Comprehenders Rationally Adapt Semantic Predictions to the Statistics of the Environment: a Bayesian Model of Trial-level N400 Amplitudes

Nathaniel Delaney-Busch[1], Emily Morgan[1], Ellen Lau[2], Gina Kuperberg[1,3]

[1] Department of Psychology, Tufts University; [2] Department of Linguistics, University of Maryland;
[3] Department of Psychiatry and the Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School

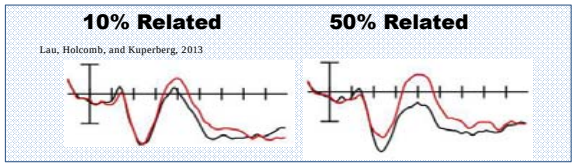
Introduction

Semantic priming:

- When semantic information is activated prior to bottom-up input (i.e. predicted), the semantic processing for the incoming word is typically facilitated, e.g. "salt" can facilitate "pepper".
- N400 component is sensitive to semantic priming: Facilitated processing → Smaller N400
- May reflect lexico-semantic surprisal (where smaller amplitudes indicate that more information was correctly predicted ahead of time, with low prediction error, see Frank et al, 2015)

Adaptation:

- The semantic priming effect is larger in experimental environments with a higher versus a lower proportion of semantically associated trials, e.g. Lau et al. (2013, 2014):
- **More** related prime-target pairs → **Stronger** predictions → **Larger** N400 effects; vs.
- **Fewer** related prime-target pairs → **Weaker** predictions → **Smaller** N400 effects



This provides evidence that we adapt to the statistical structure of the environment such that semantic prediction is enhanced in environments with higher predictive validity.

But there are an infinite number of ways this adaptation can occur.

Q: Can the **shape of the learning curve** be explained by **rational adaptation specifically**?

Methods: Reanalyze data from Lau et al., predicting trial-by-trial adaptation using a Bayesian learning model of probabilistic rational adaptation that conceives of the N400 as reflecting word surprisal (i.e. the unpredicted information content of a word).

Frank S., Otten L., Galli G., Vigliocco G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain Lang.*
Lau, E.F., Weber, K., Gramfort, A., Hamalainen, M., & Kuperberg, G. (2014). **Spatiotemporal signatures of lexical-semantic prediction.** *Cerebral Cortex.*
Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). **Dissociating N400 effects of prediction from association in single-word contexts.** *J Cogn Neurosci.*

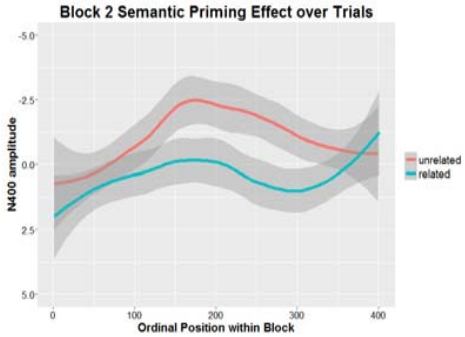
Design and Example Stimuli

10% Related (Block 1)	50% Related (Block 2)
breeze – late	tardy– late
syrup – lightning	**syrup – lightning**
blubber – groom	blubber – groom
instruct – scratch	itch – scratch
clorox – tarantula	clorox – tarantula
rye – west	rye – west
keg – beer	**keg – beer**
touchdown – bank	touchdown – bank
house – cracker	saltine – cracker
tapioca – test	quiz – test

- Task targets (identify animal words) in red
- Counterbalanced experimental items in **bold**
- Time-locked ERP measures underlined

Modeling and Results

1. The N400 semantic priming effect shows adaptation



Amplitudes estimated using nonparametric local regression.

- Semantic priming effect grows over the course of block 2, as people learn the new proportion. Adaptation effect.
- End-of-experiment convergence. Fatigue effect?

2. Building a rational adapter model of the N400

We set out to build a theoretically motivated model of how a rational adapter might incrementally update their predictions, and how those predictions might relate to the N400 ERP component (e.g. through prediction error).

a) *A rational adapter would enter block 2 with some prior expectation of the probability of seeing a related (vs. unrelated) item, but then would incrementally update this probabilistic prediction as evidence accumulates.*

- This parameterizes the "rational adaptation" to a single value: expected probability of getting a related trial, "**λ**".
- **λ** is expressed as a beta-binomial distribution over the probability of receiving a related trial. At the beginning of block 2, we assume that participants hold a prior belief that the block 1 proportion (of 10% related) will continue, with some degree of certainty. Each incoming trial then slowly changes this belief (which changes more slowly as certainty increases), eventually asymptoting to the "correct" belief of 50% related.
- The mean **λ** at any given trial is exactly specified by the number of related and unrelated trials observed so far, plus the prior. We assumed mean = 0.1, and a concentration = 50 for the prior distribution.

b) *Related and unrelated trials have different avenues for prediction.*

- Forward Association Strength might govern lexico-semantic prediction for related items, expressed as a proportion, i.e. P(item | prime)
- Word frequency might govern lexical predictions for unrelated items, expressed as proportion of total corpus.
- Given the prime, the expected probability of any given word is a mixture of these two prediction methods.

$$P(word) = \lambda(FAS) + (1- \lambda)(Frequency)$$

c) *Surprisal, a measure of prediction error, expresses how word probability might relate to the N400. Lower surprisal → smaller N400 amplitude.*

$$Surprisal = -log2(P(word))$$

All modeling uses a prior over λ of Beta(4,45), i.e. mean = 0.1, and a concentration = 50. The N400 amplitudes was specified by a time window of 300-500ms over the average of three centro-parietal channels. Data was used from block 2 (50% related).

3. "Rational Adapter" Word Surprisal predicts N400 Amplitudes

N400 ~ (1 + surprisal | Subject) + (0 + surprisal | Item) + surprisal

- Surprisal: β = -1.04, t = -5.25, p < 0.001

4. "Rational Adapter" Word Surprisal provides explanatory power above and beyond trial type

N400 ~ (1 + surprisal | Subject) + (1 + surprisal | Item) + surprisal + trialtype

- Surprisal: β = -2.01, t = -2.76, p = 0.006

5. The "Rational Adapter" Word Surprisal effect is not attributable merely to frequency and FAS

N400 ~ (1 + surprisal | Subject) + (1 + surprisal | Item)
+ surprisal + trialtype + frequency + FAS

- Surprisal: β = -2.10, t = -2.11, p = 0.036

Summary

A loess local regression of the trial-by-trial N400 amplitudes (**1**) indicated that participants rapidly adapted to the change in relatedness proportion from block 1 (10%) to block 2 (50%) as participants learned to predict more strongly.

We modeled one account of how this adaptation could have occurred (**2**). Specifically, we built a "rational adapter" model of word surprisal (2a), under the theory that the semantic processing underlying the N400 component reflects prediction error (2c), i.e. semantic content not predicted ahead of time. We set the prior for block 2 to the learned proportion in block 1 (mean λ = 0.1), but allowed for the incremental learning about the new environment as the experiment unfolded.

This "rational adapter" word surprisal significantly predicted N400 amplitudes (3), and did so even after controlling for the overall main effect of related vs. unrelated trials (4). This N400 surprisal effect is also not attributable to merely frequency and forward association strength alone (**5**).

This suggests that rational adaptation could account for within-block adaptation to environment statistics. Future work will explore alternative learning accounts, a possible "engagement/fatigue" account of the end-of-block convergence, and a fitted estimate of the learning rate.

Acknowledgements

This project was funded by NIMH-R01-MH071635 to GRK and the Sidney J. Baer Trust to GRK. Thank you to Florian Jaeger, Heather Urry, Eddie Wlotko, and Meredith Brown for input.