The neural basis of the negativity bias: Insights from computational models and spatial similarity analysis of EEG

Victoria Sharpe, Lin Wang, Nate Delaney-Busch & Gina Kuperberg

When attending to emotional valence, we devote more attentional resources to evaluating negatively-valenced stimuli than positive or neutral stimuli – the so-called negativity-bias. Neurally, this bias manifests as a larger late positivity ERP evoked by negative (versus positive and neutral) stimuli, observed from 500ms after stimulus onset. Here we asked what representations are activated or re-activated during this late stage of evaluative processing. Using semantic vectors generated by word2vec, we analyzed over 13,000 words and showed that similarity among negative words was greater than among positive or neutral words. To show that these results could not be explained by confounding variables, we also assembled a set of 467 words, which varied in valence, but were matched on arousal and lexical properties. Using word2vec and another association-based computational model, we again showed that similarity among negative words was greater than among positive or neutral words, irrespective of differences in similarity of arousal or lexical properties. We then recorded EEG as 22 healthy adults read and judged the valence of this same set of words. ERPs confirmed that negative words evoked a larger late positivity between 500 and 800 ms than either neutral or positive words. A spatial similarity analysis showed that, between 600-700ms following word onset, the spatial pattern of neural activity produced by the negative words was more similar than that produced by either the positive or the neutral words. Together, these data suggest that the prolonged neural processing associated with negatively-valenced stimuli may reflect evaluation of underlying semantic representations.