Causal Representation of Abstract Phonological Properties in Temporal Cortex

David Gow[1,2,3] , Enes Avcu[1], Adriana Schoenhaut[1], David Sorensen[4], Skyla Lynch[1], Seppo Ahlfors[1,2]
gow@helix.mgh.harvard.edu
Massachusetts General Hospital/Harvard Medical School[1], Athinoula A. Martinos Center for Biomedical Imaging[2], Salem State University[3], Harvard-MIT Division of Health Sciences and Technology [4]

The Big Questions

- How abstract are neural representations of phonological structure?
  - Do they merely encode phonemic structure?
  - Do they also encode more abstract structural representations that could support hard-to-explain instances of linguistic generativity?

Abstraction and Linguistic Generativity

- The ability to create, understand, or evaluate the grammaticality of novel constructions is a central challenge for mechanistic accounts of language processing [1,2,3].

Colorless green ideas sleep furiously

Blik > Bnik > Bdik

- Whether generativity is the product of operations over variables as in generative linguistics[1,4], or the discovery of abstract features and representations as in deep learning[5] or Bayesian inference models[2], it is clear that **the scope of representation determines the productivity of processes that act on them.**

Test case: Reduplication

- The ability to interpret novel forms that resemble familiar ones is attributable to low-level mechanisms (graded activation, overlapping distributed representation).
- **Reduplication**, a widespread natural phonological process involving the copying phonology for morphological purposes, does not depend on low level acoustic-phonetic similarity.

Does English have reduplication reduplication, or do we just sometimes repeat ourselves?

- The ability to use reduplicated forms appears to reflect a cognitive primitive, given its universal role in early language contact phenomena (e.g., pidgins)[6], productive morphological role in many diverse human languages[7], salience to human infants[8] and relative learnability in artificial grammar paradigms[9].

Hypothesis: Reduplication depends on the ability to represent patterns of repetition regardless of their specific phonemic or syllabic makeup.

Establishing "representation" representation

Our criteria for claiming causal neural representation[10]:

- Decodable in temporally plausible windows by ROIs independently associated with relevant operations
- Affect downstream processing
- Behavioral performance contingent on hypothesized representations

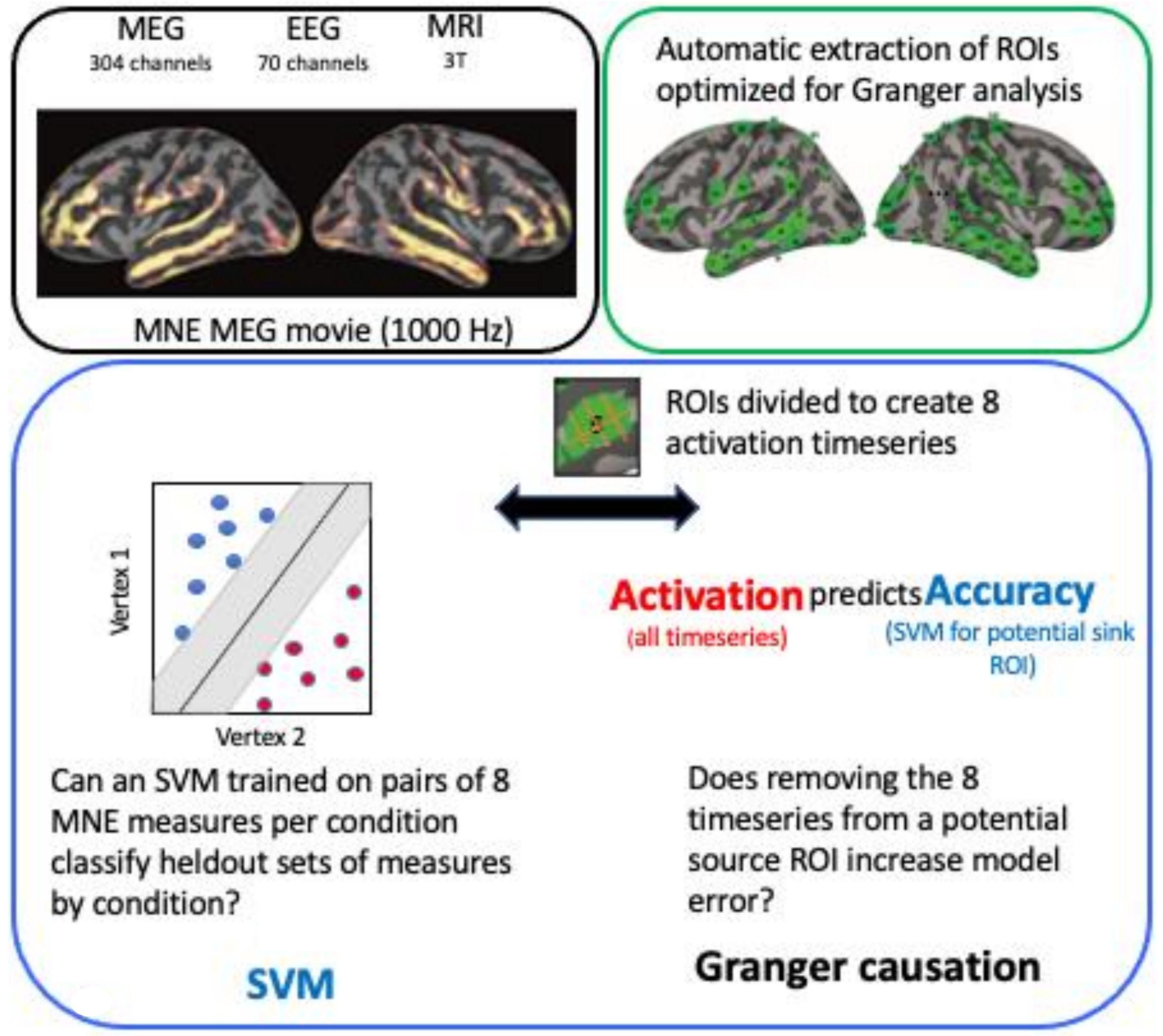
A Simple Artificial Grammar Learning/Judgment Task

- **Task**: After hearing 30 exposure CVCVCV nonsense words following a common rule (AAB, ABB or ABA) spoken by a female talker, subjects were asked to determine whether a nonoverlapping set of test CVCVCV words "came from the same imaginary language".

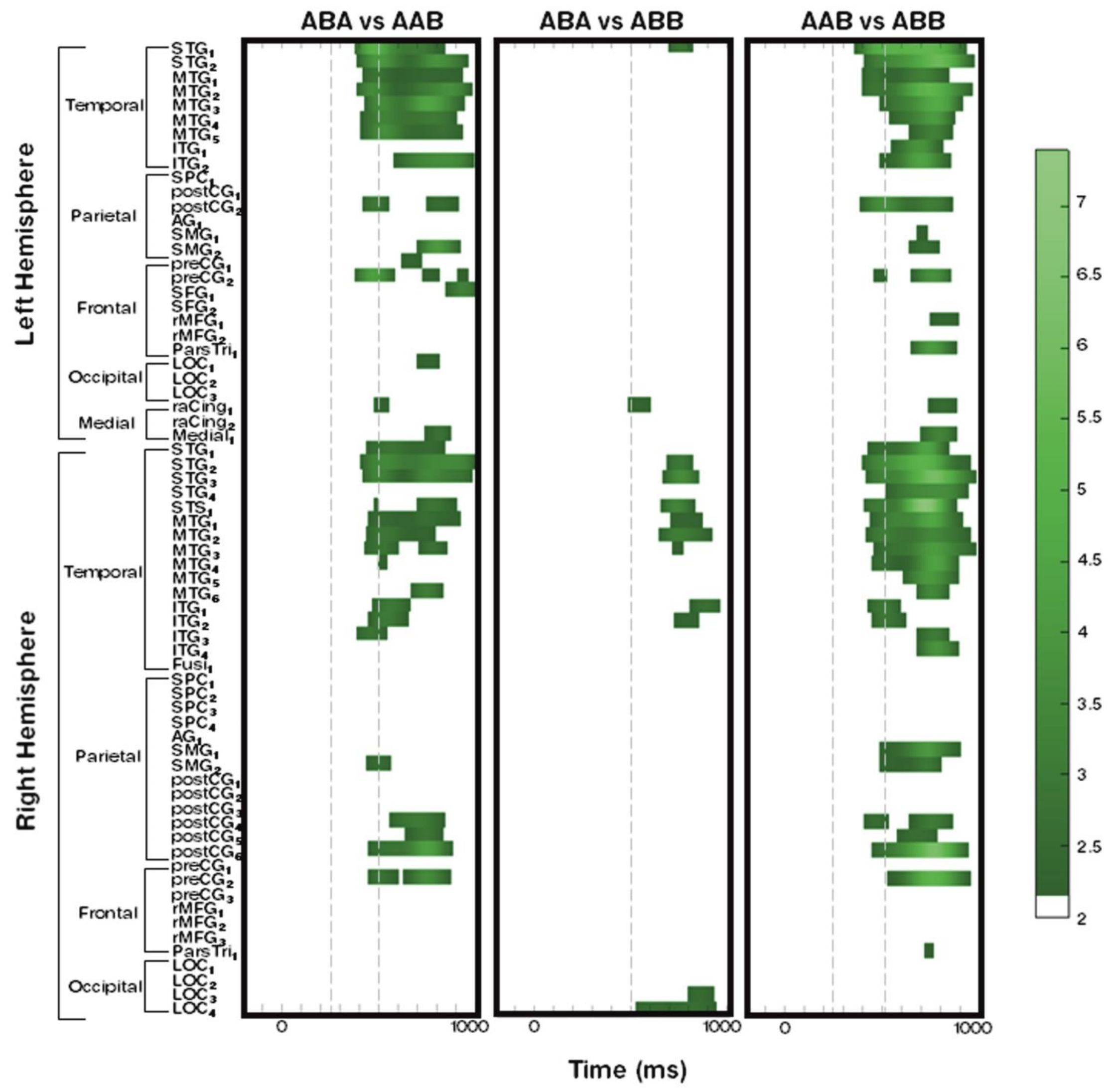
- **Stimuli**: 22 unique CV syllables were recorded, normalized for duration (200 msec), and concatenated to form 720 unique trisyllabic nonse words consistent with 3 repetition patterns: AAB, ABB, and ABA. The same CV syllables occurred an equal number of times in all syllable positions. No nonse tokens were repeated. Phonological overlap between exposure and test words was minimized.

Rule	Examples	
AAB	babapo	popoba
ABB	bapopo	pobaba
ABA	bapoba	pobapo

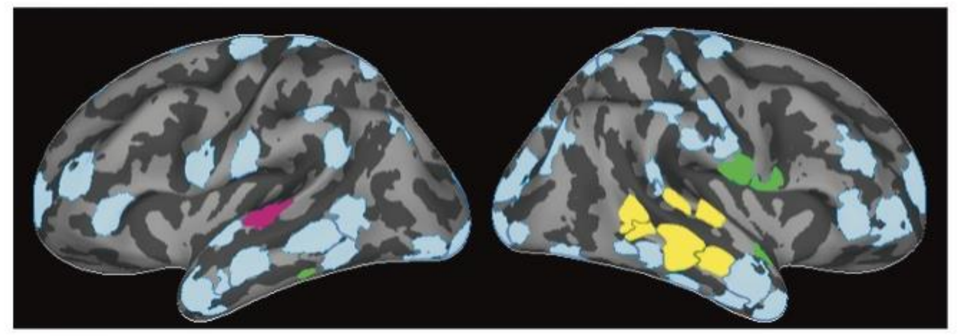
Image reconstruction, neural decoding (SVM)[11] and integrated effective connectivity (Granger causality)[12,13] analyses



Results
Pairwise SVM decoding accuracy by ROI



Distribution of ROIs showing reliable decoding versus reduplication-based activation differences across 5 contrasts

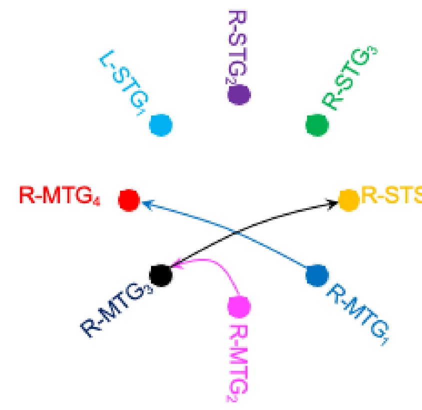


Reliable activation difference

Reliable decoding

Reliable decoding and activation difference

Significant information flow (Granger causation) between decoding ROIs in which flow was stronger for reduplicated conditions the decoding period in 2 contrasts.



Summary of Primary Results

- Subjects performed the task with high accuracy (95.9%)
- Decoding performance was dissociable from potential repetition enhancement/inhibition effects, and decoding based on response in the judgment task (secondary analyses, not shown here).
- Decoding ROIs were limited to temporal lobe structures independently associated with acoustic phonetic, lexical and morphological representation (STG, STS, MTG)
- Integrated decoding-effective connectivity analyses showed that decodable activation patterns in several temporal regions caused downstream effects on the decoding of the same contrasts in other temporal regions

Discussion & Conclusion

- Although reduplication has been cited as contrastive evidence for rule/constraint based over associative processing, we are agnostic, because **any property that can be represented should be generalizable by associative mechanisms**.
- More work is needed to determine if other forms of abstract representation underlie other forms of phonological, morphological or syntactic generativity.
- Integrated decoding-effective connectivity analyses may clarify claims about neural representation by providing a way to demonstrate that decodable information plays a causal role in downstream processing

References

[1] Chomsky, N. (1957). *Syntactic structures*. Mouton.
[2] Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, *105*(31), 10687–10692.
[3] Marcus, G. F. (2001). The algebraic mind : integrating connectionism and cognitive science.
[4] Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.
[5] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444.
[6] Muysken, P. (2013). Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition*(16), 709-773.
[7]Rubino, C. (2013). *Reduplication*. Max Planck Institute for Evolutionary Anthropology.
[8]Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77-80
[9]Tunney, R. J., & Altmann, G. T. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory and Cognition.*, *27*(3), 614-639.
[10] Dennett, D. C. (1987). *The intentional stance*. MIT Press.
[11]Cichy, R.M., Pantazis, D., & Olivia, A. (2012). Resolving human object recognition in space and time. Nature Neuroscience, 17, 455-462.
[12]Milde, T., Leistritz, L., Astolfi, L., Miltner, W. H., Weiss, T., Babiloni, F., & Witte, H. (2010). A new Kalman filter approach for the estimation of high-dimensional time-variant multivariate AR models and its application in analysis of laser-evoked brain potentials. *NeuroImage*, *50*(3), 960-969.
[13] Gow, D. W., Jr., & Caplan, D. N. (2012). New levels of language processing complexity and organization revealed by granger causation. *Frontiers in Psychology*, *3*, 506.
Look for:
Gow, D.W., Avcu, E., Schoenhaut, & Ahlfors, S. (under review). Abstract representations in temporal cortex support generative linguistic processing.

Acknowledgements: This work was supported by NIDCD grant R01DC015455 and benefited from funding from CFNT grant P41EB015896 and S10RR014978. We would like to thank Olivia Newman for assistance in data collection, analysis and visualization, and Ev Fedorenko, Ray Jackendoff, Gerry Altmann and David Caplan for feedback on related manuscripts.