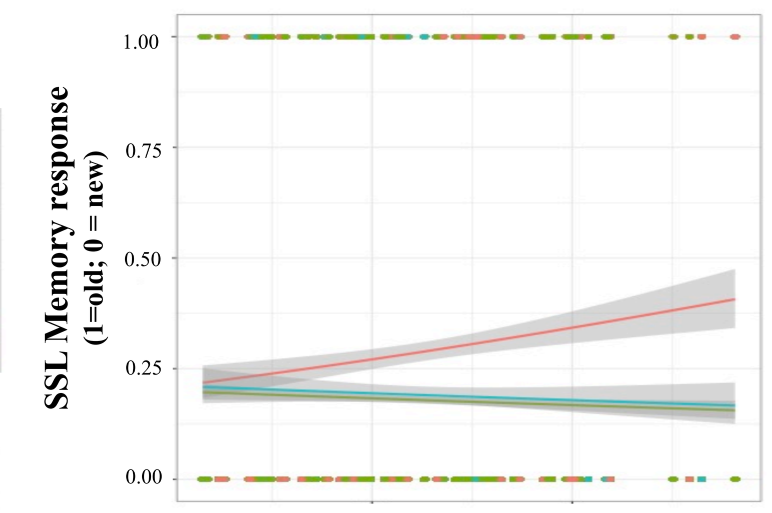




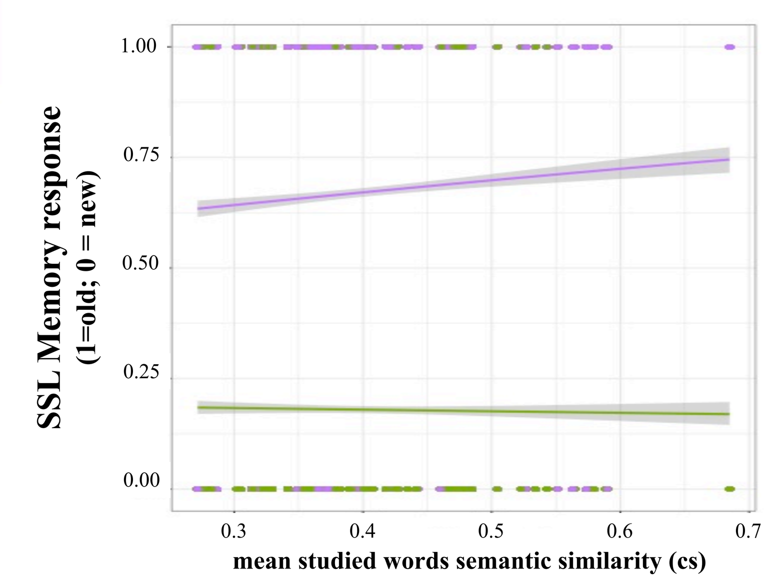
- N=62
- M.E. word identity (old, new, critical lures, new lures) (F(3,183)=246.04, p<0.001)
- M.E. list conditions (SSL, CL) (F(1,61)=19.85, p<0.001)
- Interactions between word identity and list conditions (F(3,183)=16.01, p<0.001)

- Interaction between word identity, list conditions, and confidence levels (F(3,183)=3.14, p<0.026)
- Greater high-confidence old responses for lures in SSL than in CL (t(61)= -5.57, p<0.001)

NLP–derived semantic similarity predicts false recognition and true recognition



False recognition: Interaction between semantic similarity and word identity (critical lures, new, and new lures) (new: z= -5.33, p<0.001; new lures; z = -3.96, p < 0.001)



True recognition: Interaction between semantic similarity and word identity (old and new): significant effect of semantic similarity for only old words (z=3.99, p<0.001)

Summary & Future Directions

- Semantic similarity measurements derived from an NLP model predict memory true and false recognition
- Actively collecting fMRI data while a separate group of participants view the words used in this study: we expect that cortical pattern similarity of studied and critical lure words will scale with semantic similarity. We also expect that hippocampal pattern similarity will track memory performance.

References: 1. Bower et al., 1979; 2. Kumaran et al., 2016 ; 3. Martin 2006 ; 4. Patterson et al., 2007; 5.Treves and Rolls 1994; 6. Pennington et al., 2014 ; 7. Esterman et al., 2013

Funding: Marcus and Amalia Wallenberg Foundation