'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, & Arild Hestvik*

Abstract. Previous ERP studies showed that false affirmative sentences elicited a larger N400 than their true versions, but they found the reverse pattern when the sentences were of negative form as if N400 was blind to negation. This negation-blind N400 pattern arguably constituted evidence for two-step accounts of negation processing: When processing negative sentences, a comprehender first computes an internal proposition and then considers the negation. However, the prior studies were confounded by a lexical priming relation between subject and object. Therefore, it was an open question whether or not the observed ERP pattern really reflected the two-step process. To tackle this question, we conducted an ERP experiment, using size-comparison statements where subjects and objects are semantically unrelated. This design allowed us to remove the priming confound. We predicted that if the previous negation-blind N400 pattern is unrelated to lexical priming, it would be replicated; if not, it would disappear. The result was consistent with the second prediction. This suggests that the previously observed negation-blind N400 pattern does not necessarily constitute evidence for two-step accounts of negation processing.

Keywords. sentence processing; negation; N400; two-step accounts; lexical priming

1. Introduction. Processing of negative sentences (e.g., *a whale is not a fish.*) has attracted much attention in psycholinguistics, as evidenced by many reviews (Kaup et al. 2007, Tian & Breheny 2019, Christensen 2020, Kaup & Dudschig 2020, Papeo & de Vega 2020, Dudschig et al. 2021). These reviews showcase various theories of negation processing. Among them, two-step accounts (Clark & Chase 1972, Carpenter & Just 1975, Kaup et al. 2006, 2007, Kaup & Dudschig 2007, Palaz et al. 2020) postulate that negation processing involves two steps, as schematized in (1): a comprehender first computes a to-be-negated internal proposition and then considers negation.[1]

(1) **Step 1:** ⟦A whale is a fish⟧ **Step 2:** ⟦Not⟧ (⟦A whale is a fish⟧)

Two-step accounts received neurological support from event-related potential (ERP) research, that is, a 'negation-blind' pattern of N400 results (Fischler et al. 1983, Kounios & Holcomb 1992,

*We thank Xu Qing for the conception of the experiment and Neemias Silva de Souza Filho for helpful feedback. We are also grateful to the audience at ELM3 for their comments. Authors: Daiki Asami, University of Delaware (daiasami@udel.edu), Chao Han, University of Toronto Scarborough (chao.han@utoronto.ca), Jacob Burger, University of Delaware (jburger@udel.edu), Deanna Dunlop, Gallaudet University (deanna.dunlop@gallaudet.edu), Yue Lu, University of Delaware (sunnylu@udel.edu), Effah Yahya M Morad, University of Delaware (emorad@udel.edu), Chenyue Zhao, Temple University (chenyue.zhao@temple.edu), Arild Hestvik, University of Delaware (hestvik@udel.edu). This study was conducted while all the authors were affiliated with the University of Delaware.

[1] We are agnostic about exact representations of the proposition to which *not* is applied. For this reason, we use double square brackets, following the formal semantic tradition. This does not undermine the main claims made by the previous studies.

©2025 Daiki Asami, Chao Han, Jacob Burger, Deanna Dunlop, Yue Lu et al.
Published by the LSA with permission of the author(s) under a CC BY license.

Dudschig et al. 2019, Haase et al. 2019, Palaz et al. 2020).[2] An N400 is a negative ERP component with a peak at approximately 400 ms after the onset of every content word; its amplitude is sensitive to a wide range of semantic factors (see Kutas & Federmeier (2011) for a review). Among them, content words that cause world knowledge violations increase the N400 amplitude (Hagoort et al. 2004, Hald et al. 2007, Metzner et al. 2015, Nieuwland et al. 2019). Consistent with this, prior studies found that false affirmatives in (2b) elicited a larger N400 relative to their true counterparts in (2a) (Fischler et al. 1983, Kounios & Holcomb 1992, Dudschig et al. 2019, Palaz et al. 2020).[3]

(2) a. A robin is a bird. (true affirmative) b. A robin is a tree. (false affirmative)

(Fischler et al. 1983: 402)

Crucially, the reverse pattern was observed in the negative sentences: the true sentences in (3a) elicited a larger N400 than their false versions in (3b), as if the N400 was only sensitive to the to-be-negated proposition and 'blind' to negation.

(3) a. A robin is not a tree. (true negative) b. A robin is not a bird. (false negative)

(Fischler et al. 1983: 402)

This N400 pattern falls into place under two-step accounts: when comprehending the true negative sentence (3a), only the false internal proposition (i.e., *A robin is a tree*) is computed in the N400 time window; consequently, it elicits a larger N400 than the false negative sentence (3b), in which the core proposition (i.e., *A robin is a bird*) is true.

However, as acknowledged by the original negation-blind N400 study (Fischler et al. 1983: 407–408) and later brought up by others (Nieuwland & Kuperberg 2008, Wiswede et al. 2013, Haase et al. 2019, He et al. 2022), the previous ERP studies were confounded by a lexical priming relation between subject and object, which is known to be inversely correlated with N400 amplitude (Kutas & Hillyard 1984, Bentin et al. 1985, Rugg 1985). For instance, the target word *nurse* leads to a reduced N400 when it follows the semantically related word *doctor* compared to when it follows a semantically unrelated word *cat*. This is because the prime word *doctor* not only activates the representation of itself but also pre-activates related words or concepts within the lexicon; consequently, pre-activated words require less cognitive costs than non-preactivated words, resulting in a smaller N400 amplitude.

Returning to our previous discussion on ERP studies, the true affirmative sentences exhibited the semantic relatedness between subject and object (e.g., *robin/bird* in (2a)) while their false counterparts did not (e.g., *robin/tree* in (2b)). The same contrast was also true of the comparison of the false and true negatives (e.g., *robin/bird* and *robin/tree* in (3b) and (3a), respectively). As a result, the sentences with the related pair of words (i.e., (2a) and (3b)) may attenuate the N400, relative to those with the unrelated pair of words (i.e., (2b) and (3a)). If this is correct, we may interpret the previously observed negation-blind N400 effect as a reflection of the lexical priming.

To directly test this alternative interpretation, Haase et al. (2019) conducted an ERP experiment using sentences such as (4).[4] They intended to control the semantic relatedness by increasing the

[2]Contrary to these studies, Nieuwland & Kuperberg (2008) found a 'negation-sensitive' N400 pattern such that false sentences elicited a larger N400 in both affirmative and negative sentences. See Section 4 for the relevant discussion.

[3]Throughout this paper, we underline the critical chunk after which the ERPs are averaged.

[4]The original stimuli were in German but we present only the English translations.

Truth	Sentence form: Affirmative	Sentence form: Negative
True	(a) A tiger is bigger than a guitar.	(b) A mouse is not bigger than a guitar.
False	(c) A tiger is smaller than a guitar.	(d) A mouse is not smaller than a guitar.

Table 1: Example stimuli in the four conditions, crossing the two truth values (true vs. false) and the two sentence forms (affirmative vs. negative).

semantic overlap between subject and object. Their experimental sentences used celebrity names (e.g., *George Clooney*) and profession nouns (e.g., *actor*) for subjects and objects, respectively.

(4) a. George Clooney currently is an actor in the USA. (true affirmative)
b. George Clooney currently is a singer in the USA. (false affirmative)
c. George Clooney is not a singer in the USA. (true negative)
d. George Clooney is not an actor in the USA. (false negative)

(Haase et al. 2019: 1)

They reasoned that the use of two profession nouns from the related field (e.g., *singer* and *actor*) increased the semantic relatedness between subject and object, thereby removing the priming confound. Their results revealed an N400 effect in the comparisons of false and true affirmatives ((4b) vs. (4a)) as well as true and false negatives ((4c) vs. (4d)), although the latter comparison was not statistically significant. Haase et al. (2019) took the non-significant negation-blind N400 effect as partial, if not strong, support for two-step accounts of negation processing. However, their design had an issue because a celebrity name is more related to one occupation than the other. For instance, *George Clooney* is more semantically associated with *actor* than *singer*. Thus, their experiment was still confounded by the lexical priming.

Therefore, it was still open whether the previously observed negation-blind N400 results stemmed from a process assumed by two-step accounts or lexical priming. To tackle this issue, we conducted an ERP experiment using size comparison statements (Table 1) where the subject and object differed from each other in terms of animacy and semantic category, regardless of truth values and sentence forms. This design allowed us to remove the priming confound by eliminating the semantic relatedness, rather than increasing it, unlike Haase et al. (2019).

We made two predictions, as in (5) and (6). The first prediction was that if the lexical priming was weakly associated with the previously observed ERP pattern, we would observe a larger truth-sensitive N400 in a word that makes an affirmative sentence false, compared to the same word that makes it true. Additionally, if the two-step negation processing is psychologically real, we would see a negation-blind N400 effect in the opposite subtraction (5b). The second prediction was that if the lexical priming was the source of the previous negation-blind N400 pattern, the critical words would elicit roughly equal N400 amplitude regardless of truth values and sentence forms; as a result, we would observe no N400 effect in the comparisons of interest (i.e., (6a) and (6b)).

(5) a. false affirmative − true affirmative = N400 effect

b. true negative – false negative = N400 effect

(6) a. false affirmative – true affirmative = No N400 effect
b. true negative – false negative = No N400 effect

2. Methods.

2.1. Participants. 30 undergraduate students (27 females; 3 males) participated and were compensated with extra credit. All were native speakers of English, had normal or corrected to normal vision, and reported no history of neurological diseases or medication. They gave written informed consent before data collection. The study was approved by the IRB of the University of Delaware.

2.2. Design and stimuli. The current experiment had a 2 × 2 within-subjects design. The first factor—truth value—was whether a sentence was true or false (true vs. false). The second factor—sentence form—was whether the sentence included *not* or not (affirmative vs. negative). Crossing the two factors yielded four conditions (Table 1). Experimental stimuli were comparative constructions describing a size comparison with *is (not)* {*bigger/smaller*} *than* as a predicate. We created 40 sentences for each condition (160 sentences in total).

We also created 10 sets of fillers for each condition (40 sentences in total). Half of them used five adjectives whose first letter was *b* (*brighter*, *busier*, *braver*, *broader*, and *bumpier*); the other half used those beginning with *s* (*saltier*, *sweeter*, *scarier*, *sicker*, and *softer*). These adjectives exhibit partial similarity to *bigger* and *smaller* in the experimental items in terms of orthography and phonology. We utilized this similarity to prevent participants from developing semantic satiation (e.g., (*not*) *bigger* and (*not*) *smaller* become less meaningful as a function of repetition) (Jakobovits & Lambert 1962, Smith & Klein 1990) or adopting any heuristic strategy (e.g., mapping *not bigger* into *smaller* without computing the compositional meaning).

Each participant saw all 200 sentences in a single experiment. However, we expected that results would be confounded by the facilitated processing due to the repetition priming if sentences containing the same words (e.g., *tiger*) appeared multiple times within a short interval. To rule out this confound, we distributed 200 sentences (160 experimental sentences and 40 filler sentences) into four lists using a Latin Square design. We then randomly assigned these lists to four individual blocks in each experiment so that the participants never saw more than one item from the same set within each block.

2.3. Procedure. Our experiment employed a speeded truth-value judgment task (Figure 1). To begin each trial, the participant was prompted to press 1 or 5 on the response box. Upon the button press, a fixation element appeared for 1,000–1,400 ms, and then a sentence was presented with a rapid serial visual presentation paradigm. The sentence was divided into four chunks (e.g., a tiger / is (not) / bigger than / a guitar.). Each chunk was presented for 175 ms with an 800 ms interstimulus interval. Upon the presentation of the final chunk, the participant judged whether the sentence was true or false by pressing 1 (= true) or 5 (= false) as quickly and accurately as possible. The allotted response time was 4,000 ms. Reaction time and accuracy of the response were recorded. Upon the truth-value judgment, the feedback element appeared for 1,000 ms. If the participant did not respond in 4,000 ms, the feedback *no response* was displayed and the response was recorded as a missing value. The procedure was repeated 50 times in four blocks (200 trials in total). The

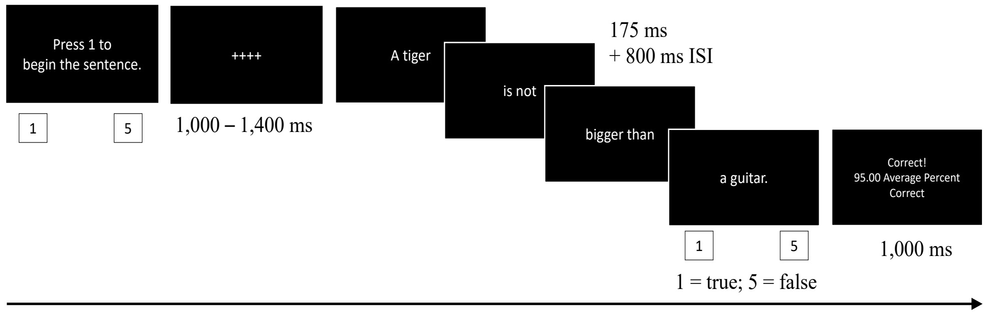

Figure 1: Trial structure.

orders of the blocks and the items were randomized for each participant. The participants had eight practice trials before the actual experiment.

2.4. EEG recording and data processing. We recorded EEG using 64-channel Hydrocel Geodesic Sensor Nets with a sampling rate of 250 Hz with Cz as the reference electrode. The continuous EEG data were run through a 0.1 Hz two-pass FIR high-pass filter, followed by a 40 Hz two-pass Butterworth low-pass filter. We segmented the data into four conditions (true affirmative, false affirmative, true negative, and false negative) excluding those with inaccurate responses. Each segment was time-locked to the onset of the object chunk and contained a 200 ms pre-stimulus period for baseline correction followed by a 1,000 ms post-stimulus period. To remove common artifacts (eye blinks, muscle movements, and saccade movements), the EEG data underwent the automated artifact correction with the Multi-Algorithm Artifact Correction (MAAC) procedure (Dien 2024). For eye blinks and saccades, each participant's data underwent independent component analysis (ICA) decomposition, and components that correlated at r = 0.95 with predefined blink and saccade templates were subtracted from the data, before the data were reconstituted by remixing the components. Bad channels were replaced with spline interpolations of the surrounding channels. After this step, the data were baseline corrected again and re-referenced to the average of all channels. Each participant's single trial data was then averaged into the four conditions.

2.5. ERP analysis plan. The aim of the experiment was to identify the brain response to true vs. false sentences and whether this response materialized as an N400 effect. The N400 is typically described as a centro-parietal negativity in the 300–500 ms time window. One approach to analysis would then be to simply average Cz, Pz, or some collection of centro-parietal midline electrodes and the 300–500 ms time window for each condition and participant and submit the resulting voltage values to an ANOVA. However, each experiment is slightly different, and the ERP may jitter temporally and/or spatially depending on the particular task demands, stimulus presentation modes, and participant populations. *A priori* defined time windows and electrode clusters may therefore not fit the observed data precisely. To improve the precision of time windows and electrode region selection relative to the data set at hand, we therefore utilized sequential temporo-spatial principle component analysis (PCA; Dien 2010, 2012, Dien & Frishkoff 2005), which decomposes the observed surface mixture data into underlying temporal and spatial components of the brain response to the experimental manipulations. Specifically, we used the two false minus

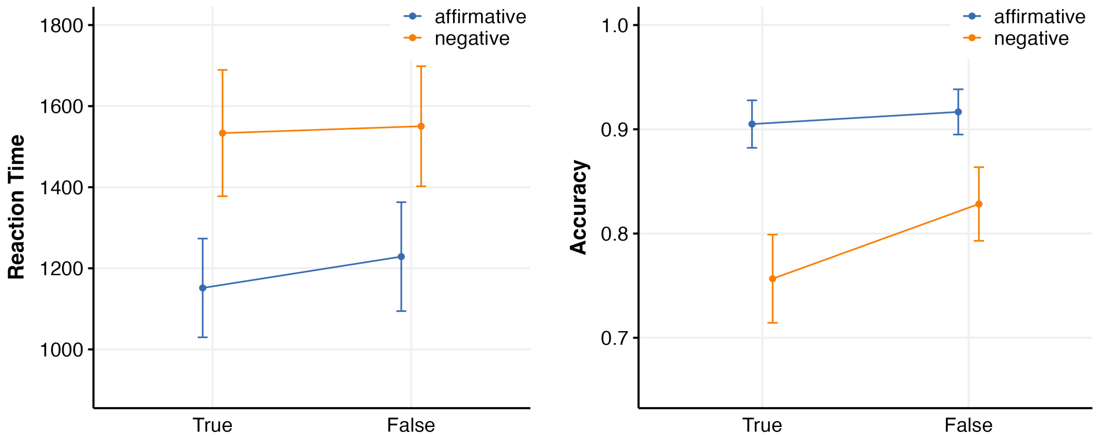

Figure 2: Mean accuracy (right) and RT (left) by condition.

true difference waves for affirmatives and negatives as input to the temporal PC, because the experiment sought to identify whether the N400 differed for the two truth value conditions. The PCA decomposes the 0–1,000 ms segment into latent time events reflecting the differential effect of truth value differences. We then spatially decompose each temporal factor with ICA to determine the main electrode source of those temporal events.

To assess the statistical significance of condition differences in the observed PCA components, we planned to use the resulting factor scores in each latent time/space component and each cell and participant as dependent measures, and analyze whether the affirmative and negative difference scores differed significantly from zero (a main effect of truth value), and from each other (an interaction with sentence form).

3. Results.

3.1. Behavioral results. We conducted a 2 x 2 repeated measures ANOVA with sentence form and truth value as independent variables. The right panel in Figure 2 shows the mean accuracy for each condition. For accuracy, we found a significant main effect of sentence form such that affirmatives were judged more accurately than negatives (91 vs. 79%; $F(1{,}29) = 93.5$, $p < 0.001$). We also found the significant main effect of truth value: false sentences were judged more accurately than true sentences (83 vs. 88%; $F(1{,}29) = 11.1$, $p < 0.01$). Finally, there was a sentence form by truth value interaction such that the difference between true and false affirmatives (91 vs 92%) was smaller than the difference between true and false negatives (76 vs. 83%; $F(1{,}29) = 8.2$, $p < 0.01$).

The left panel in Figure 2 shows the mean RT by condition. For RTs, we used data that evoked correct responses. We found the significant main effects of sentence form and truth value: affirmatives were faster to judge than negatives (1170 vs. 1516 ms; $F(1{,}29) = 131.5$, $p < 0.001$) and true sentences were faster to judge than false ones (1317 vs. 1368 ms; $F(1{,}29) = 4.59$, $p < 0.05$).

3.2. ERP results.

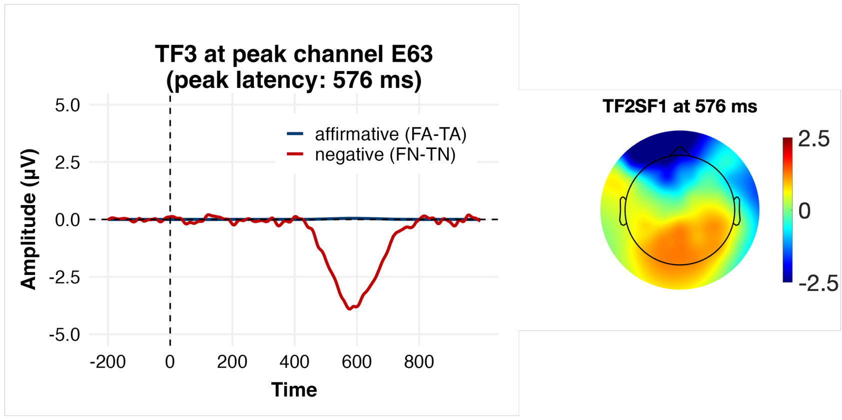

Figure 3: Waveforms (left) of microvolt-scaled factor loadings from the temporo-spatial factor TF2SF1 and the topographic map (right) of the difference between the two false-true differences: the false-true difference in the negative form minus the false-true difference in the affirmative form.

3.2.1. Temporo-spatial PCA and ICA. To analyze our ERP data without experimenter bias, we conducted a temporo-spatial PCA/ICA of all data (using all time samples and all electrodes) with the ERP PCA tool matlab toolbox (Dien 2010). This exploratory step would determine the latent effects, if any, of truth-value computation on the EEG activity. We followed the recommended analysis pipeline in the ERP PCA toolkit tutorial (Dien 2010), using temporal PCA based on the covariance of time samples. This was followed by a spatial ICA for each temporal factor based on the covariance of electrodes, to further separate out distinct spatial sources of variance within each temporal factor. To determine if the data contained a response to the difference between true and false sentences, we created a main effect of truth difference wave (i.e., (false affirmative + false negative) – (true affirmative + true negative) for each participant). This single dependent measure was then used as input for the PCA. To reduce dimensionaltiy in the time domain, we retained 13 factors from the temporal PCA solution, which determines 93% of the total variance. Following the recommendations in Dien (2010), we limit analysis to only factors that account for a sizable amount of variance, arbitrarily set to more than 6% of the total variance. This was true for the first four temporal factors (TFs): TF1 (peaked at 880 ms and accounted for 34% of the variance), TF2 (576 ms, 20%), TF3 (384 ms, 10%), and TF4 (728 ms, 7%). We next applied a spatial ICA decomposition to each of the temporal factors, to further narrow down the different spatial regions in each temporal factor according to how much variance is accounted for. Four spatial factors were retained for each temporal factor. Temporal factor 2, spatial factor 1 (F2SF1) exhibited a left anterior negativity, and was thus taken as a latent factor reflecting an LAN effect in the grand average voltage data. Figure 3 shows that the effect was carried by the negated sentence.

In contrast to a clear LAN effect, our visual inspection of the temporal and temporo-spatial factors revealed no N400 effect. For an inferential test, we reconstructed TF2SF1 in voltage space by multiplying the factor loadings with the factor scores (Dien & Frishkoff 2005), resulting in two reconstructed difference waves: affirmative (FA – TA) and negative (FN – TN). We then obtained the dependent measure by averaging the amplitudes of each difference wave over the PCA-delimited time window (506–668 ms) and 11 frontal electrodes (E5, E6, E8, E9, E10, E11, E12, E17, E55, E58, E63). We tested for the presence of a LAN effect in the affirmative and negative sentence forms separately by conducting a one-sample t-test for each sentence form to evaluate whether the difference in amplitudes between the false and true conditions was significantly different from 0. Additionally, we performed a paired-sample t-test to compare the two false-true difference amplitudes between the two sentence forms, assessing whether the LAN effect differed across the sentence forms. The results showed that the effect was not significant for the affirmative condition ($t(29) = -0.17$, $p = .885$) but was significant for the negative condition ($t(29) = -2.33$, $p = .027$). In addition, the amplitude difference between the two conditions was significant ($t(29) = 2.53$, $p = .017$). The inferential results confirmed a LAN effect driven by the false-true difference in the negative (FN — TN) condition (Figure 4).

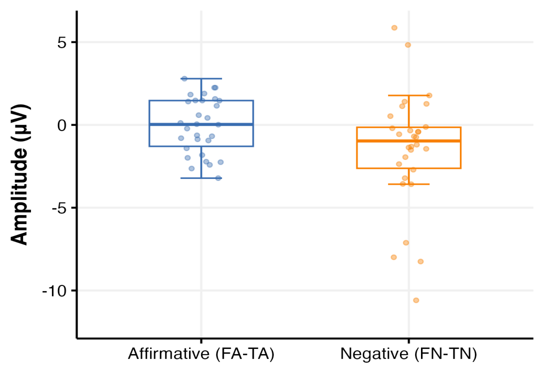

Figure 4: Boxplot of reconstructed TF2SF1 voltage data for the affirmative and negative conditions. Each dot represents an individual data point.

Figure 5 shows the waveform of the *ICA regionalized mean channel* for the main effect of truth value in the affirmative (left) and negative (right) conditions.

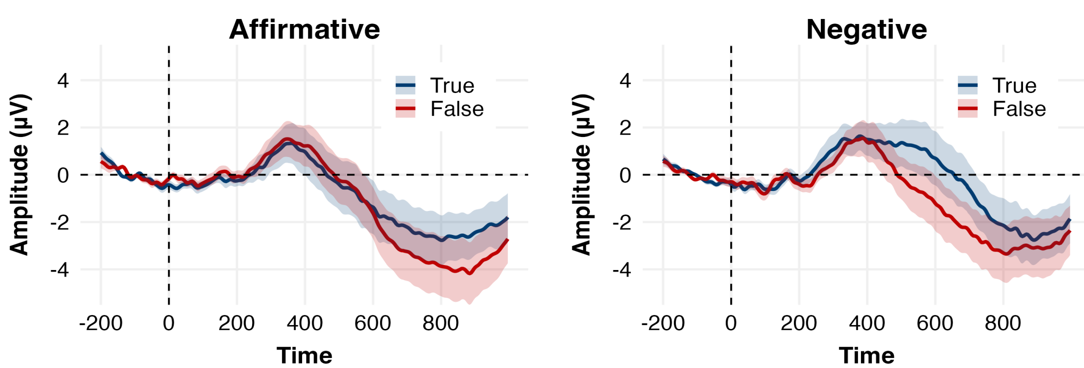

Figure 5: Waveform plots with 84% confidence intervals.

4. Discussion. Our behavioral results replicated prior findings (Clark & Chase 1972, Carpenter & Just 1975, Fischler et al. 1983, Palaz et al. 2020), indicating that our experiment tapped into a similar cognitive process as the previous studies. However, the ERP results revealed no N400 effect in the relevant comparisons (i.e., false affirmative – true affirmative or true negative – false negative). This ERP result was consistent with our second prediction: if the previously observed ERP pattern came from the lexical priming, the N400 effect would not be observed. Instead of the N400 effect, we found the main effects of truth value and sentence form, and their interaction in a LAN effect during the 504–673 ms time window. Although we must leave a precise explanation to future research, we suspect that this LAN effect has to do with the truth verification process.

No N400 effect suggests that the final word in the experimental sentences elicited a roughly equal N400 amplitude across conditions. We attribute this result to the lack of semantic relation between subject and object in our stimuli. If this is correct, the previously observed negation-blind N400 effect can be said to result from the lexical priming: the critical word attenuated the N400 when preceded by a semantically related word, compared to being preceded by an unrelated word. This lexical priming account can explain findings by some ERP studies on negation processing (Fischler et al. 1983, Kounios & Holcomb 1992, Dudschig et al. 2019, Haase et al. 2019) but it is silent about the pragmatics-related negation processing reported in Nieuwland & Kuperberg (2008) and Palaz et al. (2020). Hence, let us discuss these two studies in detail below.

Contrary to other ERP studies, Nieuwland & Kuperberg (2008) found a 'negation-sensitive' N400 pattern when negative sentences appeared under pragmatically appropriate contexts. Specifically, they found that false negative sentences elicited a larger N400 than their true versions, when preceded by contextual phrases (e.g., (7b) vs. (7a), where *with proper equipment* serves as a contextual phrase).

(7) a. With proper equipment, scuba-diving isn't very dangerous ... (true negative)
 b. With proper equipment, scuba-diving isn't very safe ... (false negative)

(Nieuwland & Kuperberg 2008: 1214)

Our lexical priming account is silent about this negation-sensitive N400 pattern because it is purely concerned with a low-level lexical process. However, this does not mean that it is flawed. Previous ERP studies showed that a high-level pragmatic process can override the low-level lexical process (Nieuwland & Van Berkum 2006, Nieuwland et al. 2010, Filik & Leuthold 2008, Hunt III et al. 2013).[5] Specifically, words that caused semantic anomaly did not elicit a large N400 when embedded under pragmatically appropriate contexts. For instance, Nieuwland & Van Berkum (2006) found that *salted* caused a larger N400 than *in love* in the sentence *the peanut was* {*salted/in love*}. embedded under a story about an amorous peanut. This finding and others by the studies cited above suggest that pragmatic information can override a low-level lexical effect. Under this consideration, the negation-sensitive N400 in Nieuwland & Kuperberg (2008) can be said to stem from the overriding effect of pragmatic information.

At first glance, the current assumption about the relationship between pragmatics and lexical semantics seems to conflict with previous findings by Palaz et al. (2020). Combining a pseudo-word learning paradigm with a truth-value judgment task, Palaz et al. (2020) conducted an ERP experiment to examine the effect of the informative use of negation on an N400. In their design,

[5] Our discussion is consistent with the original claim made by Nieuwland & Kuperberg (2008) that pragmatic effects can lead to incremental negation processing, but we resort to other studies in our discussion to avoide circularity.

negative sentences followed context sentences, as exemplified by (8). They used a pseudo-word for a subject (e.g., *sloken* or *trante*), and the context sentence enabled participants to learn what type of entity it referred to (e.g., tool or fish).[6] The false negative sentence like (8b) was assumed to be pragmatically informative because it denied the participants' belief about the referent of the pseudo-word (e.g., one assumes *trante* to be some kind of fish based on the context in (8b).)

(8) a. Context: I used a huge sloken to bang the nail in the gate.
A sloken is not a fish. (true negative)

b. Context: I saw a trante with exotic gills in the National Aquarium.
A trante is not a fish. (false negative)

(Palaz et al. 2020: 4)

Despite the contextual manipulation, Palaz et al. (2020) observed a negation-blind N400 effect: the true negative sentence (8a) elicited a larger N400 than their false version (8b). At first sight, this 'pragmatics-insensitive' N400 effect seemingly speaks against our assumption about the overriding effect of pragmatics. However, a closer look at behavioral results and a task design suggests that their manipulation of pragmatic informativeness was not sufficient enough to trigger the pragmatic effect.

First, their behavioral results showed that the negative sentences took significantly longer to judge than the affirmative sentences in the presence of contexts. Previous studies found that the processing asymmetry between the negative and affirmative sentences disappeared under appropriate pragmatic contexts (Wason 1965, Johnson-Laird & Tridgell 1972, Glenberg et al. 1999) but not under inappropriate ones (Experiment 2 in Dale & Duran 2011). Given these previous findings, the higher processing load in the negative than affirmative sentences in Palaz et al. (2020) suggests that their contextual manipulation could not trigger the strong effect of pragmatic information.

Second, their experiment had a methodological concern. Palaz et al. (2020: 3) say "negation can become informative in contexts which express beliefs or predictions" because negation can correct the false beliefs that people have. We agree with this view, but it is doubtful that their experimental task properly tapped into this pragmatic effect of negation. In the task, for instance, participants were instructed to judge whether *a trante is not a fish.* in (8b) is plausible or implausible via button press. If this negative sentence was pragmatically informative due to it denying the participants' belief about the referent of the word *trante*, the pragmatically correct response should be plausible but not implausible. However, we would judge this sentence as *implausible* if we were naive participants because it just contradicts what the context has described. In fact, Palaz et al. (2020) also seem to share this judgment, since they categorized implausible as 'accurate' in this case. Their behavioral results further suggest that this is a general intuition: they showed more than 88% 'accuracy' on the relevant condition. This strongly suggests that the participants in Palaz et al. (2020) did not behave in a way that is consistent with how the negation should pragmatically work. For this reason, it is not clear whether their task was properly designed to examine the informative use of negation. A more appropriate task would be the one in which participants read a sentence

[6]Palaz et al. (2020) conducted a control experiment with real words. The results were similar to those from the experiment with the pseudo-words. They can be explained under our lexical priming account.

and do a secondary task which does not involve plausibility judgments (e.g., a word recognition task; Nieuwland & Kuperberg 2008).

These two considerations raise the possibility that the results of Palaz et al. (2020) do not reflect the informative use of negation. Therefore, the pragmatics-insensitive N400 pattern observed in Palaz et al. (2020) is not problematic for our assumption about the effect of pragmatics.

To conclude, removing the lexical priming relation between subject and object led to no N400 effect. Although we must interpret a statistically non-significant result with caution, this suggests that the putative negation-blind N400 effect resulted from lexical priming. If this is correct, it does not necessarily constitute evidence for a strict two-step account.

References

Bentin, Shlomo, Gregory McCarthy & Charles C Wood. 1985. Event-related potentials, lexical decision and semantic priming. *Electroencephalography and clinical Neurophysiology* 60(4). 343–355. 10.1016/0013-4694(85)90008-2.

Carpenter, Patricia A & Marcel A Just. 1975. Sentence comprehension: A psycholinguistic processing model of verification. *Psychological review* 82(1). 45. 10.1016/0010-0285(72)90019-9.

Christensen, Ken Ramshøj. 2020. The neurology of negation: fMRI, ERP, and aphasia. In Viviane Déprez & M Teresa Espinal (eds.), *The Oxford handbook of negation*, 725–739. Oxford: Oxford University Press. 10.1093/oxfordhb/9780198830528.013.47.

Clark, Herbert H & William G Chase. 1972. On the process of comparing sentences against pictures. *Cognitive psychology* 3(3). 472–517. 10.1016/0010-0285(72)90019-9.

Dale, Rick & Nicholas D Duran. 2011. The cognitive dynamics of negated sentence verification. *Cognitive science* 35(5). 983–996. 10.1111/j.1551-6709.2010.01164.x.

Dien, Joseph. 2010. The ERP PCA toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of neuroscience methods* 187(1). 138–145. 10.1016/j.jneumeth.2009.12.009.

Dien, Joseph. 2012. Applying principal components analysis to event-related potentials: A tutorial. *Developmental neuropsychology* 37(6). 497–517. 10.1080/87565641.2012.697503.

Dien, Joseph. 2024. Multi-algorithm artifact correction (MAAC) procedure part one: Algorithm and example. *Biological Psychology* 188. 108775. 10.1016/j.biopsycho.2024.108775.

Dien, Joseph & Gwen A Frishkoff. 2005. Principal components analysis of event-related potential datasets. In Handy Todd C (ed.), *Event-related potentials: A methods handbook*, 189–208. MIT press Cambridge.

Dudschig, Carolin, Barbara Kaup, Mingya Liu & Juliane Schwab. 2021. The processing of negation and polarity: An overview. *Journal of Psycholinguistic Research* 50(6). 1199–1213. 10.1093/oxfordhb/9780198830528.013.47.

Dudschig, Carolin, Ian Grant Mackenzie, Claudia Maienborn, Barbara Kaup & Hartmut Leuthold. 2019. Negation and the N400: Investigating temporal aspects of negation integration using semantic and world-knowledge violations. *Language, Cognition and Neuroscience* 34(3). 309–319. 10.1080/23273798.2018.1535127.

Filik, Ruth & Hartmut Leuthold. 2008. Processing local pragmatic anomalies in fictional contexts: Evidence from the N400. *Psychophysiology* 45(4). 554–558. 10.1111/j.1469-

8986.2008.00656.x.

Fischler, Ira, Paul A Bloom, Donald G Childers, Salim E Roucos & Nathan W Perry Jr. 1983. Brain potentials related to stages of sentence verification. *Psychophysiology* 20(4). 400–409. 10.1111/j.1469-8986.1983.tb00920.x.

Glenberg, Arthur M, David A Robertson, Jennifer L Jansen & Mina C Johnson-Glenberg. 1999. Not propositions. *Cognitive Systems Research* 1(1). 19–33. 10.1016/S1389-0417(99)00004-2.

Haase, Viviana, Maria Spychalska & Markus Werning. 2019. Investigating the comprehension of negated sentences employing world knowledge: An event-related potential study. *Frontiers in psychology* 10. 2184. 10.3389/fpsyg.2019.02184.

Hagoort, Peter, Lea Hald, Marcel Bastiaansen & Karl Magnus Petersson. 2004. Integration of word meaning and world knowledge in language comprehension. *Science* 304(5669). 438–441. 10.1126/science.1095455.

Hald, Lea A, Esther G Steenbeek-Planting & Peter Hagoort. 2007. The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain research* 1146. 210–218. 10.1016/j.brainres.2007.02.054.

He, Yifei, Johanna Sommer, Silvia Hansen-Schirra & Arne Nagels. 2022. Multivariate pattern analysis of EEG reveals nuanced impact of negation on sentence processing in the N400 and later time windows. *Psychophysiology* e14491. 10.1111/psyp.14491.

Hunt III, Lamar, Stephen Politzer-Ahles, Linzi Gibson, Utako Minai & Robert Fiorentino. 2013. Pragmatic inferences modulate N400 during sentence comprehension: Evidence from picture–sentence verification. *Neuroscience Letters* 534. 246–251. 10.1016/j.neulet.2012.11.044.

Jakobovits, Leon A & Wallace E Lambert. 1962. Mediated satiation in verbal transfer. *Journal of Experimental Psychology* 64(4). 346. https://psycnet.apa.org/doi/10.1037/h0044630.

Johnson-Laird, Philip N & JM Tridgell. 1972. When negation is easier than affirmation. *Quarterly Journal of Experimental Psychology* 24(1). 87–91. 10.1080/14640747208400271.

Kaup, Barbara & Carolin Dudschig. 2007. The experiential view of language comprehension: How is negation represented? In Schmalhofer Franz & Perfetti Charles A. (eds.), *Higher level language proceesses in the brain. Iinference and comprehension processes*, 255–288. London: Lawrence Erlbaum Associates Publishers.

Kaup, Barbara & Carolin Dudschig. 2020. Understanding negation: Issues in the processing of negation. In Viviane Déprez & M Teresa Espinal (eds.), *The Oxford handbook of negation*, 635–655. Oxford: Oxford University Press. 10.1093/oxfordhb/9780198830528.001.0001.

Kaup, Barbara, Jana Lüdtke & Rolf A Zwaan. 2006. Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics* 38(7). 1033–1050. 10.1016/j.pragma.2005.09.012.

Kaup, Barbara, Richard H Yaxley, Carol J Madden, Rolf A Zwaan & Jana Lüdtke. 2007. Experiential simulations of negated text information. *Quarterly journal of experimental psychology* 60(7). 976–990. 10.1080/17470210600823512.

Kounios, John & Phillip J Holcomb. 1992. Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. *Journal of experimental psychology: General* 121(4). 459. https://psycnet.apa.org/doi/10.1037/0096-3445.121.4.459.

Kutas, Marta & Kara D Federmeier. 2011. Thirty years and counting: Finding meaning in the

N400 component of the event-related brain potential (ERP). *Annual review of psychology* 62. 621–647. 10.1146/annurev.psych.093008.131123.

Kutas, Marta & Steven A Hillyard. 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307(5947). 161–163. 10.1038/307161a0.

Metzner, Paul, Titus von der Malsburg, Shravan Vasishth & Frank Rösler. 2015. Brain responses to world knowledge violations: A comparison of stimulus-and fixation-triggered event-related potentials and neural oscillations. *Journal of Cognitive Neuroscience* 27(5). 1017–1028. 10.1162/jocn$_{a0}$0731.

Nieuwland, Mante S, Dale J Barr, Federica Bartolozzi, Simon Busch-Moreno, Emily Darley, David I Donaldson, Heather J Ferguson, Xiao Fu, Evelien Heyselaar, Falk Huettig et al. 2019. Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society B* 375(1791). 20180522. 10.1098/rstb.2018.0522.

Nieuwland, Mante S, Tali Ditman & Gina R Kuperberg. 2010. On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of memory and language* 63(3). 324–346. 10.1016/j.jml.2010.06.005.

Nieuwland, Mante S & Gina R Kuperberg. 2008. When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science* 19(12). 1213–1218. 10.1111/j.1467-9280.2008.02226.x.

Nieuwland, Mante S & Jos JA Van Berkum. 2006. When peanuts fall in love: N400 evidence for the power of discourse. *Journal of cognitive neuroscience* 18(7). 1098–1111. 10.1162/jocn.2006.18.7.1098.

Palaz, Bilge, Ryan Rhodes & Arild Hestvik. 2020. Informative use of "not" is N400-blind. *Psychophysiology* 57(12). e13676. 10.1111/psyp.13676.

Papeo, Liuba & Manuel de Vega. 2020. The neurobiology of lexical and sentential negation. In Viviane Déprez & M Teresa Espinal (eds.), *The Oxford handbook of negation*, 740–756. Oxford: Oxford University Press. 10.1093/oxfordhb/9780198830528.013.44.

Rugg, Michael D. 1985. The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology* 22(6). 642–647. 10.1111/j.1469-8986.1985.tb01661.x.

Smith, Lee & Raymond Klein. 1990. Evidence for semantic satiation: Repeating a category slows subsequent semantic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16(5). 852. 10.1037/0278-7393.16.5.852.

Tian, Ye & Richard Breheny. 2019. Negation. In Cummins Chris & Katsos Napoleon (eds.), *The Oxford handbook of experiemental semantics and pragmatics*, 195–207. Oxford: Oxford University Press. 10.1093/oxfordhb/9780198791768.013.29.

Wason, Peter C. 1965. The contexts of plausible denial. *Journal of verbal learning and verbal behavior* 4(1). 7–11. 10.1016/S0022-5371(65)80060-3.

Wiswede, Daniel, Nicolas Koranyi, Florian Müller, Oliver Langner & Klaus Rothermund. 2013. Validating the truth of propositions: Behavioral and ERP indicators of truth evaluation processes. *Social Cognitive and Affective Neuroscience* 8(6). 647–653. 10.1093/scan/nss042.