

Prediction

Brian Dillon
Shota Momma

University of Massachusetts, Amherst
Department of Linguistics

3/7/2022

Why care?



To explain the robust and fast nature of comprehension

Why care?

- Processing involves Prediction,
- Prediction is Production,
- Prediction leads to Prediction error,
- Prediction error creates Priming,
- Priming is imPlicit learning,
- imPlicit learning is the mechanism for acquisition/adaptation of Processing, Prediction and Production, and
- Production provides the input for training Processing.

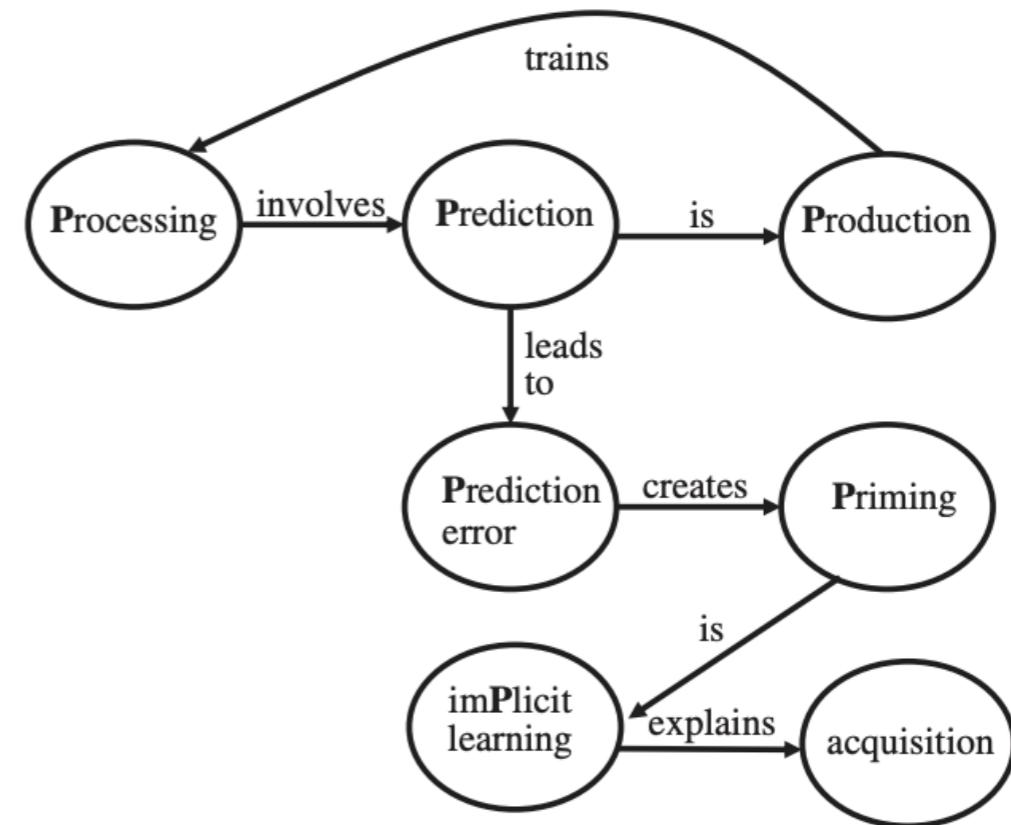
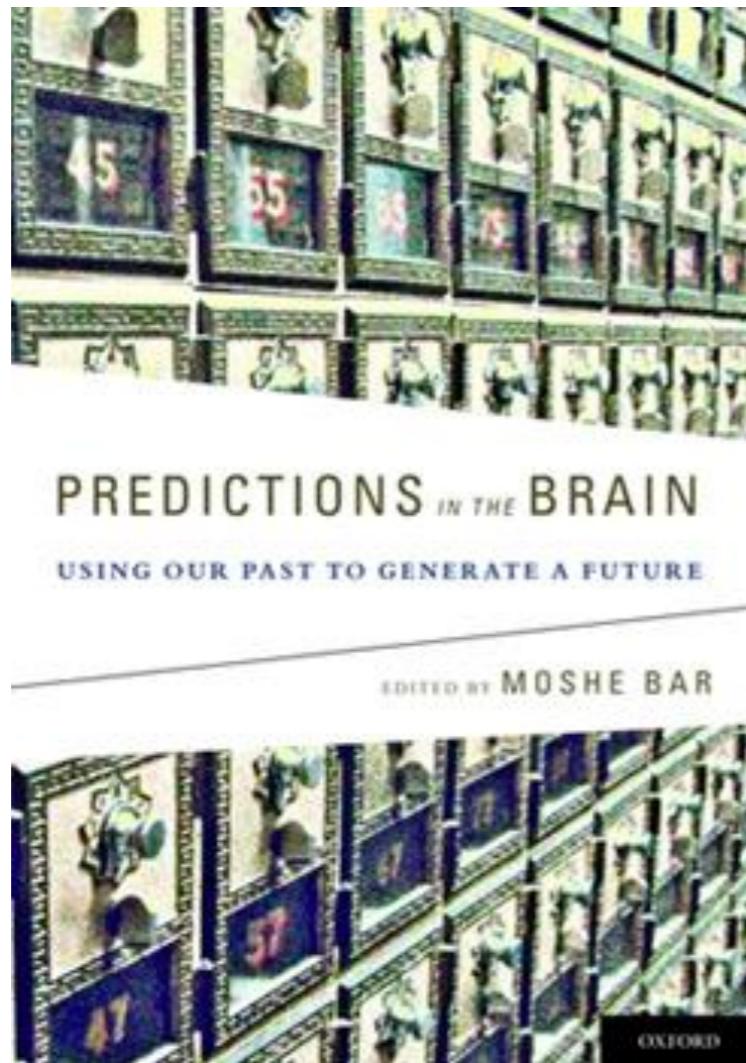


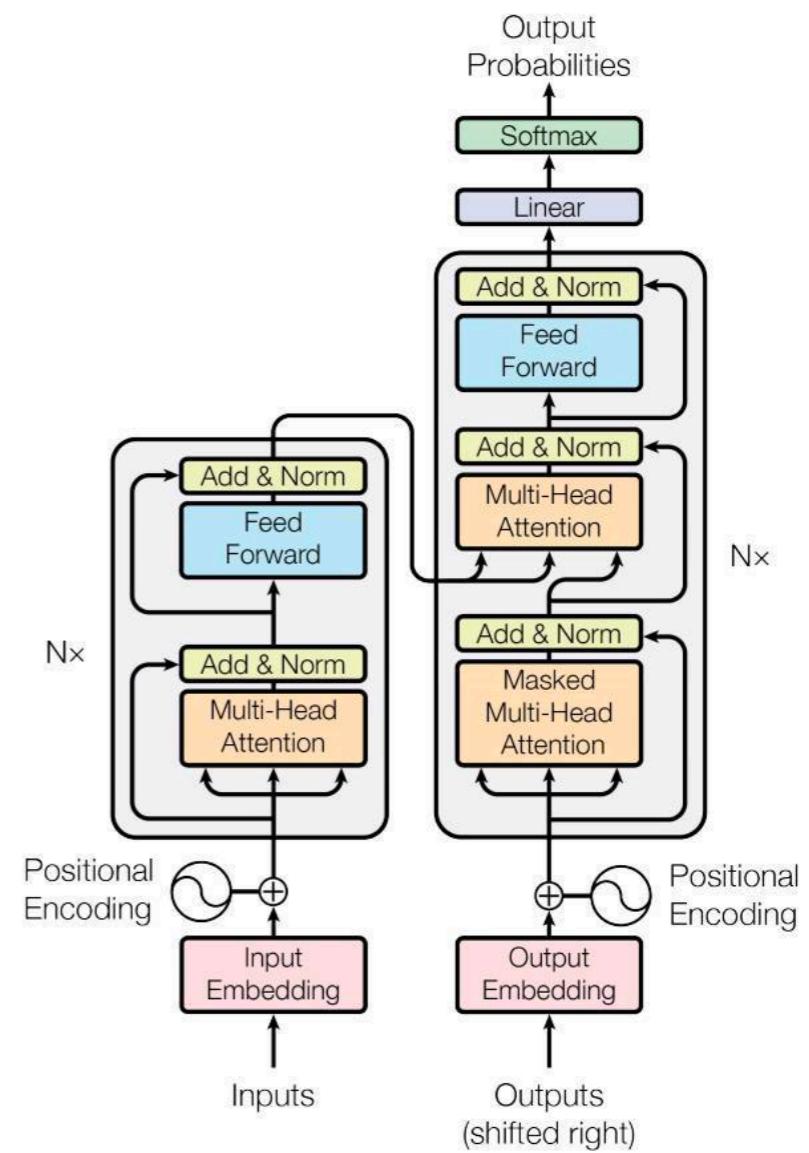
Figure 1. The P-chain framework for psycholinguistics.

An integrated theory of comprehension, production and acquisition (with prediction as the central component)?

Why care?



Prediction as a unifying framework of cognition



Language models trained on prediction tasks

Why care?

What do RNN Language Models Learn about Filler–Gap Dependencies?

Ethan Wilcox, Roger Levy, Takashi Morita, Richard Futrell

Mechanisms for handling nested dependencies in neural-network language models and humans[☆]

Yair Lakretz^{a,*}, Dieuwke Hupkes^c, Alessandra Vergallito^b, Marco Marelli^b, Marco Baroni^{c,d,1}, Stanislas Dehaene^{a,e,1}

Neural Language Models Capture Some, But Not All, Agreement Attraction Effects

Suhas Arehalli
Johns Hopkins University

Tal Linzen
Johns Hopkins University

Accounting for Agreement Phenomena in Sentence Comprehension with Transformer Language Models: Effects of Similarity-based Interference on Surprisal and Attention

Soo Hyun Ryu, Richard L. Lewis

Assessing the Ability of LSTMs to Learn Syntax-Sensitive Dependencies

Tal Linzen^{1,2} **Emmanuel Dupoux¹**
LSCP¹ & IJN², CNRS,
EHESS and ENS, PSL Research University
{tal.linzen,
emmanuel.dupoux}@ens.fr

Yoav Goldberg
Computer Science Department
Bar Ilan University
yoav.goldberg@gmail.com

Why care?

- Processing involves Prediction,
- Prediction is Production,
- Prediction leads to Prediction error,
- Prediction error creates Priming,
- Priming is imPlicit learning,
- imPlicit learning is the mechanism for acquisition/adaptation of Processing, Prediction and Production, and
- Production provides the input for training Processing.

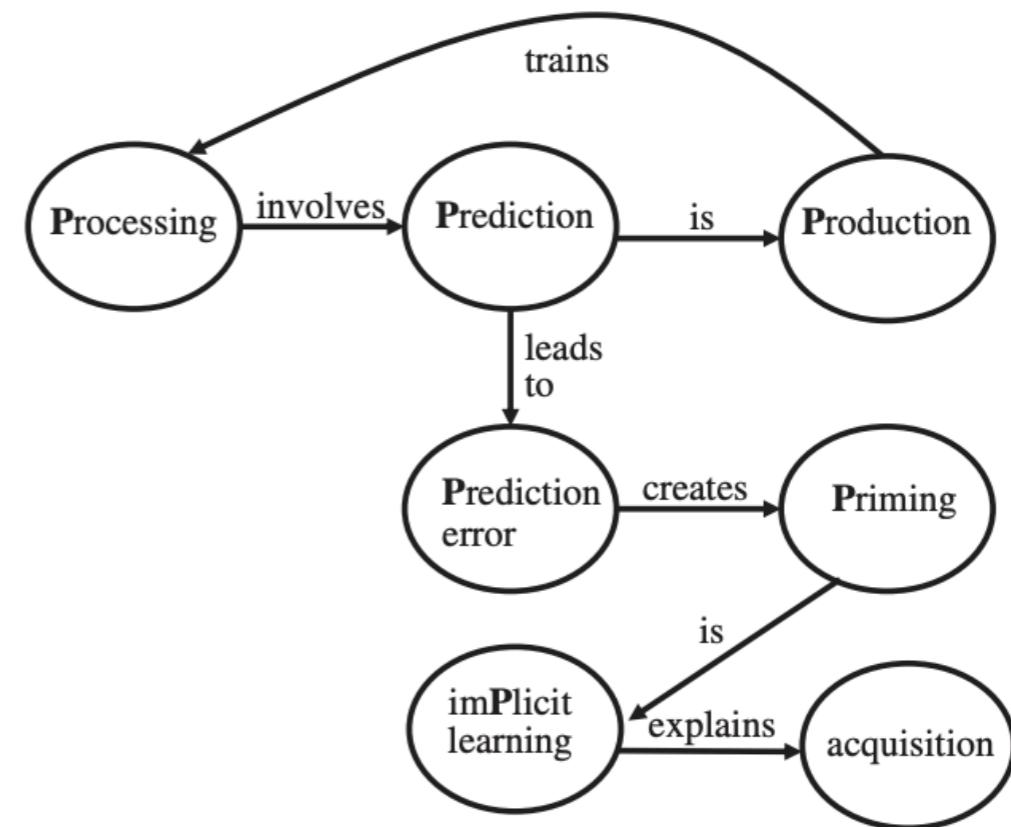


Figure 1. The P-chain framework for psycholinguistics.

An integrated theory of comprehension, production and acquisition (with prediction as the central component)?

Implicit learning

The girl gave...



Predict DO or PD structure

If you've predicted PD/DO structure, you get "surprised" to see a DO/PD structure. This 'surprise' triggers learning (prediction-error based learning)

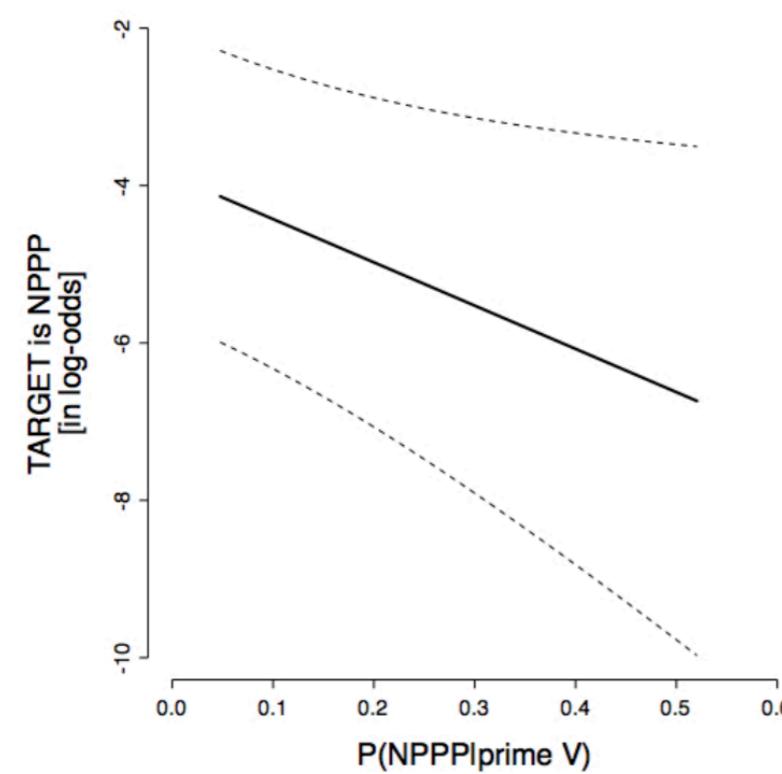
- > 'tune' your system of 'generating' structure used for both prediction and production
- > next time, you are more likely to use that structure in production (syntactic priming) or to predict it (adaptation effect in comprehension)

Inverse frequency: Verb bias

Switch board corpus:

Prime = preceding dative relative to target dative

Dative



Active/passive

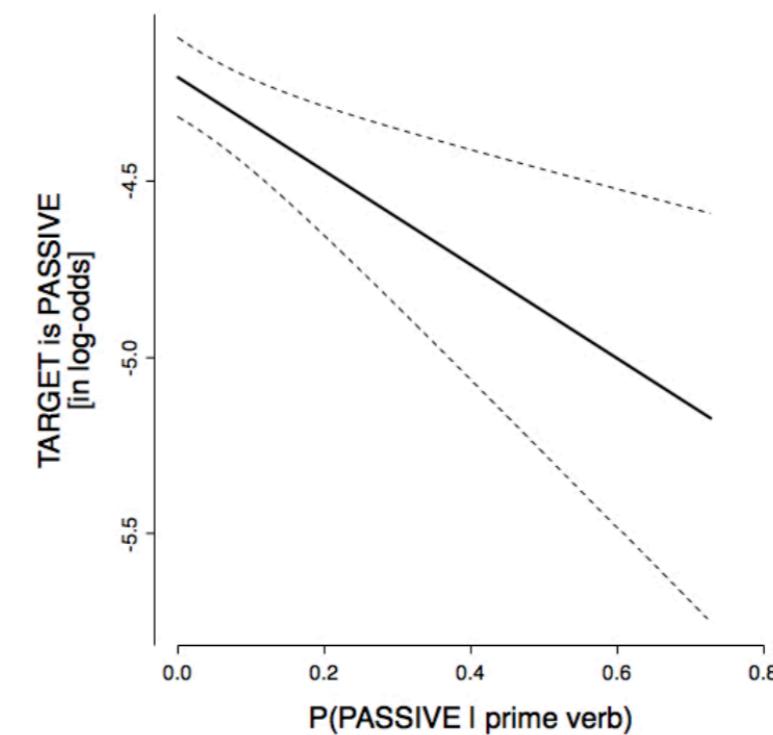
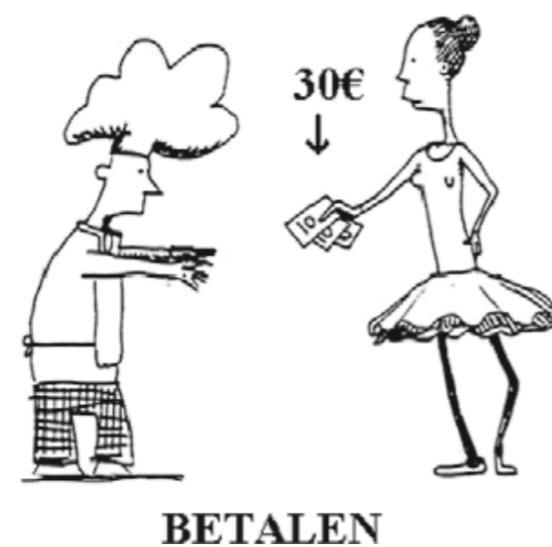


Figure 1: Prime surprisal (based on prime verb's NPPP bias)

Figure 2: Prime surprisal (based on prime verb's passive bias)

Inverse frequency: Verb bias



(a) De kok biedt de dokter een pistool aan
[The cook offers the doctor a gun]

DO-dative

(b) De kok biedt een pistool aan aan de dokter
[The cook offers a gun to the doctor]

PO-dative

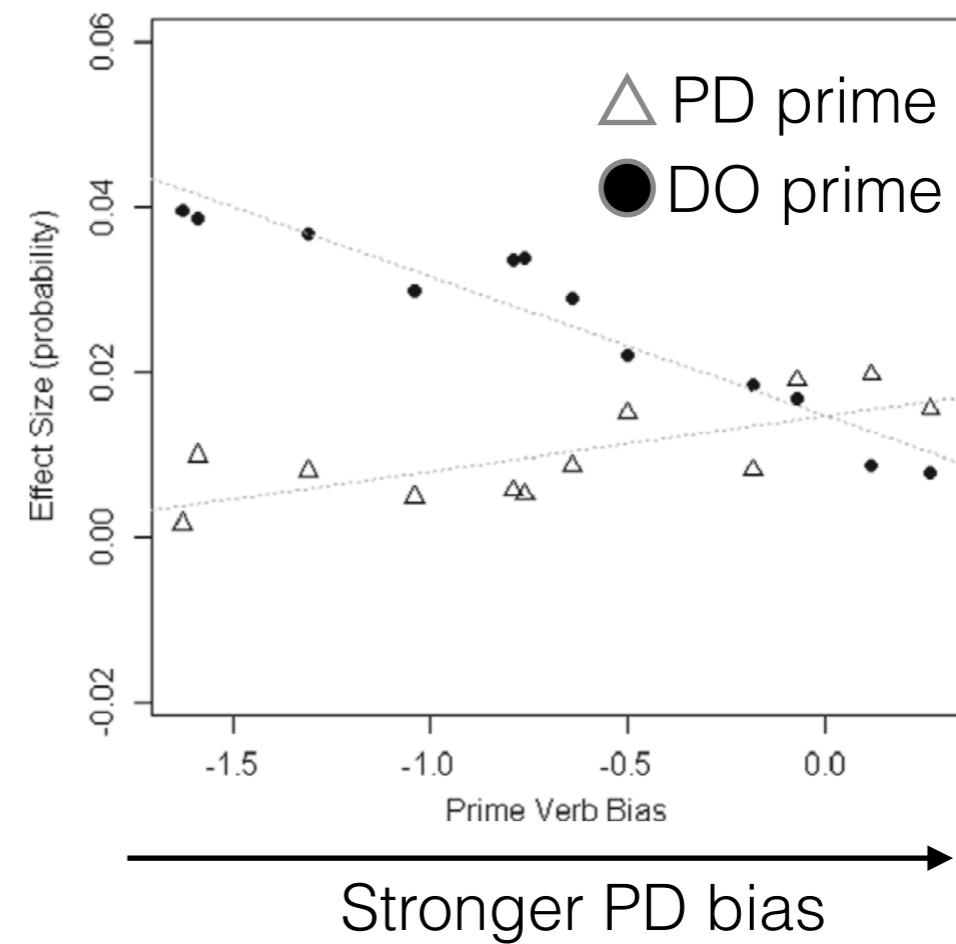
(c) De kok scheldt de dokter uit
[The cook scolds the doctor]

Baseline

Table 1

Raw numbers of DO- and PO-responses and percentage of PO responses in all conditions of the priming experiment.

	DO-responses	PO-responses	% PO
Baseline	211	668	76.0
DO-prime	311	572	64.8
PO-prime	173	705	80.3



Note: Effects of DO- priming (circles) and PO-priming (triangles) for the different prime verbs, ordered according to their alternation bias (strongest bias towards PO on left hand side).

Fig. 3. Effects of DO- and PO-priming (vs. baseline) for each prime verb.

Long-term priming

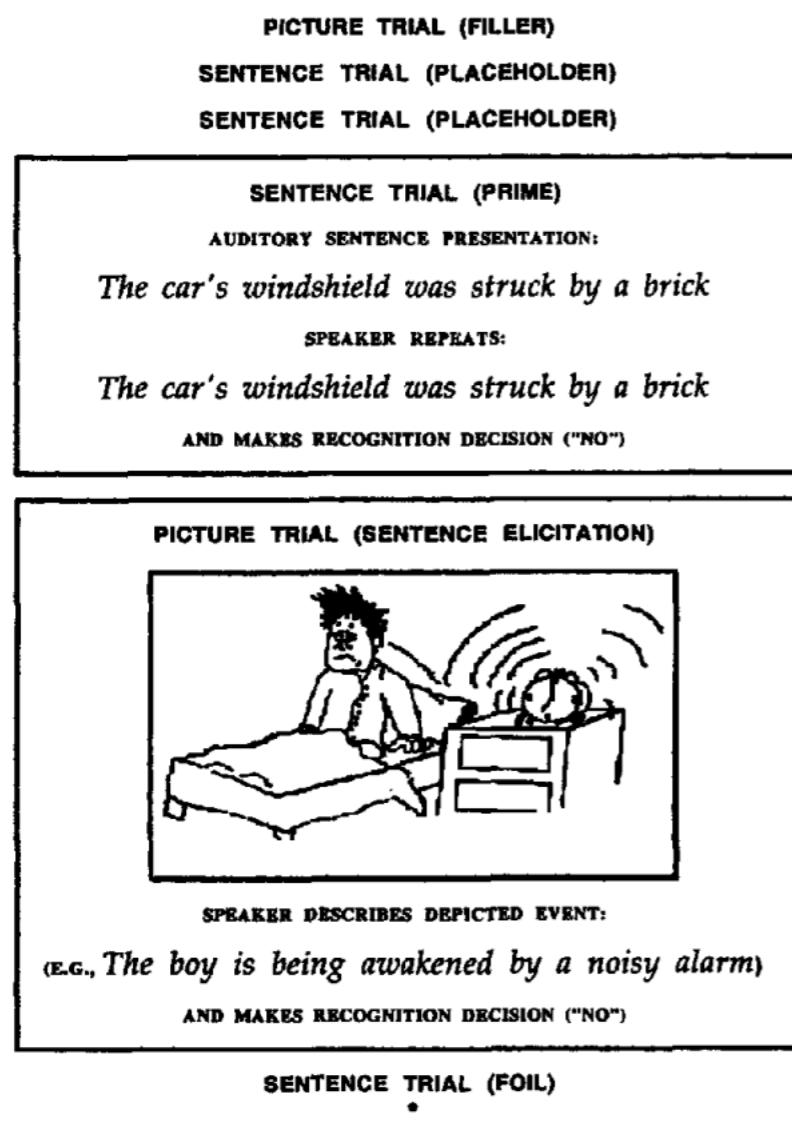


Figure 1. Sequence of events on a sample Lag-0 structural priming trial.

Table 1
Proportions of Target Syntactic Structures Produced After Target and Alternative Primes Over Three Lags (Experiment 1)

Utterance and priming form	Lag			
	0	1	2	M
Dative				
Prepositional dative (target)	.48	.54	.61	.54
Double-object dative (alternative)	.43	.42	.47	.44
Transitive				
Passive transitive (target)	.55	.58	.54	.56
Active transitive (alternative)	.49	.54	.54	.52

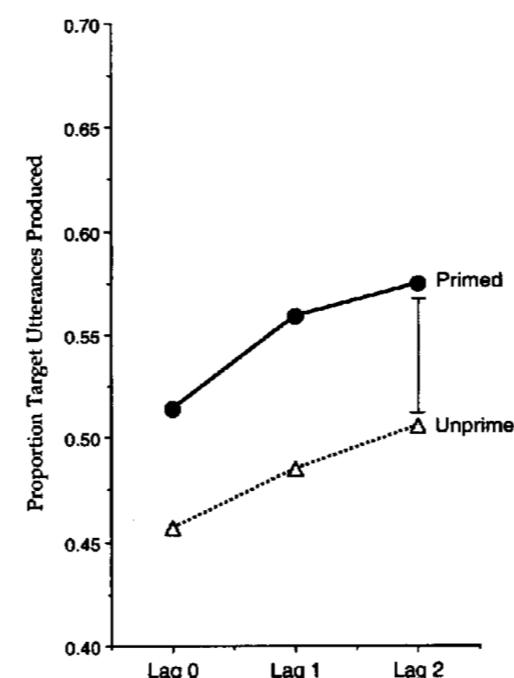


Figure 2. Structural priming across 0, 1, or 2 intervening sentences (lag) after structure-matching (primed) or structure-mismatching (unprimed) priming sentences. The error bar represents the 95% confidence interval for a pairwise planned comparison (.062), as calculated from the error term for the interaction between prime type and lag in the analysis of variance by participants.

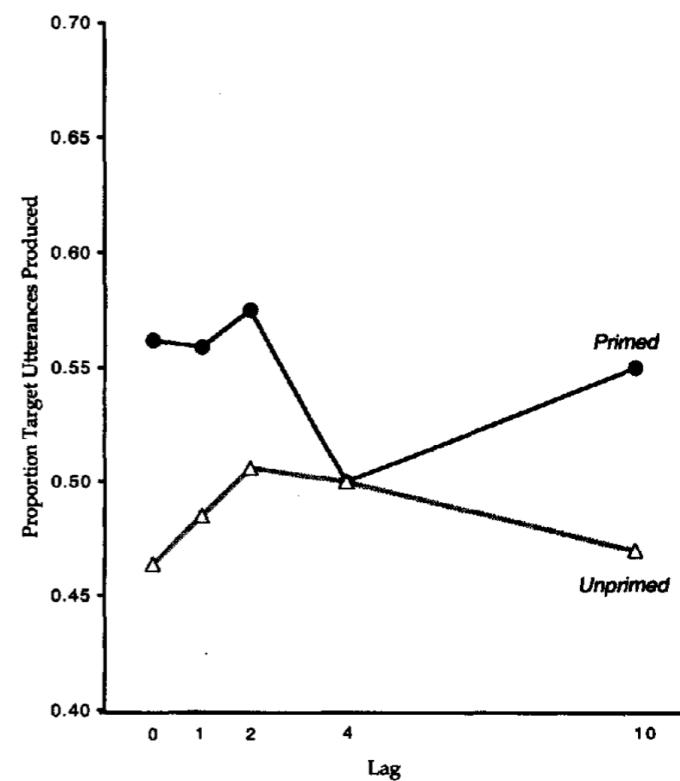
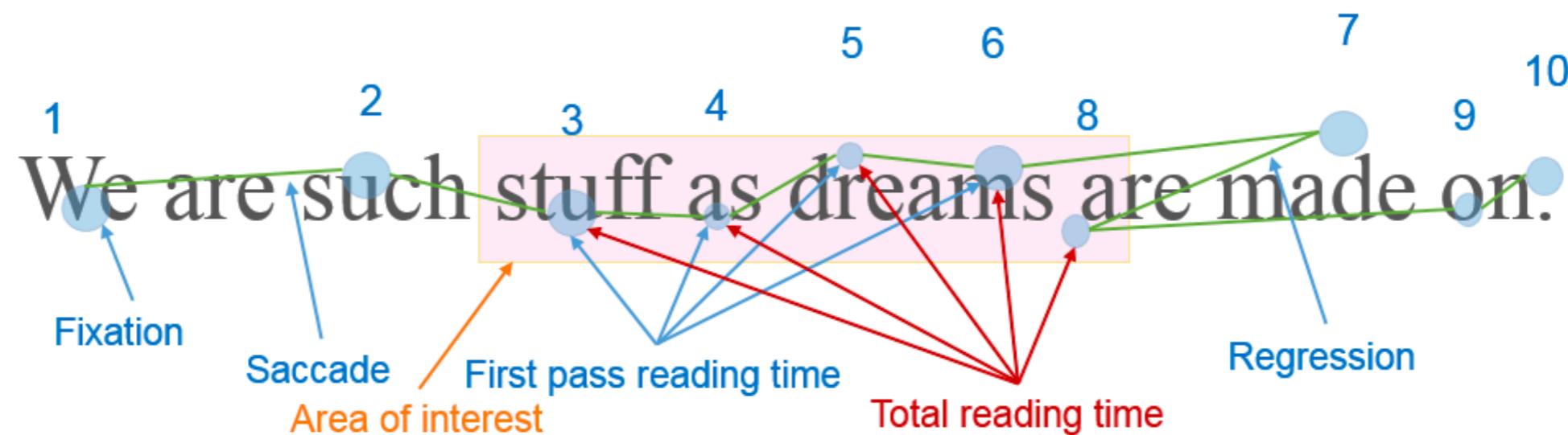


Figure 4. Overview of structural priming at lags spanning 0 to 10 intervening sentences.

Predictability effect

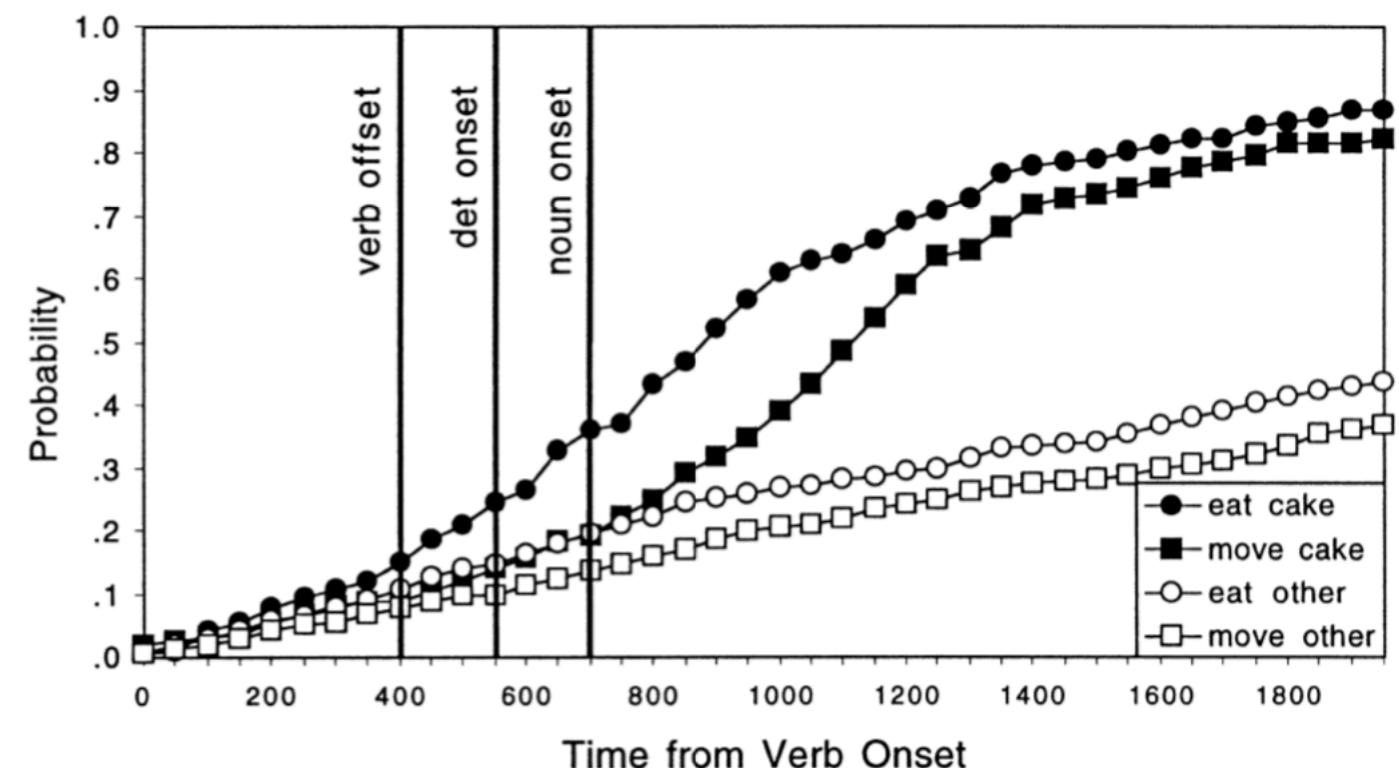
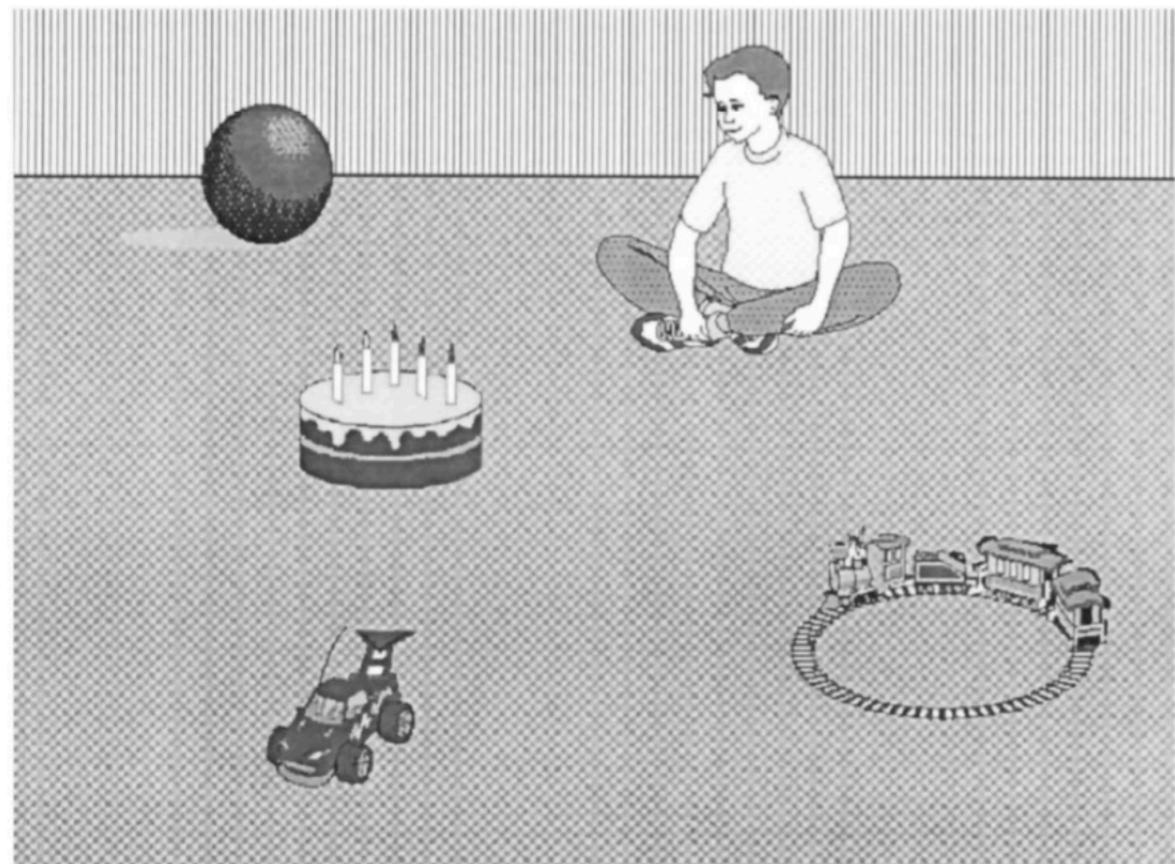


Britney always drives way too recklessly. She just got into an *accident* that wrecked her car.

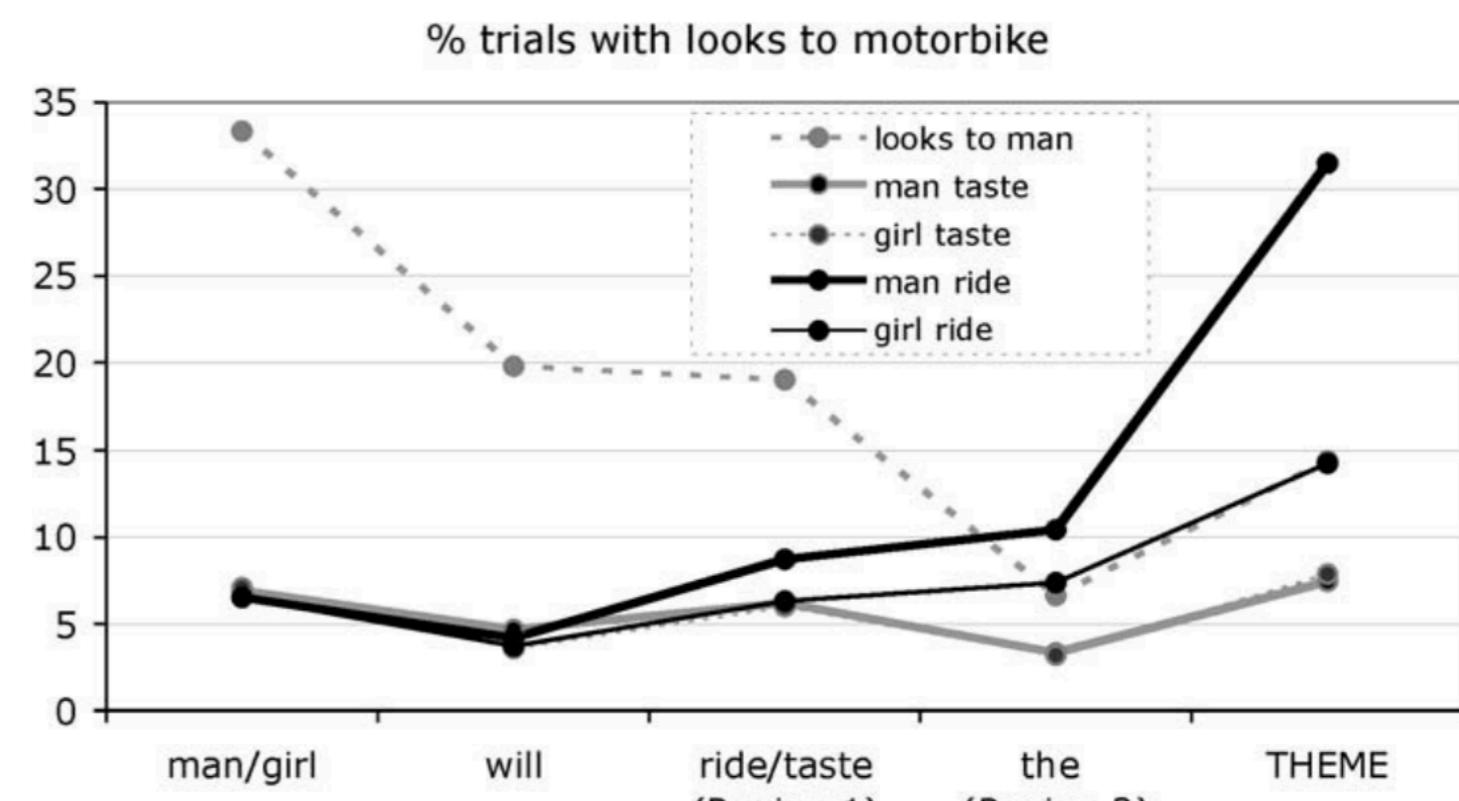
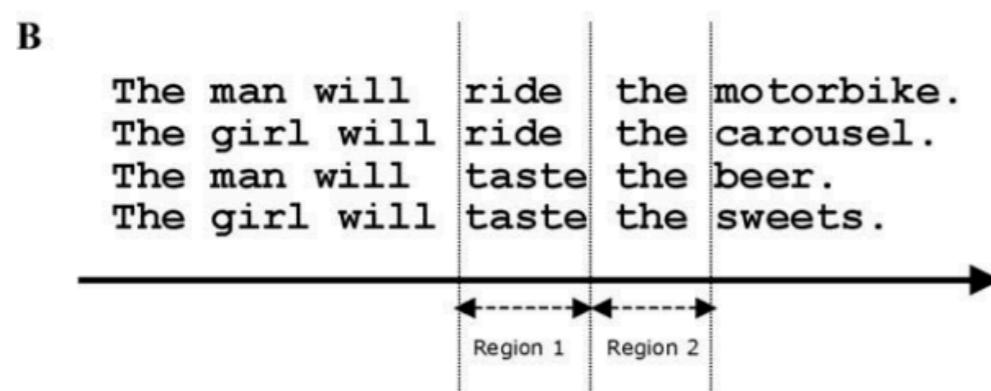
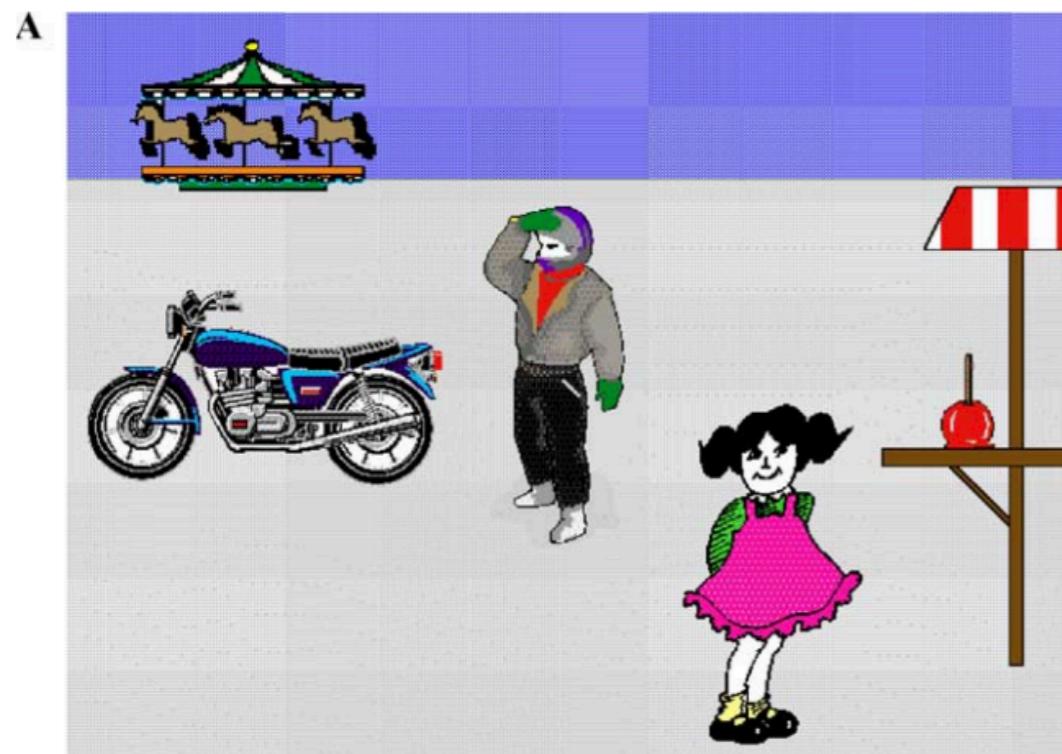
Britney is in a really bad mood today. She just got into an *accident* that wrecked her car.

More predictable ->
higher skipping rate,
shorter first fixation

Evidence for prediction



Evidence for prediction



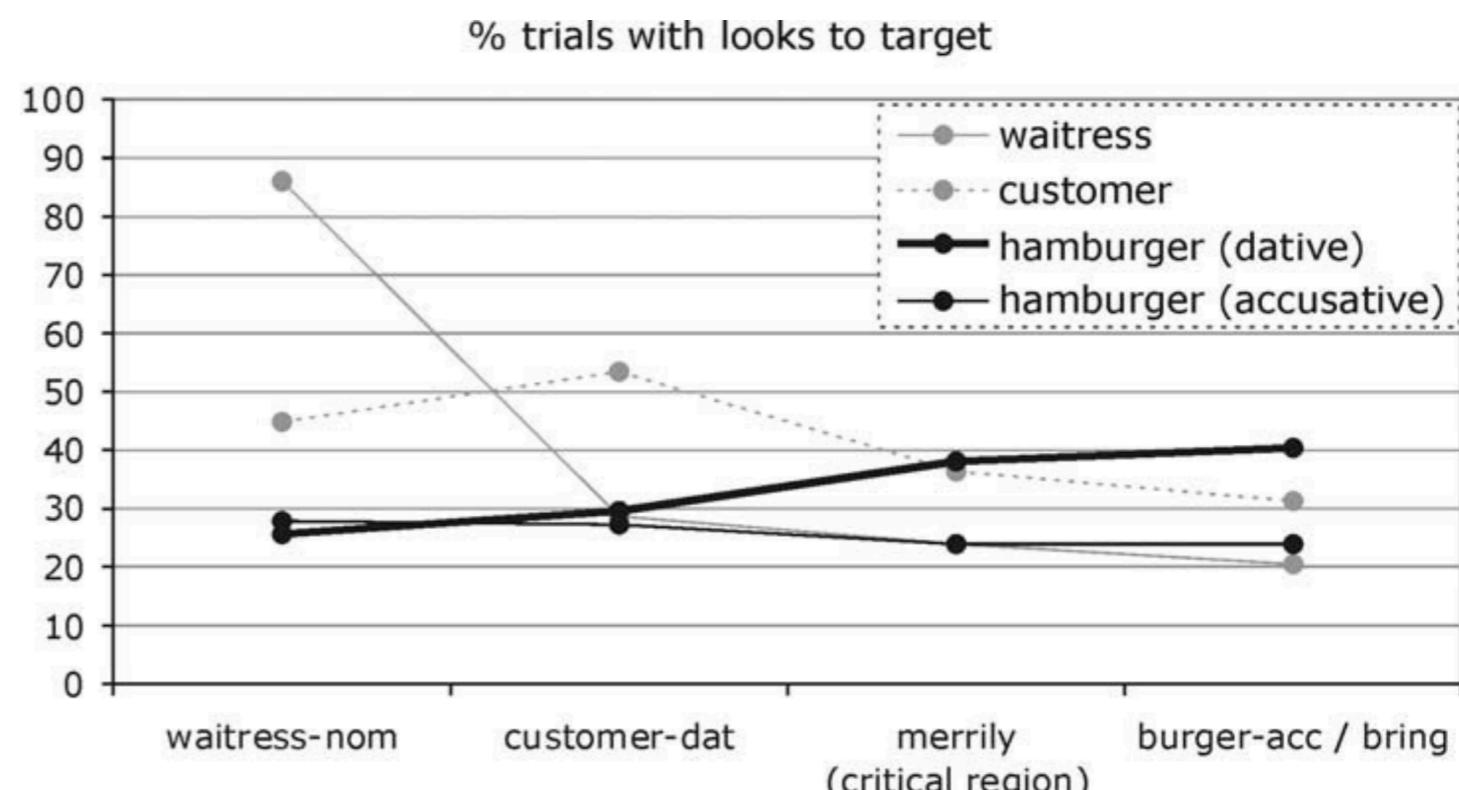
Evidence for prediction

A



B

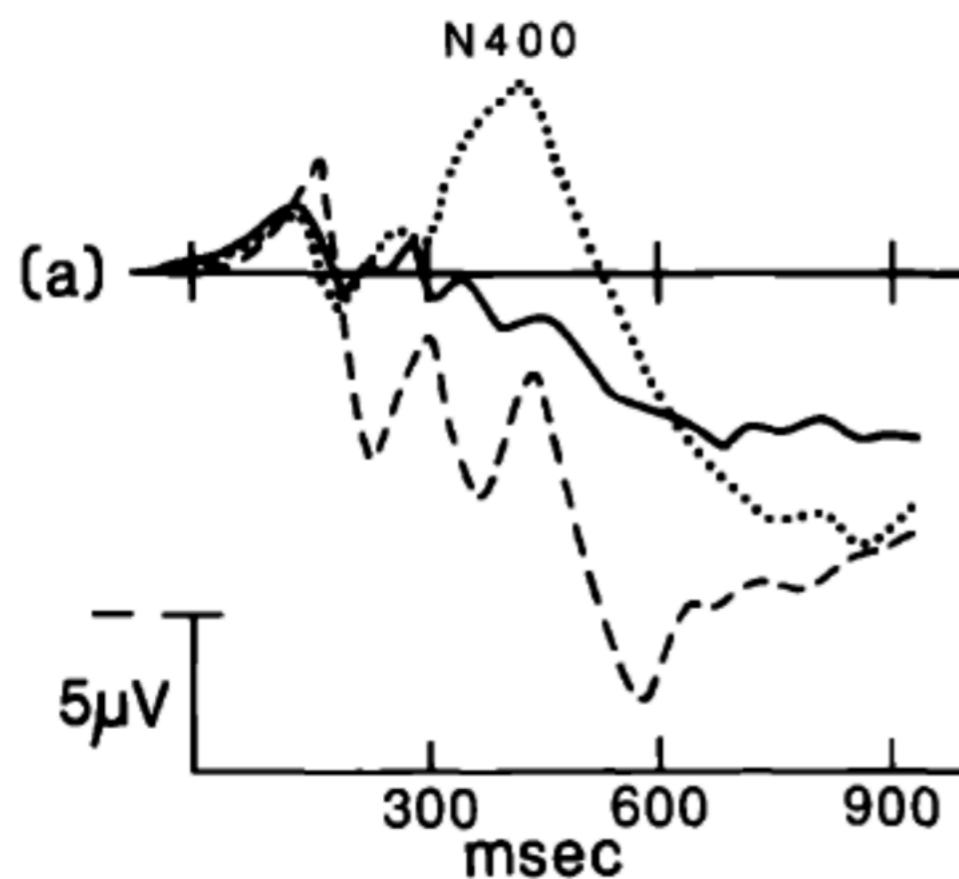
waitress-nom	customer-dat	merrily	hamburger-acc bring.
ウェイトレスが	客に	楽し気に	ハンバーガーを 運ぶ。
←→			
ウェイトレスが	客を	楽し気に	からかう。
waitress-nom	customer-acc	merrily	tease.



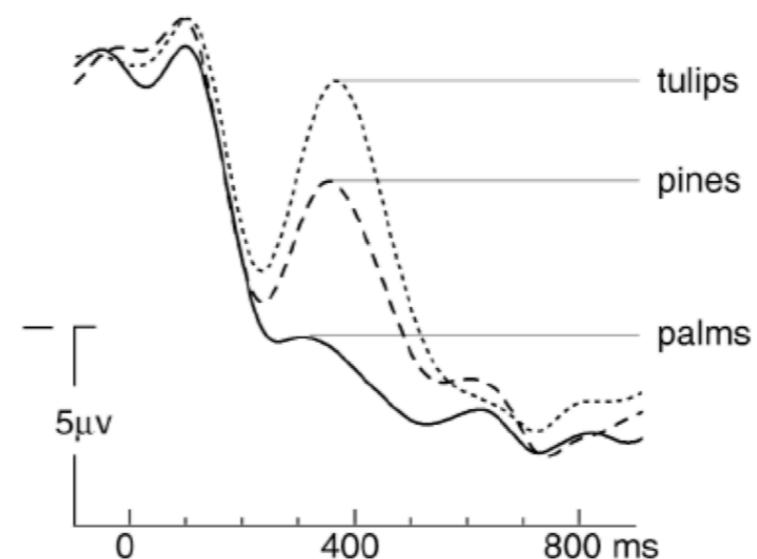
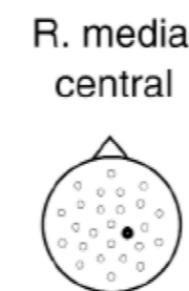
N400 and prediction

ERPs

- He spread the warm bread with butter.
- - - - He spread the warm bread with BUTTER.
- He spread the warm bread with socks.



“They wanted to make the hotel look more like a tropical resort.
So along the driveway they planted rows of _____.”

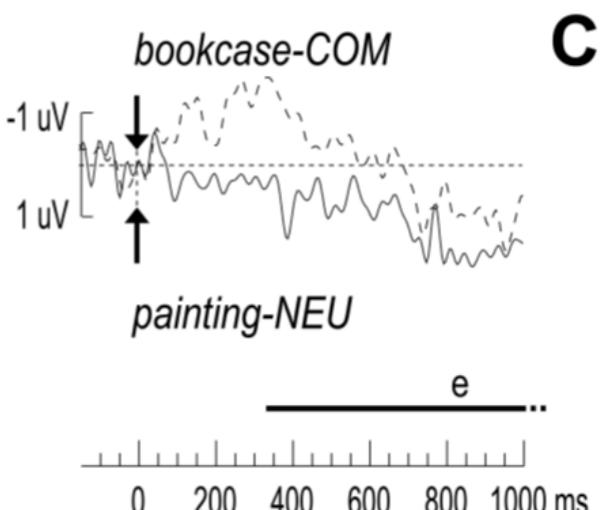
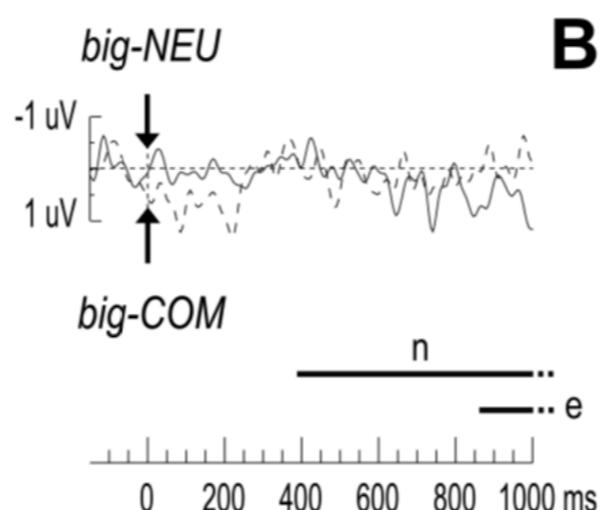


Kutas & Hillyard (1983)

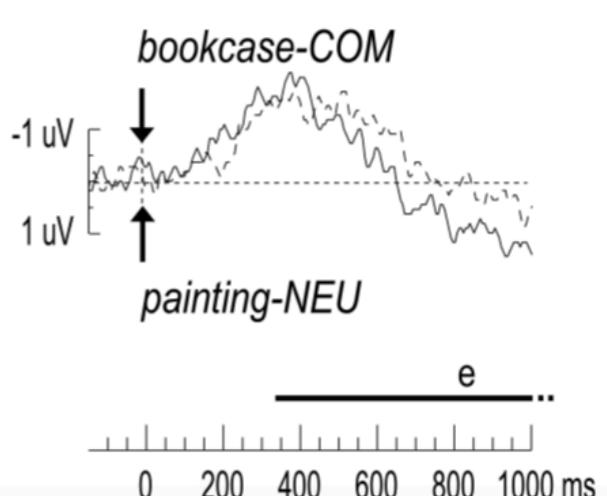
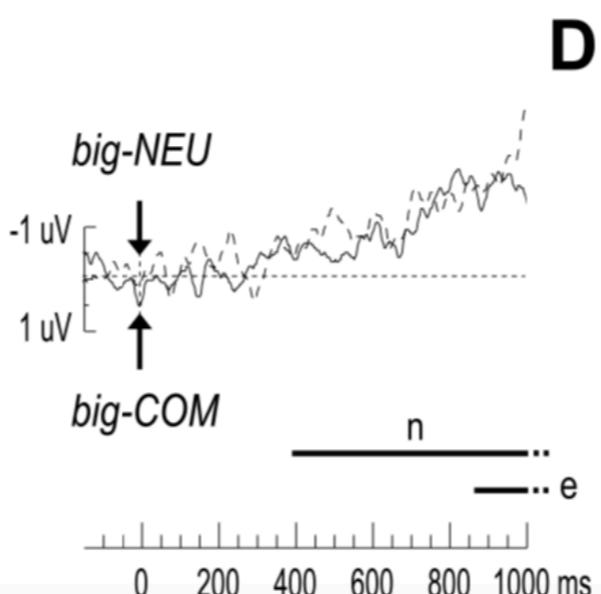
Federmeier & Kutas (1999)

Evidence for prediction

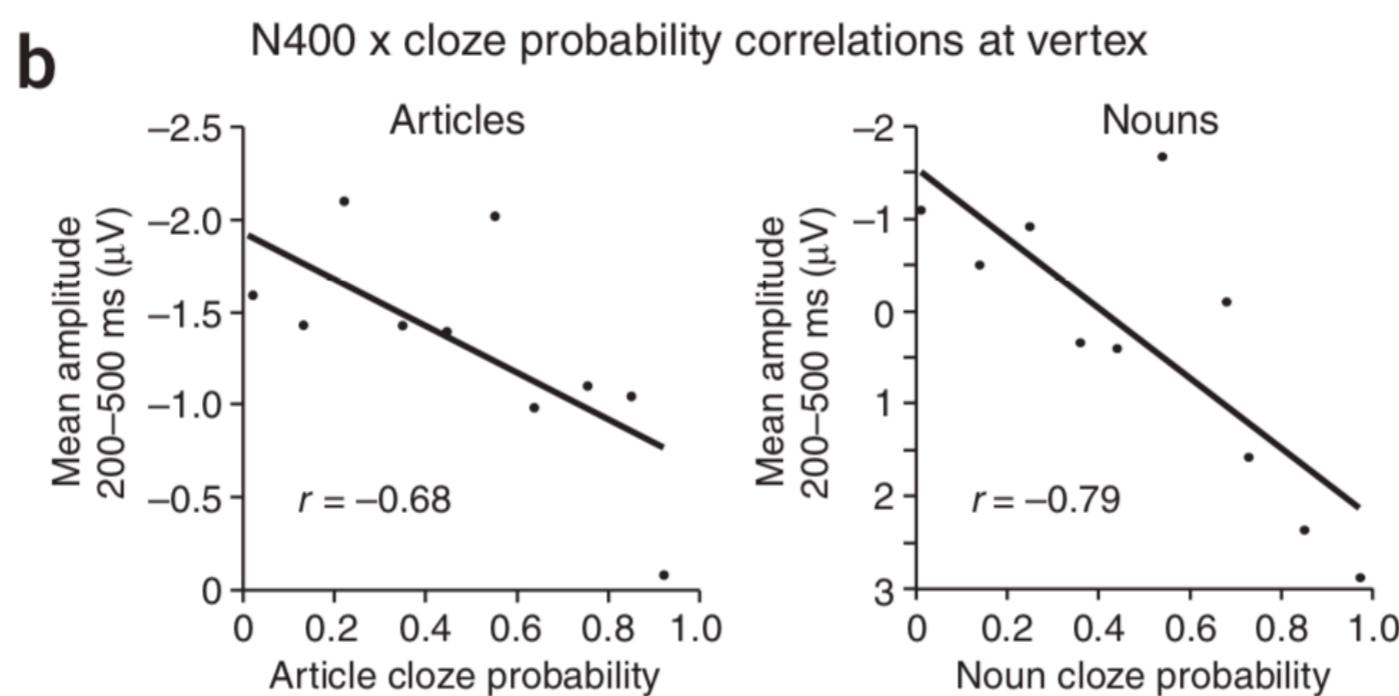
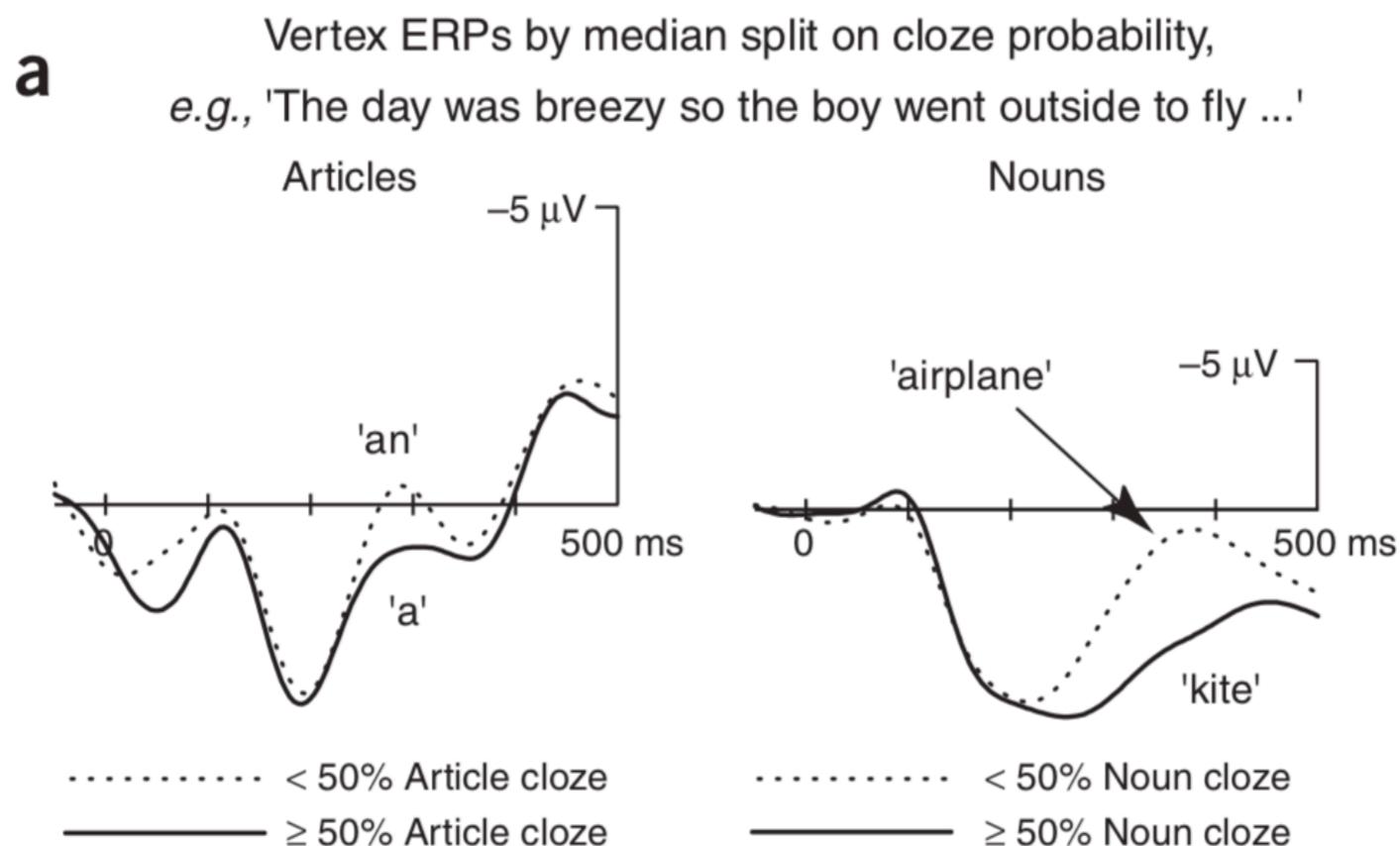
*The burglar had no trouble whatsoever
to locate the secret family safe.
Of course, it was situated behind a...*



[no predictive discourse context]
Of course, it was situated behind a...



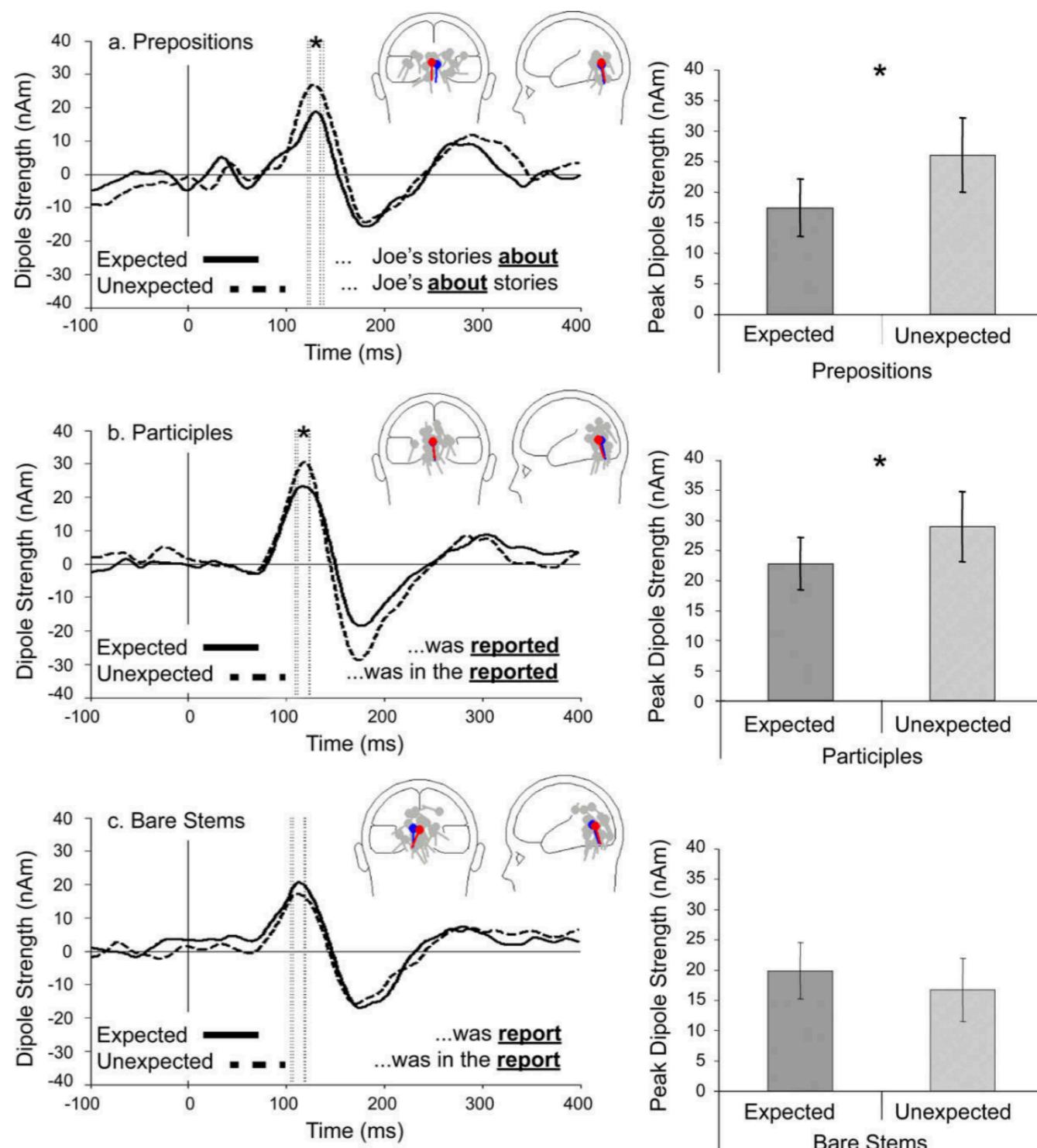
Evidence for prediction



Evidence for prediction

Condition	Expected	
	Example sentence	Cloze-P
Preposition	The boys heard Joe's stories about Africa	0.44
Participle	The discovery was reported	0.29
Bare stem	The discovery was in the report	0.91

Condition	Unexpected	
	Example sentence	Cloze-P
	The boys heard Joe's about stories Africa	0
	The discovery was in the reported	0.002
	The discovery was report	0.002



Evidence for prediction



NEUROSCIENCE



Large-scale replication study reveals a limit on probabilistic prediction in language comprehension



Mante S Nieuwland , Stephen Politzer-Ahles, Evelien Heyselaar, Katrien Segaert, Emily Darley, Nina Kazanina, Sarah Von Grebmer Zu Wolfsturn, Federica Bartolozzi, Vita Kogan [see all »](#)

Max Planck Institute for Psycholinguistics, Netherlands; University of Edinburgh, United Kingdom; The Hong Kong Polytechnic University, Hong Kong; University of Oxford, United Kingdom; University of Birmingham, United Kingdom; University of Bristol, United Kingdom; University of Glasgow, United Kingdom; University of Kent, United Kingdom; University College London, United Kingdom [see all »](#)

How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects

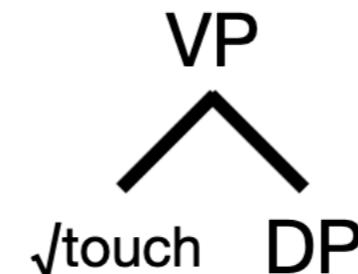
Aine Ito, Andrea E. Martin & Mante S. Nieuwland

Prediction across levels

Semantic

[[touch]]

Syntactic



Lexical

/tʌtʃ/

Phonetic

