fMRI reveals language-specific predictive coding during naturalistic sentence comprehension

Cory Shain[1], Idan Blank[2], Marten van Schijndel[3], Evelina Fedorenko[2], William Schuler[1] [1]OSU, [2]MIT, [3]JHU

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 [39, 42, 60, 49, 25, 52, 58, 53, 38], 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, 22], or (2) domain-general circuits, namely, the bilateral "Multiple Demand" (MD) network [19], which supports top-down executive control across both linguistic and non-linguistic tasks [21]. 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" [36, 51] locally implemented in domain-specific circuits [44, 47, 1, 9, 2, 61, 51, 36]. On the other hand, given that the MD network encodes predictive signals across domains and relays them as feedback to other regions [59, 12, 20, 62, 11], it might be recruited to predict upcoming words [65] (hypothesis 2). This outcome would be consistent with the general scaling of MD activity with cognitive effort, since surprisal reliably indexes such effort [53].

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 [16, 30].[1] Despite the growing interest in fMRI studies of naturalistic comprehension [56, 66, 55, 64, 63, 29, 33, 8, 54, 7, 31, 17, 13, 15, 3], only a handful of such studies have directly investigated effects of *n*-gram surprisal [65, 8, 41].[2] Further, the conclusions from these studies are complicated by reverse inference from anatomy back to function [46]. To circumvent this issue, here we used robustly validated "localizer tasks" to functionally define the language and MD networks in each individual subject [24, 23, **?**]. Moreover, to our knowledge, ours is the largest fMRI investigation to-date (78 subjects) of prediction effects in naturalistic comprehension.

Subjects' responses in functionally localized regions of interest (fROI) was recorded while they listened to materials from the Natural Stories corpus [26]. Hemodynamic response functions (HRF) by region were estimated using deconvolutional time series regression (DTSR) [50]. Mixed-effects DTSR models were fitted to responses from each network individually and to the combined response from both networks.

Ablative out-of-sample paired permutation tests showed a significant response to 5-gram surprisal (sequential prediction) in the language network (p=0.0001***) but not the MD network (p=1.0) in individual network models and a significantly increased response to surprisal in the language network over that of the MD network in the combined model (p=0.0001***).[3] Results thus suggest that predictive coding for upcoming words is primarily a canonical computation carried out by domain-specific cortical circuits, rather than by feedback from higher, domain-general executive control circuits (hypothesis 1). This finding joins previous evidence for a high degree of compartmentalization in the language processing architecture [22, 45, 5] and contrasts with experimental results showing (1) engagement of language regions in non-linguistic tasks [14, 37, 6] and (2) engagement of executive control regions in language processing [35, 34, 32], results which might have been influenced by task artifacts that can be minimized using our naturalistic paradigm [30, 10]. By showing that prediction can be locally implemented for high-level language reasoning in humans, this finding also complements recent work on predictive coding in the mammalian brain, which has largely focused on low-level perceptual processing [9, 36, 51]

[1]Minimizing such artifacts is crucial in studies of the MD network, which is highly sensitive to task variables [43, 57, 18].

[2][8] and [31] 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 [25], leaving differentiation of lexicalized vs. unlexicalized prediction to future work.

[3]While this study does not directly investigate locality effects [28, 40, 48], dependency locality integration cost [27] is not well correlated with *5-gram surprisal* in our stimuli ($\rho = 0.183$), suggesting that such effects do not drive our results.

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] Idan Blank and Evelina Fedorenko. *Journal of Neuroscience*, 2017.
[6] Sheila E Blumstein. *Language and Linguistics Compass*, 2009.
[7] Jonathan Brennan. *Language and Linguistics Compass*, 2016.
[8] Jonathan Brennan, Edward P Stabler, Sarah E Van Wagenen, et al. *Brain and language*, 2016.
[9] Andreja Bubic, D Yves Von Cramon, and Ricarda I Schubotz. *Frontiers in human neuroscience*, 2010.
[10] Karen L Campbell and Lorraine K Tyler. *Current opinion in behavioral sciences*, 2018.
[11] Zenas C Chao, Kana Takaura, Liping Wang, et al. *Neuron*, 2018.
[12] Tamara C Cristescu, Joseph T Devlin, and Anna C Nobre. *Neuroimage*, 2006.
[13] Wendy A de Heer, Alexander G Huth, Thomas L Griffiths, et al. *Journal of Neuroscience*, 2017.
[14] Stanislas Dehaene, Elizabeth Spelke, Philippe Pinel, et al. *Science*, 1999.
[15] Morteza Dehghani, Reihane Boghrati, Kingson Man, et al. *Human brain mapping*, 2017.
[16] Vera Demberg and Frank Keller. *Cognition*, 2008.
[17] Rutvik H Desai, Wonil Choi, Vicky T Lai, and John M Henderson. *Journal of Neuroscience*, 2016.
[18] Mark D'Esposito and Bradley R Postle. *Annual review of psychology*, 2015.
[19] John Duncan. *Trends in cognitive sciences*, 2010.
[20] Tobias Egner, Jim M P Monti, Emily H Trittschuh, et al. *Journal of Neuroscience*, 2008.
[21] Evelina Fedorenko. *Frontiers in psychology*, 2014.
[22] Evelina Fedorenko, Michael K Behr, and Nancy Kanwisher. *Proceedings of the National Academy of Sciences*, 2011.
[23] Evelina Fedorenko, John Duncan, and Nancy Kanwisher. *Proceedings of the National Academy of Sciences*, 2013.
[24] Evelina Fedorenko, Po-Jang Hsieh, Alfonso Nieto-Castañón, et al. *Journal of neurophysiology*, 2010.
[25] Stefan L Frank and Rens Bod. *Psychological Science*, 6 2011.
[26] 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.
[27] Edward Gibson. In *Image, language, brain: Papers from the first mind articulation project symposium*, 2000.
[28] Daniel J Grodner and Edward Gibson. *Cognitive Science*, 2005.
[29] John Hale, David Lutz, Wen-Ming Luh, and Jonathan Brennan. In *Proceedings of the 6th workshop on cognitive modeling and computational linguistics*, 2015.
[30] Uri Hasson and Christopher J Honey. *NeuroImage*, 2012.
[31] John M Henderson, Wonil Choi, Matthew W Lowder, and Fernanda Ferreira. *Neuroimage*, 2016.
[32] Nina S Hsu and Jared M Novick. *Psychological science*, 2016.
[33] Alexander G Huth, Wendy A de Heer, Thomas L Griffiths, et al. *Nature*, 4 2016.
[34] David January, John C Trueswell, and Sharon L Thompson-Schill. *Journal of Cognitive Neuroscience*, 2009.
[35] Edith Kaan and Tamara Y Swaab. *Trends in cognitive sciences*, 2002.
[36] Georg B Keller and Thomas D Mrsic-Flogel. *Neuron*, 2018.
[37] Stefan Koelsch, Thomas C Gunter, D Yves v Cramon, et al. *Neuroimage*, 2002.
[38] Gina R Kuperberg and T Florian Jaeger. *Language, cognition and neuroscience*, 2016.
[39] M Kutas and S A Hillyard. *Nature*, 1984.
[40] Roger Levy and Edward Gibson. *Frontiers in Psychology*, 2013.
[41] Alessandro Lopopolo, Stefan L Frank, Antal den Bosch, and Roel M Willems. *PloS one*, 2017.
[42] Maryellen C MacDonald, Neal J Pearlmutter, and Mark S Seidenberg. *Psychological Review*, 1994.
[43] Earl K Miller and Jonathan D Cohen. *Annual review of neuroscience*, 2001.
[44] P Read Montague, Peter Dayan, and Terrence J Sejnowski. *Journal of neuroscience*, 1996.
[45] Martin M Monti, Lawrence M Parsons, and Daniel N Osherson. *Psychological science*, 2012.
[46] Russell A Poldrack. *Neuron*, 2011.
[47] Rajesh P N Rao and Dana H Ballard. *Nature neuroscience*, 1999.
[48] Nathan E Rasmussen and William Schuler. *Cognitive science*, 2018.
[49] Keith Rayner, Jane Ashby, Alexander Pollatsek, and Erik D Reichle. *Journal of Experimental Psychology: Human Perception and Performance*, 2004.
[50] Cory Shain and William Schuler. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018.
[51] Yosef Singer, Yayoi Teramoto, Ben D B Willmore, et al. *eLife*, 2018.
[52] Nathaniel J Smith and Roger Levy. In *Proceedings of the 33rd CogSci Conference*, 2011.
[53] Nathaniel J Smith and Roger Levy. *Cognition*, 2013.
[54] Mariam R Sood and Martin I Sereno. *Human brain mapping*, 2016.
[55] Nicole K Speer, Jeremy R Reynolds, Khena M Swallow, and Jeffrey M Zacks. *Psychological science*, 2009.
[56] Nicole K Speer, Jeffrey M Zacks, and Jeremy R Reynolds. *Psychological Science*, 2007.
[57] Kartik K Sreenivasan, Clayton E Curtis, and Mark DEsposito. *Trends in cognitive sciences*, 2014.
[58] Adrian Staub and Ashley Benatar. *Psychonomic Bulletin \& Review*, 2013.
[59] Bryan A Strange, Andrew Duggins, William Penny, et al. *Neural Networks*, 2005.
[60] Michael K Tanenhaus, Michael J Spivey-Knowlton, Kathleen M Eberhard, and Julie C E Sedivy. *Science*, 1995.
[61] Catherine Wacongne, Jean-Pierre Changeux, and Stanislas Dehaene. *Journal of Neuroscience*, 2012.
[62] Catherine Wacongne, Etienne Labyt, Virginie van Wassenhove, et al. *Proceedings of the National Academy of Sciences*, 2011.
[63] Leila Wehbe, Brian Murphy, Partha Talukdar, et al. *PloS one*, 2014.
[64] Carin Whitney, Walter Huber, Juliane Klann, et al. *Neuroimage*, 2009.
[65] Roel M Willems, Stefan L Frank, Annabel D Nijhof, et al. *Cerebral Cortex*, 2015.
[66] Tal Yarkoni, Nicole K Speer, and Jeffrey M Zacks. *Neuroimage*, 2008.