How rational is the brain?
A probabilistic generative framework of language comprehension
Gina Kuperberg MD PhD

A large body of evidence suggests that comprehenders are able to use the preceding context, in combination with their stored linguistic and non-linguistic knowledge, to probabilistically predict upcoming inputs (Kuperberg & Jaeger, 2016). So long as our probabilistic knowledge mirrors the statistics of the linguistic input, this anticipatory processing provides a rational way of ensuring that comprehension is both fast and accurate. But what exactly does rational mean, given that our brains have limited resources, and that we sometimes encounter inputs that are completely unexpected?

One prominent theory of rational comprehension — surprisal theory— argues that we use the prior context to pre-activate even very low probability continuations (Hale, 2001; Levy, 2008; Smith & Levy, 2013). According to this theory, a word's processing is predicted by its negative log probability, given the context. It therefore assumes that we devote more resources to pre-activating information that is less likely to appear in the bottom-up input. It also assumes that the difficulty of lexical access and the difficulty of building a message-level representation can be collapsed into a single processing stage. I will present evidence that challenges these two assumptions.

First, I will present data from a series of controlled behavioral and event-related potential (ERP) experimental studies, which show that the relationship between the probability of an incoming word and its processing is, in fact, linear rather than logarithmic (Brothers & Kuperberg, 2019). This contradicts the corpus-based findings reported by Smith & Levy (2013). This linear relationship is also seen in a meta-analysis of existing eye tracking studies. These findings suggest that we probabilistically predict upcoming words in proportion to their likelihood of actually appearing. Second, I will present neural data (ERP, MEG and fMRI) showing that, in some situations, low probability words in discourse contexts produce additional later responses that are not seen to higher probability words (e.g. Kuperberg, Brothers, & Wlotko, in press). These data suggest that lexical access and message-level integration are not always mediated by the same neural mechanism.

I will argue that these findings can be understood within a *hierarchical generative framework* of language comprehension (Kuperberg & Jaeger, 2016; Kuperberg, Brothers, & Wlotko, in press). Within this framework, the comprehender draws upon a hierarchical generative model that she believes mirrors the statistical structure of her current communicative environment. Probabilistic predictions are passed down from higher to lower levels of the hierarchy in proportion to the certainty of her high-level beliefs. Bottom-up information that matches these predictions is 'explained away', leading to facilitated processing. Unpredicted/unexplained information that cannot be explained at lower levels drives belief updating at higher levels of the hierarchy, or, if it conflicts with the structure of the current generative model, triggers reanalysis and adaptation. This architecture can be understood within a bounded rational framework that assumes that language evolved to support communication across brains with limited metabolic resources, in communicative environments that vary systematically across situations.

Brothers, T., & Kuperberg, G. R. (2019). Linear (not logarithmic) effects of lexical predictability: A challenge to surprisal theory. Presented at the 32nd Annual CUNY Conference on Human Sentence Processing, Boulder, CO.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. Presented at the Proceedings of the North American Chapter of the Association for Computational Linguistics on Language technologies (NAACL '01), Pittsburgh, PA.

Kuperberg, G. R., Brothers, T., & Wlotko, E. (in press). A Tale of Two Positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. Journal of Cognitive Neuroscience.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? Language, Cognition and Neuroscience, 31(1), 32-59. doi:10.1080/23273798.2015.1102299

Levy, R. (2008). Expectation-based syntactic comprehension. Cognition, 106(3), 1126-1177. doi:10.1016/j.cognition.2007.05.006

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. Cognition, 128(3), 302-319. doi:10.1016/j.cognition.2013.02.013