fMRI reveals language-specific predictive coding during naturalistic sentence comprehension

What are the neuro-cognitive mechanisms supporting predictive language processing? In particular, to what extent does prediction during comprehension recruit language-specific mechanisms, and to what extent does it rely on general cognitive mechanisms supporting prediction across domains? While decades of psycholinguistic research have advanced our understanding of predictive language processing [31, 34, 51, 40, 21, 43, 49, 44, 30], this evidence has largely been obtained through behavioral (e.g. eye-tracking) or electrophysiological (e.g., EEG/MEG) measures, which can reliably identify global response patterns but are not ideal for disentangling the respective contributions to prediction of functionally distinct mechanisms. In this study, we therefore used fMRI to determine whether a signature of predictive coding during language comprehension — positive response to *n*-gram surprisal — is primarily evident in (1) domain-specific cortical circuits, namely, the left fronto-temporal language network [4, 18], or (2) domain-general circuits, namely, the bilateral "Multiple Demand" (MD) network [15], which supports top-down executive control across both linguistic and non-linguistic tasks [17]. On the one hand, given that the language network stores linguistic knowledge, including plausibly the statistics of language input, it might directly carry out predictive processing (hypothesis 1). This result would be consistent with a growing body of research in cognitive neuroscience supporting prediction as a "canonical computation" [29, 42] locally implemented in domain-specific circuits [36, 38, 1, 7, 2, 52, 42, 29]. On the other hand, given that the MD network encodes predictive signals across domains and relays them as feedback to other regions [50, 9, 16, 53, 8], it might be recruited to predict upcoming words [56] (hypothesis 2). This outcome would be consistent with the general scaling of MD activity with cognitive effort, since surprisal reliably indexes such effort [44].

To distinguish between these hypotheses, we scanned subjects while they passively listened to stories. This naturalistic paradigm complements previous work on linguistic prediction that has relied on carefully constructed stimuli, which may introduce task artifacts that do not generalize to everyday cognition [12, 26].[1] Despite the growing interest in fMRI studies of naturalistic comprehension [47, 57, 46, 55, 54, 25, 28, 6, 45, 5, 27, 13, 10, 11, 3], only a handful of such studies have directly investigated effects of *n*-gram surprisal [56, 6, 33].[2] Further, the conclusions from these studies are complicated by reverse inference from anatomy back to function [37]. To circumvent this issue, here we used "localizer tasks" to functionally define the language and MD networks in each individual subject. Moreover, to our knowledge, ours is the largest fMRI investigation to-date (78 subjects) of prediction effects in naturalistic comprehension.

Functional regions of interest (fROIs) were identified with a previously validated reading task of sentences and nonword lists, using the sentences $>$ nonwords contrast for language fROIs and the opposite contrast for MD fROIs [20, 19]. We then recorded BOLD signal time-series from these fROIs while subjects listened to materials from the Natural Stories corpus [22].[3] We applied a recently-developed deconvolutional time series regression (DTSR) model [41] to overcome limitations of naturalistic language stimuli for hemodynamic response function (HRF) discovery rather than assuming a canonical HRF shape (cf. e.g. [56, 6, 33]). Prediction effects in the two networks were evaluated via ablative paired permutation testing of models containing a fixed effect for word prediction (5-gram surprisal) against an empirically-motivated baseline[4] on half the data (held-out from training), pooling over all fROIs in each network.

We found a positive effect of prediction in the language network ($p = 0.0001$) but not the MD network ($p = 0.5$), despite greater power to detect effects in MD because it contained more fROIs (20 vs. 6). When analyzing individual language fROIs, we found prediction effects in 4 regions (anterior and posterior middle temporal gyrus, inferior and middle frontal gyrus) but not in 2 others (angular gyrus, orbital part of the inferior frontal gyrus),[5] suggesting that prediction effort is distributed across multiple non-adjacent language regions, with potential differentiation within the language network in the degree of recruitment for prediction.

Conclusion: Our results support hypothesis 1, i.e., predictive coding in language is carried by a language-specific (cf. domain-general) mechanism.

[1] Minimizing such artifacts is crucial in studies of the MD network, which is highly sensitive to task variables [35, 48, 14].

[2] [6] and [27] also studied and found positive effects of unlexicalized *syntactic surprisal* in some regions, which has been interpreted as evidence of a specifically syntactic prediction mechanism. In this study, we are targeting arguably more basic effects of word prediction [21], leaving differentiation of lexicalized vs. unlexicalized prediction to future work.

[3] While this study does not directly investigate locality effects [24, 32, 39], dependency locality integration cost [23] is not well correlated with *5-gram surprisal* in our stimuli ($\rho = 0.183$), suggesting that such effects do not drive our results.

[4] *Word rate*, *sound power*, and *unigram log probability*, with by-subject random intercepts

[5] By ROI threshold P $<$ 0.05

References

[1] Arjen Alink, Caspar M Schwiedrzik, Axel Kohler, et al. *Journal of Neuroscience*, 2010.
[2] Andre M Bastos, W Martin Usrey, Rick A Adams, et al. *Neuron*, 2012.
[3] Shohini Bhattasali, John Hale, Christophe Pallier, et al. *Proceedings of the Society for Computation in Linguistics (SCiL) 2018*, 2018.
[4] Jeffrey R Binder, Julie A Frost, Thomas A Hammeke, et al. *Journal of Neuroscience*, 1997.
[5] Jonathan Brennan. *Language and Linguistics Compass*, 2016.
[6] Jonathan Brennan, Edward P Stabler, Sarah E Van Wagenen, et al. *Brain and language*, 2016.
[7] Andreja Bubic, D Yves Von Cramon, and Ricarda I Schubotz. *Frontiers in human neuroscience*, 2010.
[8] Zenas C Chao, Kana Takaura, Liping Wang, et al. *Neuron*, 2018.
[9] Tamara C Cristescu, Joseph T Devlin, and Anna C Nobre. *Neuroimage*, 2006.
[10] Wendy A de Heer, Alexander G Huth, Thomas L Griffiths, et al. *Journal of Neuroscience*, 2017.
[11] Morteza Dehghani, Reihane Boghrati, Kingson Man, et al. *Human brain mapping*, 2017.
[12] Vera Demberg and Frank Keller. *Cognition*, 2008.
[13] Rutvik H Desai, Wonil Choi, Vicky T Lai, and John M Henderson. *Journal of Neuroscience*, 2016.
[14] Mark D'Esposito and Bradley R Postle. *Annual review of psychology*, 2015.
[15] John Duncan. *Trends in cognitive sciences*, 2010.
[16] Tobias Egner, Jim M P Monti, Emily H Trittschuh, et al. *Journal of Neuroscience*, 2008.
[17] Evelina Fedorenko. *Frontiers in psychology*, 2014.
[18] Evelina Fedorenko, Michael K Behr, and Nancy Kanwisher. *Proceedings of the National Academy of Sciences*, 2011.
[19] Evelina Fedorenko, John Duncan, and Nancy Kanwisher. *Proceedings of the National Academy of Sciences*, 2013.
[20] Evelina Fedorenko, Po-Jang Hsieh, Alfonso Nieto-Castañón, et al. *Journal of neurophysiology*, 2010.
[21] Stefan L Frank and Rens Bod. *Psychological Science*, 6 2011.
[22] Richard Futrell, Edward Gibson, Harry J . Tily, et al. In Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 5 2018.
[23] Edward Gibson. In *Image, language, brain: Papers from the first mind articulation project symposium*, 2000.
[24] Daniel J Grodner and Edward Gibson. *Cognitive Science*, 2005.
[25] John Hale, David Lutz, Wen-Ming Luh, and Jonathan Brennan. In *Proceedings of the 6th workshop on cognitive modeling and computational linguistics*, 2015.
[26] Uri Hasson and Christopher J Honey. *NeuroImage*, 2012.
[27] John M Henderson, Wonil Choi, Matthew W Lowder, and Fernanda Ferreira. *Neuroimage*, 2016.
[28] Alexander G Huth, Wendy A de Heer, Thomas L Griffiths, et al. *Nature*, 4 2016.
[29] Georg B Keller and Thomas D Mrsic-Flogel. *Neuron*, 2018.
[30] Gina R Kuperberg and T Florian Jaeger. *Language, cognition and neuroscience*, 2016.
[31] M Kutas and S A Hillyard. *Nature*, 1984.
[32] Roger Levy and Edward Gibson. *Frontiers in Psychology*, 2013.
[33] Alessandro Lopopolo, Stefan L Frank, Antal den Bosch, and Roel M Willems. *PloS one*, 2017.
[34] Maryellen C MacDonald, Neal J Pearlmutter, and Mark S Seidenberg. *Psychological Review*, 1994.
[35] Earl K Miller and Jonathan D Cohen. *Annual review of neuroscience*, 2001.
[36] P Read Montague, Peter Dayan, and Terrence J Sejnowski. *Journal of neuroscience*, 1996.
[37] Russell A Poldrack. *Neuron*, 2011.
[38] Rajesh P N Rao and Dana H Ballard. *Nature neuroscience*, 1999.
[39] Nathan E Rasmussen and William Schuler. *Cognitive science*, 2018.
[40] Keith Rayner, Jane Ashby, Alexander Pollatsek, and Erik D Reichle. *Journal of Experimental Psychology: Human Perception and Performance*, 2004.
[41] Cory Shain and William Schuler. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018.
[42] Yosef Singer, Yayoi Teramoto, Ben D B Willmore, et al. *eLife*, 2018.
[43] Nathaniel J Smith and Roger Levy. In *Proceedings of the 33rd CogSci Conference*, 2011.
[44] Nathaniel J Smith and Roger Levy. *Cognition*, 2013.
[45] Mariam R Sood and Martin I Sereno. *Human brain mapping*, 2016.
[46] Nicole K Speer, Jeremy R Reynolds, Khena M Swallow, and Jeffrey M Zacks. *Psychological science*, 2009.
[47] Nicole K Speer, Jeffrey M Zacks, and Jeremy R Reynolds. *Psychological Science*, 2007.
[48] Kartik K Sreenivasan, Clayton E Curtis, and Mark DEsposito. *Trends in cognitive sciences*, 2014.
[49] Adrian Staub and Ashley Benatar. *Psychonomic Bulletin \& Review*, 2013.
[50] Bryan A Strange, Andrew Duggins, William Penny, et al. *Neural Networks*, 2005.
[51] Michael K Tanenhaus, Michael J Spivey-Knowlton, Kathleen M Eberhard, and Julie C E Sedivy. *Science*, 1995.
[52] Catherine Wacongne, Jean-Pierre Changeux, and Stanislas Dehaene. *Journal of Neuroscience*, 2012.
[53] Catherine Wacongne, Etienne Labyt, Virginie van Wassenhove, et al. *Proceedings of the National Academy of Sciences*, 2011.
[54] Leila Wehbe, Brian Murphy, Partha Talukdar, et al. *PloS one*, 2014.
[55] Carin Whitney, Walter Huber, Juliane Klann, et al. *Neuroimage*, 2009.
[56] Roel M Willems, Stefan L Frank, Annabel D Nijhof, et al. *Cerebral Cortex*, 2015.
[57] Tal Yarkoni, Nicole K Speer, and Jeffrey M Zacks. *Neuroimage*, 2008.