'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. Prior ERP studies of truth-value and negation computation (Fischler et al. 1983; Palaz et al. 2020; among others) have argued for a classic two-step account of negation processing (Clark & Chase 1972). Their evidence comes from an interaction between sentence form (presence or absence of "not") and truth value in an N400 effect. Specifically, false affirmative (FA, 1b) sentences yielded a larger N400 compared to true affirmatives (TA, 1a)—the truth-sensitive N400 effect (Hagoort et al. 2004)—whereas the inverse was observed for negative sentences: true negative (TN, 2a) sentences caused a larger N400 than the false negative (FN, 2b) as if the N400 was 'blind' to negation and reflected only the truth value of the internal positive proposition (i.e., 'a robin is a tree').

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

In the two-step model of negation processing, when comprehending a negative sentence such as (2a), the meaning of the to-be-negated proposition, which is false, is computed in the first step, and then negation is applied to flip its truth value in the second step (Clark & Chase 1972). Under the assumption that the N400 is elicited during the first step, the negation-blind N400 effect follows.

However, the prior studies arguably contained a confound: the stimuli that generated the larger N400 contain no lexical priming relation between subject and object, whereas the control stimuli contained the lexical priming relation between subject and object. The N400 effect is known as an inverse index of priming: when a semantically related word pair (e.g., 'doctor' and 'nurse') is compared to an unrelated pair (e.g., 'car and 'nurse'), the primed word generates a reduced N400 compared to the unprimed word (Holcomb 1988). The subject and object in (1a/2b) are semantically related, but the subject and object in (2a/1b) where the truth-sensitive N400 effect was observed, are semantically unrelated, thereby providing an independent source of the observed N400 effect. The goal of the current study was to examine if the negation-blind N400 pattern persists even when this priming confound was removed. To this end, we conducted an ERP experiment with comparative constructions where the subject and object were unrelated in terms of animacy as well as semantic category (Table. 1). We predicted that if the previously observed interaction was unrelated to priming, the negation-blind N400 patten would replicate; if not, it would disappear. Our result was consistent with the second prediction.

Methods. 30 people participated in our ERP experiment with the 2x2 within-subject design, manipulating truth value (true vs. false) and sentence form (affirmative vs. negative). Each condition had 40 sentences. We also used 40 fillers (160 + 40 = 200 sentences).

	Sentence form	
Truth value	Affirmative	Negative
True	A tiger is bigger than a guitar.	A mouse is not bigger than a guitar.
False	A tiger is smaller than a guitar.	A mouse is not smaller than a guitar.

Table 1: Sample stimuli in the four conditions (two truth values × two sentence forms).

Each stimulus was visually presented in four chunks (e.g., A tiger / is / bigger than / a book) with 175ms duration and 800ms ISI. Participants made a speeded truth value judgment via button press at the object chunk. EEG was timelocked to the object with -200-to-1000ms epochs.

Behavioral Results. For accuracy, we observed main effects of truth value and sentence form: the false condition had higher accuracy than the true condition (87% vs. 83%, $F(1,29)=11$, $p<0.01$)

and the affirmative condition had higher accuracy than the negative condition (91.1 vs. 79.3%, $F(1,29)=93.5$, $p<.0001$). Truth value interacted with sentence form such that FN accuracy was higher than TN accuracy (82.8 vs. 75.7%, $F(1,29)=8.14$, $p<.01$). RT analysis revealed main effects of truth value and sentence form: the true condition was judged faster than the false condition (1342 vs 1390 ms, $F(1,29)=4.79$, $p<0.05$) and the negative condition took shorter to judge than the affirmative condition (1190 vs. 1542 ms, $F(1,29)=132.4$, $p<0.0001$). The results mirrored prior findings (Clark & Chase 1972; Carpenter & Just., 1976; Fischler et al. 1983; Palaz et al. 2020).

ERP Results. In looking for the N400, we used a data-driven sequential PCA technique (Dien, 2010, 2012) to identify the temporal and spatial components of the brain response to truth value. We used the two difference waves, false-minus-true for affirmatives and negatives, as inputs. Inspection of the resulting temporal factors corresponding to the difference between true and false sentences revealed no N400, but instead a late left-anterior negativity. We next used the factor loadings to constrain selection of a time window of 504-680ms and a left-anterior electrode cluster and calculated the mean voltage per cell and subject as dependent measures for a 2x2 repeated measure ANOVA. This revealed a main effect of negation ($F(1, 29) = 5.52$, $p = .026$), main effect of truth value ($F(1, 29) = 6.12$, $p = .020$), and interaction between the truth value and negation ($F(1, 29) = 5.53$, $p = .026$), driven by a greater difference for negatives. Figure 1 shows the mean waveforms for the regionalized channels:

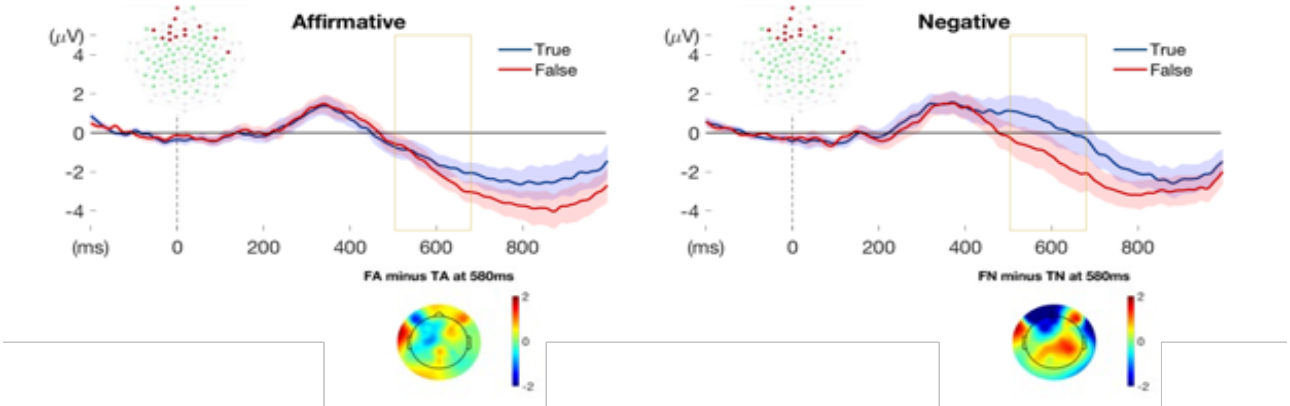

Fig. 1: Waveforms with 84% CI for main effect of truth; difference wave topoplot at peak latency

Discussion. The main finding is that when controlling for a lexical priming relation between subject and object, no N400 index to truth value was observed, and consequently no negation-blind N400. This suggests that previous N400 evidence for the two-step negation processing was wholly due to the lexical priming confound. Despite the lack of the N400 effect, we did observe a statistically significant brain response modulation by truth value, which suggests that the true ERP index of truth value computation is not the N400 but the LAN. This matches the findings of Hagoort et al., (2004) who identified the left inferior prefrontal cortex as being related to truth-value computation based on world knowledge. We attribute the relative lateness of this LAN to the relatively more difficult judgment task, which is seen by the longer RTs than those in previous studies (e.g., Fischler et al. 1983).

Selected References. Carpenter, P. A. & Just M.A (1975) in *Psychological Review*; Clark, H., & Chase, W. G. (1972) in *Cognitive Psychology*; Dien, J. (2010) in *Journal of Neuroscience Methods*; Dien, J. (2012) in *Developmental Neuropsychology*; Fischler, I. et al. (1983) in *Psychophysiology*; Hagoort, P. et al. (2004) in *Science*; Palaz, B. et al. (2020) in *Psychophysiology*