Linear (not logarithmic) effects of lexical predictability: A challenge to surprisal theory
Trevor Brothers1,2, Gina Kuperberg1,2
1 Dept of Psychology, Tufts University, 2 Massachusetts General Hospital

Introduction
While it is clear that unpredictable linguistic content incurs processing costs, different models of sentence comprehension predict different linking functions relating context-based word probabilities and processing difficulty. Surprisal Theory (Levy, 2008) and Bayesian Reader (Norris, 2009) predict a logarithmic relationship, with low probability words incurring additional processing costs. Proportional pre-activation accounts predict a linear relationship between probability and RTs, assuming 1) finite pre-activation and 2) behavioral facilitation proportional to the strength/degree of activation across features.

To test these models, we developed 216 sentence triplets which manipulated the semantic predictability of a single critical word.

High cloze: 90% ± 0.4%, range: 100% - 60%
Mod. cloze: 20% ± 0.6%, range: 50% - 7%
Low cloze: 1% ± 0.1%, range: 0% - 5%

H: For Halloween, they were carving a large pumpkin at the farmer’s market.
M: For the pie, they bought a large pumpkin at the farmer’s market.
L: Before the party, they bought a large pumpkin at the farmer’s market.

Experiment 1: Self-paced Reading (N = 216)
Participants read sentences (self-paced moving window) and answered comprehension questions. RTs were analyzed for a three-word critical region.

Experiment 2: Cross-modal Picture Naming (N = 36)
Participants heard sentence contexts and named pictures corresponding to the CW, appearing 250ms after sentence offset.

Experiment 3: ERP Study (N = 19)
Participants read sentences (self-paced RSVP) as EEG was recorded from the scalp. The amplitude of the N400 was measured, 300-500ms after CW onset.

Summary of Results
• Across all three experiments, we saw a robust linear relationship between word probability and processing difficulty.
• Log-transformed cloze resulted in significant reductions in model fit, and produced non-linear GAM fits (mgcv).
• The pattern of effects (90% vs. 20% > 20% vs. 1%) was exactly opposite the predictions of surprisal/Bayesian Reader.
• In Experiment 2, at the single-trial level, behavioral responses for all 36 participants were better fit by a linear function (mean r = -.435) than a logarithmic function (mean r = -.399, p < .001).
• These findings are incompatible with the two main claims of surprisal theory: 1) that RTs covary with -log(prob|context), and 2) lexical probabilities can provide a unitary “causal bottleneck” for explaining all forms of comprehension difficulty.

FAQs
Q: Isn’t cloze bad at estimating probabilities below 1%?
A: We may have overestimated the probability of some very low-cloze items (by replacing values of 0% with 0.5%). But if so, a log function would provide even worse empirical fit to the data.
Q: Can corpus-based probabilities account for additional variance in the very low end of the probability scale?
A: No, BNC trigram/LSTM probabilities (log or linear) accounted for no additional variance above and beyond the effects of cloze.
Q: Is natural language actually this predictable?
A: Yes. The mean cloze of words in connected texts is ~18% with the vast majority of words falling within the 1%-100% cloze range (Luke & Christianson, 2015).
Q: Should I log-transform my probability measures?
A: No. There is no compelling experimental data supporting a logarithmic linking function between word probability and RTs.