'Negation-blind' N400 effect disappears when lexical priming is controlled

Daiki Asami, Chao Han, Jacob Burger, Deanna Dunlop, Yue Lu, Effah Yahya M Morad, Chenyue Zhao, and Arild Hestvik

University of Delaware

Introduction: 2-step models of negation[1] says that when people comprehend a negative sentence, they initially compute the truth value of the to-be-negated proposition and then compute the effect of negation at a subsequent stage, where negation occurs in a sentential scope position. Evidence for this model has come from ERP studies, where the “truth-value N400” inverts in negated sentences[2]. False affirmative sentences as in (1b) compared to true affirmatives as in (1a) trigger a “truth-value N400” at the sentence-final object. Negated sentences exhibit the inverse pattern: here, it is true negatives (2a) that trigger N400 compared to false negatives (2b). This can be explained if the N400 is elicited at the initial stage before negation is computed, such the N400 is elicited by “tree” in (2a) for “a robin is a tree”—thereby the term “negation blind N400.” However, prior ERP studies contained a priming confound: sentences like (1a/2b) contain a class-inclusion relation and consequently priming between the subject and object via lexical spreading activation, whereas the critical sentences (1b/2a) have no such priming relation. Because the N400 is known to be an inverse index of priming[3], this presents a confound because negation-blind N400 can also be explained as list priming in (2b) compared to (2a), triggering the inverse N400 priming index. The goal of the current study was therefore to examine whether the negation-blind N400 persists when this priming confound is removed. If the truth-value N400 in prior experiments was unrelated to the inverse priming N400, the N400 blindness pattern should persist. If previous results were wholly due to the “inverse of priming N400”, negation-blind N400 should disappear.

Methods: 30 undergraduate students volunteered in an experiment with a 2x2 within-subject design, with truth value (true vs. false) crossed with sentence form (affirmative vs. negative), using 40 trials in each cell, and 40 filler items=200 trials (Table 2). To control the priming relation between subject and object, we used sentences that contained simple true/false size comparisons between familiar physical objects, differing in animacy and semantic category. The sentences were visually presented in four centered 175ms chunks (e.g., A tiger/is/bigger than/a book) with an 800ms ISI. Participants made timed truth value judgment after each sentence. EEG was time-locked to the onset of the object; epochs were-200ms to 1000ms.

Results: We found higher accuracy for false sentences compared to true sentences in negative forms (83% vs. 76%, p=.002), but not in affirmative forms (92% vs. 91%, p=.263). Participants were significantly slower in judging false sentences than true sentences (p=.024), as well as in judging negative sentences compared to affirmatives (p<.001). This behavioral pattern replicated previous studies[2,4], except the magnitude of the RTs was greater in our experiment. Turning to the ERP results, we identified the truth-value ERP effect by decomposing the two false-true difference waves with temporo-spatial Principal Component Analysis, which is a data-driven approach to determining temporal and spatial regions of interest. [5] No N400 component reflecting truth value differences emerged, but a late left-anterior negativity (LAN) was observed. Using the PCA/ICA factor loadings we constructed a 504-680 ms time window and a left-anterior electrode cluster, and calculated the mean voltage for this time/space region as the dependent variable (Fig 2). ANOVA revealed a main effect of negation ($F(1, 29) = 5.52$, $p = .026$), a main effect of truth value ($F(1, 29) = 6.12$, $p = .020$), and an interaction between the truth value and negation ($F(1, 29) = 5.53$, $p = .026$), such that the “truth-value LAN" amplitude was greater for negatives than for affirmatives (see Fig. 1-2).

Discussion: When controlling for priming between subject and object, no N400 index to truth value and consequently no negation-blind N400 was observed. Instead, we observed a LAN component with a main effect for truth value, which matches a prior finding identifying the left inferior prefrontal cortex as being related to truth-value computation. [6] Notably, no “negation blind” inversion was observed for this LAN. The theoretical implication is that once priming is controlled, there is no “negation-blind” N400 evidence for 2-step models of non-incremental processing of negation. Instead, LAN emerges as the critical ERP index of truth-value processing, whether “not” is present or absent.

(1)	a. A robin is a bird. (TA)	b. A robin is a tree. (FA) ← N400
(2)	a. A robin is not a tree. (TN) ← N400	b. A robin is not a bird. (FN)

Table 1: Sample stimuli of a two-step model of negation processing in [6].

	Affirmative	Negative
True	A tiger is bigger than a guitar. (TA)	A mouse is not bigger than a guitar. (TN)
False	A tiger is smaller than a guitar. (FA)	A mouse is not smaller than a guitar. (FN)

Table 2: Example stimuli in the four cells.

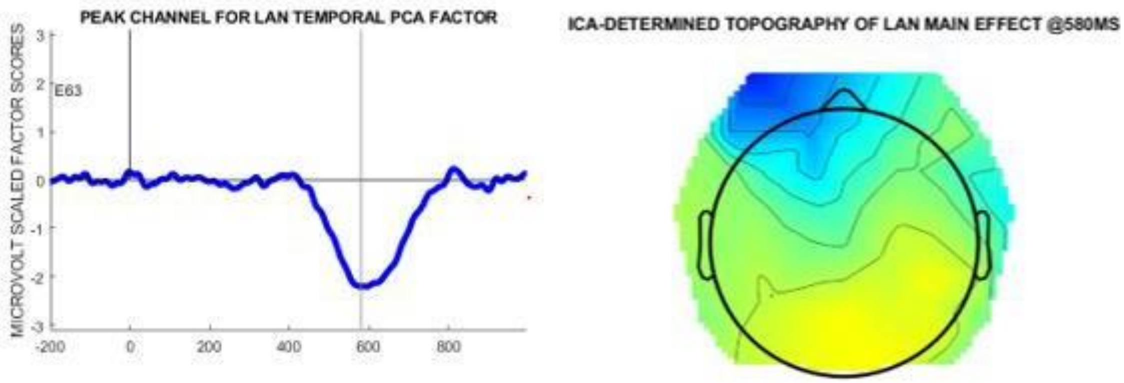

Figure 1: Time course and topography of main effect of truth, temporo-spatial LAN factor

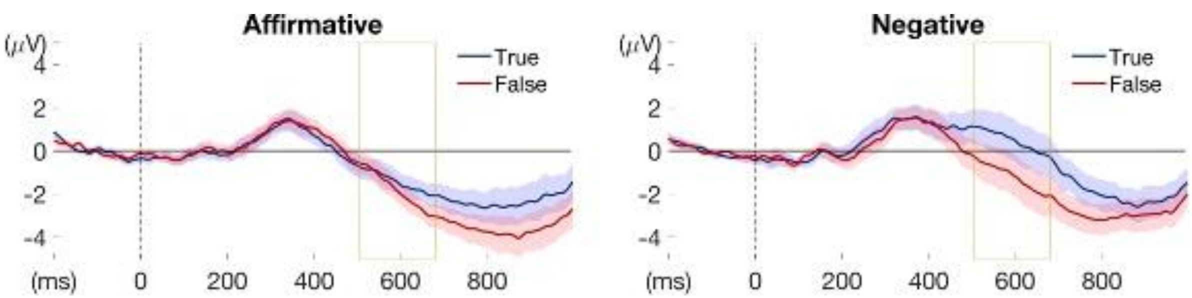

Figure 2: Mean voltage waveforms for ICA-determined LAN electrode cluster with 84% CI

1. Clark, H., & Chase, W. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*.
2. Fischler, I., Bloom, P.A., Childers, D.G., Roucos, S.E. and Perry, N.W., Jr. (1983), Brain Potentials Related to Stages of Sentence Verification. *Psychophysiology*.
3. Holcomb, P. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language*.
4. Palaz, B., Rhodes, R., & Hestvik, A. (2020). Informative use of “not” is n400-blind. *Psychophysiology*.
5. Dien J.(2010). The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data. *J Neurosci Methods*.
6. Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*.