C E N T E R   F O R   R E S E A R C H   I N   L A N G U A G E

December 2008 **Vol. 20, No. 3**

CRL Technical Reports, University of California, San Diego, La Jolla CA 92093-0526
Tel: (858) 534-0714 • E-mail: editor@crl.ucsd.edu • WWW: http://crl.ucsd.edu/newsletter/current/Newsletter/articles.html

TECHNICAL REPORT

Negation Processing in Context Is Not (Always) Delayed

Jenny Staab [ab], Thomas P. Urbach [c], and Marta Kutas [bc]

[a] Joint Doctoral Program in Language and Communicative Disorders,
San Diego State University & University of California, San Diego

[b] Center for Research in Language, University of California, San Diego

[c] Department of Cognitive Science, University of California, San Diego

EDITOR'S NOTE

This newsletter is produced and distributed by the **CENTER FOR RESEARCH IN LANGUAGE**, a research center at the University of California, San Diego that unites the efforts of fields such as Cognitive Science, Linguistics, Psychology, Computer Science, Sociology, and Philosophy, all who share an interest in language. We feature papers related to language and cognition (distributed via the World Wide Web) and welcome response from friends and colleagues at UCSD as well as other institutions. Please visit our web site at `http://crl.ucsd.edu`.

SUBSCRIPTION INFORMATION

If you know of others who would be interested in receiving the Newsletter and the Technical Reports, you may add them to our email subscription list by sending an email to `majordomo@crl.ucsd.edu` with the line `"subscribe newsletter <email-address>"` in the body of the message (e.g., `subscribe newsletter jdoe@ucsd.edu`). Please forward correspondence to:

Jamie Alexandre, Editor
Center for Research in Language, 0526
9500 Gilman Drive, University of California, San Diego 92093-0526
Telephone: (858) 534-0714 • E-mail: editor@crl.ucsd.edu

Back issues of the the CRL Newsletter are available on our website. Papers featured in recent issues include:

A Phonetic Study of Voiced, Voiceless, and Alternating Stops in Turkish
Stephen M. Wilson
Neuroscience Interdepartmental Program, UCLA
Vol. 15, No. 1, April 2003

New corpora, new tests, and new data for frequency-based corpus comparisons
Robert A. Liebscher
Cognitive Science, UCSD
Vol. 15, No.2; December 2003

The relationship between language and coverbal gesture in aphasia
Eva Schleicher
Psychology, University of Vienna & Cognitive Science, UCSD
Vol. 16, No. 1, January 2005

In search of Noun-Verb dissociations in aphasia across three processing tasks
Analía Arévalo, **Suzanne Moineau**
Language and Communicative Disorders, SDSU & UCSD, Center for Research in Language, UCSD
Ayşe Saygin
Cognitive Science & CRL, UCSD
Carl Ludy
VA Medical Center Martinez
Elizabeth Bates
Cognitive Science & CRL, UCSD
Vol. 17, No. 1, March 2005

Meaning in gestures: What event-related potentials reveal about processes underlying the comprehension of iconic gestures
Ying C. Wu
Cognitive Science Department, UCSD
Vol. 17, No. 2, August 2005

What age of acquisition effects reveal about the nature of phonological processing
Rachel I. Mayberry
Linguistics Department, UCSD
Pamela Witcher
School of Communication Sciences & Disorders, McGill University
Vol. 17, No.3, December 2005

Effects of Broca's aphasia and LIPC damage on the use of contextual information in sentence comprehension
Eileen R. Cardillo
CRL & Institute for Neural Computation, UCSD
Kim Plunkett
Experimental Psychology, University of Oxford
Jennifer Aydelott
Psychology, Birbeck College, University of London)
Vol. 18, No. 1, June 2006

Avoid ambiguity! (If you can)
Victor S. Ferreira
Department of Psychology, UCSD
Vol. 18, No. 2, December 2006

Arab Sign Languages: A Lexical Comparison
Kinda Al-Fityani
Department of Communication, UCSD
Vol. 19, No. 1, March 2007

The Coordinated Interplay Account of Utterance Comprehension, Attention, and the Use of Scene Information
Pia Knoeferle
Department of Cognitive Science, UCSD
Vol. 19. No. 2, December 2007

Doing time: Speech, gesture, and the conceptualization of time
Kensy Cooperrider, Rafael Núñez
Depatment of Cognitive Science, UCSD
Vol. 19. No. 3, December 2007

Auditory perception in atypical development: From basic building blocks to higher-level perceptual organization
Mayada Elsabbagh
Center for Brain and Cognitive Development, Birkbeck College, University of London
Henri Cohen
Cognitive Neuroscience Center, University of Quebec
Annette Karmiloff-Smith
Center for Brain and Cognitive Development, Birkbeck College, University of London
Vol. 20. No. 1, March 2008

The Role of Orthographic Gender in Cognition
Tim Beyer, Carla L. Hudson Kam
Center for Research in Language, UCSD
Vol. 20. No. 2, June 2008

NEGATION PROCESSING IN CONTEXT IS NOT (ALWAYS) DELAYED

Jenny Staab [ab], Thomas P. Urbach [c], and Marta Kutas [bc]

[a] Joint Doctoral Program in Language and Communicative Disorders, San Diego State University & University of California, San Diego

[b] Center for Research in Language, University of California, San Diego

[c] Department of Cognitive Science, University of California, San Diego

Abstract

Although most linguistic cues are thought to affect subsequent processing (almost) immediately after they are encountered, negation has traditionally been viewed as an operator that has its effects only after the negated sentence has been processed. Consequently, most tests for effects of negation have been post-sentential. One prior study using event-related brain potentials (ERPs) to detect negation effects on the processing of subsequent words within the same sentence failed to observe any. We maintain that this failure was due to the use of isolated sentences in which negation was not pragmatically licensed and did not change the expectancy for the sentence endings. To make negation-induced expectation changes detectable, we embedded affirmative and negative sentences in discourse contexts in which negation impacted the expectancy for and plausibility of a continuation; i.e., expectancies for negative and affirmative sentences differed. We conducted a series of three experiments. One used the event-related brain potential (ERP) methodology, especially the N400 to the sentence-final words as the main index of word expectancy. The N400 results revealed that negation can affect expectancies for sentence continuations. The ERP study was complemented by two verification experiments, that differed in the presentation mode for the target sentence (word-by-word vs. whole-sentence). The comparison of verification times indicated that for negation-induced expectation changes to occur readers must have enough time and available processing capacity. In sum, when pragmatically licensed and supported by processing resources, the effects of negation can – like other operators – be (almost immediate) and intra-sentential.

Introduction

Negation has many effects. Apart from changing the meaning of a sentence, it affects the way the sentence and its constituents are processed. In experimental settings, however, negation effects have varied widely. Much of the observed variation can be attributed to differences in the time points at which negation effects have been probed and, importantly, differences in the presence or absence and nature of discourse contexts used.

Recent psycholinguistic research has focused on the representations of negated concepts (e.g., Giora, Balaban, Fein, & Alkabets, 2004; Kaup, 1997, 2001; Lüdtke, Friedrich, De Filippis, & Kaup, 2008; MacDonald & Just, 1989). More specifically, these studies have examined the question of whether or not negation reduces the activation of concepts to which it applies. It has been proposed, for example, that negation acts as a corrective device, shifting attention by suppressing an activated element and allowing an alternative to be activated instead (De Mey, 1971). A number of studies have in fact reported evidence for reduced activation of negated concepts (Kaup, 1997, 2001; Kaup, Lüdtke, & Zwaan, 2006; MacDonald & Just, 1989). MacDonald and Just (1989), for instance, found that probes that had been negated in a preceding sentence (1a) were named more slowly than were non-negated probes (1b).

(1) Almost every weekend, Elizabeth bakes no bread but only cookies for the children.
   a. bread
   b. cookies

This effect does not seem to be obligatory, however, as others have failed to observe suppression due to negation (Giora et al., 2004; Kaup, Lüdtke, & Burkert, 2006, Kaup, 2007; Lüdtke et al., 2008). Overall, it appears that suppression effects are more likely to be detected in certain experimental settings: when alternatives to the negated concept are readily available; when sentences are presented outside of a global context that would necessitate the retention of the negated information to establish coherence; and when negation effects are probed after longer delays (Giora, 2006, 2007; Giora, Fein, Aschkenazi, & Alkabets-Zlozover, 2007; Staab, 2007).

The most well documented and almost universal negation effect, however, is the increase in processing difficulty associated with negation: Numerous studies of both comprehension and production have shown that negation leads to longer response times and higher error rates (Carpenter & Just, 1975; Carpenter, Just, Keller, Eddy, & Thulborn, 1999; Clark, 1976; Kim, 1985; Sherman, 1976; Trabasso, Rollins, & Shaughnessy, 1971; Wason, 1959, 1961). The classic paradigm used to assess these effects has been the sentence-picture verification task, in which subjects are asked to judge whether an affirmative or negative sentence is true or false with respect to a simple visual display. Clark and Chase (1972), for instance, presented pictures of a star and a plus ($\overset{*}{+}$ or $\overset{+}{*}$) along with an sentence describing a particular spatial configuration of the two objects. The sentences could be affirmative or negative and true or false:

[$\overset{*}{+}$]

(2) The star is above the plus.
(star above plus) *True Affirmative (TA)*

(3) The plus is above the star.
(plus above star) *False Affirmative (FA)*

(4) The star isn't above the plus.
(not (star above plus)) *False Negative (FN)*

(5) The plus isn't above the star.
(not (plus above star)) *True Negative (TN)*

This experiment produced the response time pattern most frequently observed in verification studies: TA < FA < FN < TN. Affirmative sentences were verified faster than negative ones, but while TA elicited faster response times than FA, TN led to slower responses than FN. Clark and Chase's explanation of this pattern was based on the assumption that negative sentences like (5) are represented as a positive inner proposition *(plus above star)* embedded in a negative outer proposition *(not())*. Under this assumption, part of the difference in the response times between negative and affirmative sentences is due to the additional time it would take to encode the negative outer proposition.

The construction of the mental representations would be followed by the comparison of the sentence representation with the picture representation, proceeding from the innermost to the outermost proposition. Any mismatch between propositions would flip the truth index (initially set to *true*), which in turn would increase the verification response time. In the example above, the picture would be represented as *(star above plus)*. The representation of the TA (2) is identical to that of the picture; the response index thus need not change, and the verification time consequently would not increase. For the FA (3), the comparison with the picture results in a mismatch of the inner propositions. The response index therefore would thus flip to *false*, leading to longer verification times than for TA. Verifiying FN (4) would lead to even longer response latencies. Its inner proposition *(star above plus)* matches that of the picture, but its outer proposition *(not ())* does not. The time needed to flip the response index (to *false*) would be added to the extra time for encoding the negative proposition, resulting in the observed increase in verification time. For the TN (5), both the inner and the outer propositions conflict with those of the picture. Consequently, the response index first changes to *false* and then back to *true*. Combined with the additional encoding time for negation, this condition would yield the longest verification response times.

The notion of negation as an embedding proposition is consonant with Klima's (1964) classic analysis of negation as well as its role in the propositional logic tradition, in which it is an operator applied to an entire sentence or proposition (Frege, 1884, 1919). This view predicts that negation can only be processed after the processing of the embedded affirmative proposition is complete. Consequently, negation effects would only be observed after the negated sentence or clause has come to an end. Indeed, all the negation effects mentioned above were assessed only post-sentence, and the probability of observing effects of negation increased with the delay.

Delayed processing of negation, however, would seem to conflict with our intuitions as we do not consider the negated information as initially true but become false after a phrase or sentence. More relevant, delayed processing of negation also is at odds with all the ever-mounting evidence for incremental language processing. Psycholinguistic research has demonstrated that all sorts of linguistic information are used as soon as they become available (Altmann & Kamide, 1999; Crocker &

Brants, 2000; Kamide, Altmann, & Haywood, 2003; van Berkum, Koornneef, Otten, & Nieuwland, 2007), and it is not obvious why negation should be an exception.

If the negation marker is processed incrementally and integrated into the sentence representation as soon as it is encountered, then it also should have an observable impact on how subsequent information is processed within the remainder of that negative sentence. Presumably, by altering the semantic context in which a subsequently received word needs to be integrated, negation should affect how well the word fits with this context. In fact, it should influence a person's expectations about upcoming lexical items or at least semantic characteristics thereof (for arguments concerning on-line prediction see DeLong, Urbach, & Kutas, 2005; van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004).

The impact of negation on semantic context could take at least two different forms. Negation could change the incremental interpretation of the sentence such that a different lexical item is anticipated as more likely to occur and/or easier to integrate when it does occur. Alternatively (or simultaneously), negation could merely reduce the degree of sentential constraint. For the affirmative fragment "The capital of France is…", for example, only one possible completion can be reasonably expected. The corresponding negative fragment has hundreds of possible endings. Thus, the effects of negation can be qualitative, changing the most expected ending, or quantitative, decreasing sentence constraint.

The amplitude of a component of the ERP - the N400 - has been shown to vary as a function of semantic expectancy and fit to semantic context . Although not specific to words, the N400 has been used extensively to study sentence processing at the level of meaning. N400 amplitude varies with the semantic expectancy and semantic fit of a word within a context, with highly expected words that fit well within a context eliciting smaller N400s than those that are less expected and less good fits (Kutas & Hillyard, 1980, 1984).

If negation makes a word incongruous, or less probable (since other words would be plausible, too) for a given sentence, these changes in expectancy and associated integrative ease should be reflected in N400 amplitude. More than fifteen years ago, two studies used these properties of the N400 to study the effect of negation on the processing of semantic relationships (Fischler, Bloom, Childers, Roucos, & Perry, 1983; Kounios & Holcomb, 1992). No effects of negation on N400 amplitude were observed. Only response times in the verification task proved sensitive to the changes in meaning and truth-value due to negation.

Fischler and colleagues (1983) presented "class inclusion" statements, such as (6) through (9), and asked participants to verify the sentences. Each of these statements began with a concrete noun (e.g., *a robin*) and ended with a superordinate category name (e.g., *a bird*). The two nouns were connected either by *is*, or *is not*, yielding affirmative or negative sentences, respectively. The relationship between the concrete noun and the category also was varied: the concrete noun was either an exemplar of the category or not. The truth-value of the sentence thus depended on the combination of category relationship and the form of the statement. An affirmative statement was true when the first noun was a member of the category (6) and false when it was not (7). Negative sentences, by contrast, were true when the concrete noun was not a category member (9), and false when it was (8).

(6) A robin is a bird. (TA)
(7) A robin is a vehicle. (FA)
(8) A robin is not a bird. (FN)
(9) A robin is not a vehicle. (TN)

The dependent variable was the N400 amplitude to the final word (category label). Since it directly followed the negation marker, any N400 difference could be taken as evidence for an immediate effect of negation. To the extent that participants had updated their expectations for the category name based on the negation marker, the N400 to the final word would have been greater in (8) than in (6). The results, however, showed no such effect. N400 amplitude was determined exclusively by the relationship between the first and the second noun, in both affirmative and negative sentences. That is, if *robin* was the sentence subject, the N400 to *vehicle* was larger than that to *bird*, regardless of whether the sentence was true or false. While the ERPs did not appear to be sensitive to truth-value or negation, verification times showed the expected interaction between the two factors. Participants obviously processed the negation, but its effect were limited to late (post-N400) presumably interpretive processes. Fischler et al. explained the lack of a negation effect on N400 with reference to Clark's (1976) model of sentence verification: N400 reflected the computation of the positive inner proposition. The outer negation was processed later (than N400 operations), such that its effects were detected only later in the verification times.

There is, however, an alternative interpretation for these results. Fischler et al. (1983) used isolated sentences with pairs of nouns that were either very strongly related (e.g., *robin* and *bird*), or completely unrelated (e.g., *robin* and *vehicle*). Given this, the observed data pattern is not surprising given the pragmatics of negation. Negation is typically used to deny a supposition (Givón, 1979, 1984, 1989, 1993; Horn, 1989; Jespersen, 1917; Strawson, 1952; Wason, 1965), and in the absence of discourse context, this supposition must be grounded in general knowledge. That is, in isolation negation is used to deny something that is part of an invoked schema (Fillmore, 1985). Fischler's isolated experimental sentences evoke the schema associated with the first noun (e.g., *robin*). Consequently, only elements of that schema can be negated (e.g., *bird*). By contrast, unrelated items (e.g., *vehicle*), are not part of any invoked schema, do not constitute an acceptable completion, and cannot be facilitated or expected to be negated – just as the experimental data show.

It thus seems relatively unlikely that negation effects can be detected when sentences are used outside of context and the negation is not pragmatically licensed. In isolation, negative sentences can only deny stereotypical facts or assumptions – namely, the same information and lexical items that are associated with the affirmative sentence. As a result, the expected completions for affirmative and negative sentences are indistinguishable. In order to detect effects of negation on expectations about upcoming words, it would appear necessary to embed the experimental sentences in wider contexs – contexts that can provide suppositions or possibilities that can be plausibly denied and that are independent of stereotypical associations that seem to affect negative and affirmative sentences equally.

The type of context in which negative sentences occur doesn't just impact whether suppression effects can be observed. A context can also decrease the processing difficulty associated with negation. Wason (1965), for example, showed that sentences were easier to produce when they described an exception to a rule, that is, when they denied the assumption that every item behaved like the majority. The sentence "Circle number 4 is not red", for instance, was easier to produce in a situation in which the majority of circles was red compared to a situation when only one out of seven circles was red. Similarly, Glenberg and colleagues (Glenberg, Robertson, Jansen, & Johnson-Glenberg, 1999) as well as Lüdtke and Kaup (2006) found that the comprehension of negative sentences was facilitated when the sentences occurred in a context that provided for a hypothesis that the negative sentence denied. The context dependence of negation thus has been established theoretically as well as experimentally, and negation-induced changes in expectations for sentence continuations may be another example of a context-dependent negation effect.

The primary goal of the current study was to test the hypothesis that negation can have observable effects on the processing of subsequent words within the same sentence or clause. This hypothesis implies that the negation marker is integrated into the sentence on-line contra earlier proposals according to which negation is not considered until after the processing of the affirmative inner proposition has been completed (Carpenter & Just, 1975; Clark, 1976; Fischler et al., 1983).

As just argued, in order to detect such early negation effects in an experimental setting, it may be necessary to embed the negative sentences in a context that provides suppositions that can be plausibly denied. We have therefore employed choice scenarios as in Example (10-11).

(10) *Introduction*
During his long flight Joe needed a snack. The flight attendant could only offer him pretzels and cookies.

(11) *Affirmative bias*
- a. Joe wanted something salty.
- b. Joe wanted something sweet.

(12) *Target sentence*
- i. So he bought the pretzels.
- ii. So he bought the cookies.
- iii. So he didn't buy the pretzels.
- iv. So he didn't buy the cookies.

All stories were constructed according to the same pattern: The first two sentences (10) introduced two options. The following bias sentence (11), which in this set of experiments was always affirmative, provided information about the agent's preferences. Finally, the target sentence (12) presented the scenario outcome, i.e. the character's choice. The target sentence could be either affirmative or negative, and its final word was one of the two options initially introduced. In both the affirmative and negative case, the (correct) ending (which made the final sentence consistent with the preceding information) was completely predictable, as both options had been introduced, and favoring one (e.g., *salty* implied *pretzels*) excluded the other (i.e. not *cookies*). These stimuli thus differed importantly from those used in previous ERP experiments (Fischler et al., 1983; Lüdtke et al., 2008), where no clear prediction was possible for negative sentences (though see Nieuwland & Kuperberg, 2008)

Experiment 1: ERP analysis of negation in written sentences

The primary experimental support of the view that negation operates only after completed processing of the embedded affirmative proposition has come from Fischler and colleagues (1983). In that study, the N400 to the sentence-final word was the main variable of interest as its amplitude was independent of the presence of a negation marker in the sentence, and thus offered as support for no intra-sentential effect of negation. Experiment 1 was designed to refute Fischler's conclusions. To this end, choice scenarios like Example (10-11) were used in a verification paradigm, with the main dependent variable of interest being the N400 to the sentence final word. As in Fischler et al.'s experiment, sentential truth could only be determined upon perception of the final word.

Fischler and colleagues took the N400 to reflect a process of monitoring the consistency or validity of propositions. Although few researchers would subscribe to this particular functional interpretation of the N400, it is a good indicator of how expected (and in most cases how plausible) the eliciting word is within (or at the end of) a given sentence context. In fact, numerous investigations have shown that the N400 is sensitive to the match between a word and its context at different levels: lexical associations, sentential, and discourse. The N400 is reduced in amplitude when the eliciting word is preceded by a semantically related lexical item. This word level priming effect has been observed for word lists (Bentin, McCarthy, & Wood, 1985) as well as for word pairs embedded in sentences (Van Petten, Weckerly, McIsaac, & Kutas, 1997). N400 amplitude also depends on the fit between a word and overall sentence meaning (e.g. Van Petten et al., 1997). When a word is a good fit or highly expected in a sentence context, it elicits a smaller N400 than when it fits less well or is less expected (DeLong et al., 2005; Federmeier, Wlotko, De Ochoa-Ewald, & Kutas, 2007; Friederici, Pfeifer, & Hahne, 1993; Hagoort & Brown, 2000; Kuperberg et al., 2003; Kutas & Hillyard, 1984). The global discourse-context in which a sentence is embedded provides further constraints that can affect N400 amplitude. Words that fit equally well in an isolated sentence (e.g., *Fortunately, I didn't lose all my files/friends*.) will elicit smaller N400s if they are more consistent with the wider discourse context (i.e., *files* in *My computer system suddenly broke down.*), and larger N400s if they violate the discourse constraints (Salmon & Pratt, 2002; van Berkum, Hagoort, & Brown, 1999; van Berkum, Zwitserlood, Hagoort, & Brown, 2003). The N400 is thus sensitive to different forms of semantic context: lexical associations as well as sentence- and discourse-level information.

In the current experiment, both the lexical and the message level information were manipulated and thus expected to affect target N400 amplitude. At the lexical level, final words (such as *pretzels*) that were related to the bias sentence (…*salty*.) should be facilitated and elicit a smaller N400 compared to unrelated final words (*cookies*). At the same time, the presence or absence of negation changed the sentence- and discourse-based expectancies, as it determined the consistency (or truth) of the final word and sentence with the story. A final sentence ending in *pretzels*, for instance, was consistent with the bias sentence *He wanted something salty*. (11a) in its affirmative form (12i), but not if it was negative (12iii).

(11) *Affirmative bias*
  a. Joe wanted something salty.

(12) *Target sentence*
  i. So he bought the pretzels. (TA)
  ii. So he bought the cookies. (FA)
  iii. So he didn't buy the pretzels. (FN)
  iv. So he didn't buy the cookies. (TN)

The N400 to the final word should therefore not only depend on that word's relatedness to the bias sentence, but also on whether the sentence was affirmative or negative – if as we propose the negation had been integrated into the ongoing sentence representation. If our proposal is wrong, then our results should parallel Fischler's, with small N400 to related endings (TA and FN) and large N400s to unrelated endings (FA and TN). However, if neaation is incorporated into the sentence representation and affects expectations for upcoming words, as hypothesized, a different pattern should emerge: the smallest N400 should be observed for TA (12i) as they were both related to the bias and true, and the largest N400 should be observed for FA (12ii) as they were both unrelated to the bias and false. The two negative sentences should lead to intermediate N400 amplitudess, as each receives facilitation from either truth or relatedness but not both: FN (12iii) are related, but false, while TN (12iv) are unrelated, but true. The expected order of FN relative to TN depends on the relative strength of the truth and relatedness effects: If truth is more important than relatedness, then TN should elicit smaller N400s than FN; if relatedness has a stronger effect, then the opposite order should be observed.

In addition to the N400, a late positive component (LPC) was also expected to vary with sentence truth. The LPC is a type of P3, a domain-general component elicited by unexpected task-relevant stimuli (Duncan-Johnson & Donchin, 1977) that has been linked to event categorization (Kok, 2001) or the updating of working memory representations as a function of newly received information (Donchin & Coles, 1988). P3-like positivities to complex stimuli in higher cognitive tasks such as language processing are usually referred to as LPC or P600 (although their membership in the P3 family is contested if they are elicited by a syntactic manipulation: see Kutas, Federmeier, Staab, & Kluender, 2007). LPCs have been observed in response to semantic or pragmatic anomalies, following an N400 in some (albeit not all) cases (Faustmann, Murdoch, Finnigan, & Copland, 2005; Münte, Heinze, Matzke, Wieringa, & Johannes, 1998). In it thus not unreasonable to expect to observe a larger LPC to false compared to true sentences in the current experiment, in particular because of the use of a verification decision, which clearly made the truth of the sentence task relevant.

Last but not least, negation effects might be observed at the verb of the target sentence, which in the negative modality was preceded by the negative contraction *didn't*. If the negation marker was immediately integrated into the sentence representation and perhaps used to change expectations about possible sentence continuations, signs of these processes might be visible in the ERPs to any of the words following the negation. Lüdtke and colleagues (2008), for example, observed a sustained negativity on the word following the German negation marker *kein/e* (*no*) compared to the same words in the affirmative sentence version. Fischler et al. (1983) also observed a slight negativity toward the end of the ERP to the negative *is not* compared to the affirmative *is* frame, although the difference was not further analyzed. We thus planned to test for the presence of negation effects, namely a (sustained) negativity, in the ERP to the target sentence verb.

The sentence ERP data were complemented by a number of behavioral measures. Response times and accuracy were recorded to compare the result pattern with the ERP findings. Although not an absolute match to the N400 pattern, the RTs in Fischler et al. (1983) also showed a significant truth x negation interaction: TN led to longer RTs than FN, paralleling the N400 findings. If the current experimental setup does indeed reveal negation effects in the ERP, we would expect to find similar changes in the verification RTs. Accuracy was expected to be high overall as it was emphasized over speed in the instructions. If effects are found, they should parallel the RT pattern, as is typical for verification studies.

Finally, we administered a Stroop test (Stroop, 1935), a measure of inhibition or cognitive control, as well as two tests of print exposure, which correlate with linguistic ability (Stanovich, West, & Harrison, 1995). The purpose of these tests was to collect data on individual differences in overall cognitive and linguistic aptitude, which might help explain potential variability in the ERP and RT results, as this is not uncommon with more complex language processing tasks.

Method

Subjects

Thirty-two subjects (19 women) with a mean age of 20.1 years (range 18-24 years) participated for academic credit or a cash payment of $7/hour. All were right-handed native speakers of English with normal or corrected to normal vision and no history of neurological disorders.

Tests

Handedness was assessed with the Edinburgh handedness inventory (Oldfield, 1971). Version 4 of Stanovich and West's Author and Magazine Recognition Tests (Stanovich et al., 1995) was used to assess print exposure. For the Stroop test, two test sheets were created: one with colored strings of four Xs, one with color words. Each sheet contained 60 strings, arranged in four columns. Four ink colors – red, green, blue, and pink – were used, each fifteen times per sheet. The neutral version contained the strings of Xs printed in the different colors. In the interference condition, the same four color words appeared, always printed in an ink color that did not match the word. All word-ink combinations occurred equally often. See Appendix B for samples of all testing materials.

Stimuli

One-hundred-twenty scenarios such as Example (10-12), consisting of a two-sentence introduction, a bias, and a target sentence, were created. The introduction always remained the same, but there were two different versions of the bias and four versions of the target sentence. Each subject saw all 120 scenarios (with the same two introductory sentences), but different subjects saw different versions of the bias and target sentences. The two versions of the bias each referred to one of the two previously introduced options. One version of each scenario was assigned to one of two lists. For the target sentence, there were two affirmative and two negative versions, each

		During his long flight Joe needed a snack. The flight attendant could only offer him pretzels and cookies. Joe wanted something… salty.	sweet.
So he bought the…	pretzels.	True Affirmative (TA)	False Affirmative (FA)
	cookies.	False Affirmative (FA)	True Affirmative (TA)
So he didn't buy the…	pretzels.	False Negative (FN)	True Negative (TN)
	cookies.	True Negative (TN)	False Negative (FN)

Table 1. Sample stimuli. All bias-target combinations and resulting experimental sentence types are shown.

ending in a word related to one of the two different bias sentences. The resulting 480 target sentences were distributed over four lists in a counterbalanced fashion. Thus, half of the subjects were shown the first bias list combined with on of the four target lists, and the other subjects saw the second bias list with one of the target lists. The combination with the bias sentence determined the truth (or consistency) of the target sentence and, obviously, the relationship between target ending and bias sentence. So, endings that were true and related to the bias for one group of subjects were false and unrelated for the other group. Table 1 (above) demonstrates how combinations of bias and target versions result in the four different sentence types.

Procedure

Having given informed consent to participate in the study, subjects completed the Edinburgh handedness inventory (Oldfield, 1971) as well as the Author and Magazine Recognition Questionnaires (ART and MRT; Stanovich & Cunningham, 1992). Next, the Stroop test was administered: Subjects were first instructed to name the color of each string of letters on the first sheet as fast as possible, and the time to complete the sheet was recorded. They completed the interference condition in the same manner, after they were told to not read the color words but to name the color of the ink.

After the application of the electrodes to the head, subjects completed the experiment in a sound-proof, electrically shielded chamber. They were seated in a comfortable chair approximately 75 cm in front of a computer screen. Subjects were told that they would be reading short stories describing choices different people made. Their task was to decide whether the final sentence of the story was consistent with the information previously received. No timing instructions were given for the verification task or the self-paced reading of the introduction and bias sentences. Subjects were given a sample story and examples of consistent (true) and inconsistent (false) endings. The session began with a practice run of four scenarios, including one of each of the four target sentence types.

Each new trial was initiated by the subject's button press. After a 1000 ms blank screen, the two introductory sentences appeared together on the screen, where they remained until the next button press. Then the bias sentence was presented as a whole until the subject pressed a button. It was followed by a row of three crosses ("+++") to orient the subject's attention to the center of the screen. Following a 200 ms blank screen, the final sentence was presented word by word with a Stimulus Onset Asynchrony (SOA) of 500 ms and a word duration of 200 ms. Following the final word, the screen remained blank until 1000 ms after the subject had pressed a response button. The sentence "Please press a button to read the next story." was then shown until the subject initiated the next trial.

The trials were grouped into six blocks. After each block, subjects were encouraged to take a break. Usually, subjects completed a block in less than ten minutes, and the experiment rarely lasted more than an hour.

EEG Recording

The EEG was recorded from 26 tin electrodes geodesically arranged in an electrode-cap. The left mastoid served as reference. To control for blinks and horizontal eye movements, additional electrodes were on the outer canthi of the eyes (referenced to the left canthus) and on the right infraorbital ridge (referenced to the left mastoid). All impedances were kept below 5 kΩ. The EEG was bandpass filtered (0.01 – 100 Hz) and continuously digitized at a rate of 250 Hz.

Data Analysis

Accuracy

Data about response accuracy were submitted to a mixed-effects logistic regression with three main effects, trial number, truth, negation (including the interaction between truth and negation), as well as two random factors, Subject and Item. Trial number was included to reduce error variance, while truth and negation were the main factors of interest. In case of a significant truth x negation interaction, pairwise comparisons were carried out by running separate regression models with trial and Sentence Type as fixed effects. The p-values derived from these post-hoc models were adjusted for multiple comparisons using Hochberg's improved Bonferroni procedure (Hochberg, 1988).

Response Times

To improve the normality of their distribution, response times were logarithmically transformed (base 10). All statistics were performed on these log-transformed values. For easier comprehension, descriptive statistics were back-transformed (via exponentiation) for presentation in figures.

Trials with incorrect responses were excluded from all analyses. Furthermore, outliers were eliminated: Means and standard deviations were computed for each subject and sentence type, and data points whose distance from their corresponding mean was more than a certain number of standard deviations were rejected, with the cutoff depending on the number of valid trials for a given subject and condition (Van Selst & Jolicoeur, 1994). Approximately 4% of trials were excluded from the analyses.

The resulting data were analyzed with a mixed-effects model including trial, truth, negation, and the truth by negation interaction as fixed effects, as well as Subject and Item as random effects. A significant truth x negation interaction was resolved by performing pairwise comparisons. These comparisons were carried out as simultaneous hypothesis tests based on the normal approximation to the multivariate t-distribution (cf. Bretz, Hothorn, & Westfall, 2002; Westfall, Tobias, Rom, Wolfinger, & Hochberg, 1999). To correct for multiple comparisons, p-values were adjusted following Hochberg's (1988) method.

ERPs

EEG data were re-referenced off-line to the algebraic mean of the two mastoids. Trials contaminated by eye-movements, excessive muscle activity, or amplifier blocking were excluded. For ERPs to sentence-final targets, trials on which subjects made an incorrect verification decision were also excluded. Overall, 8% of trials were lost for target words, and 2% for verbs.

ERPs to the verb and the final word of the target sentence were computed by averaging epochs ranging from 100 ms before until 920 ms post word onset, after subtraction of a 100 ms pre-stimulus baseline. For ERPs to verbs, mean amplitudes were computed for a time-window ranging from 100 to 920 ms. For sentence-final target words, the time-windows were 150–200 ms (P2), 200–400 ms (N400), 400–600 ms (LPC), and 600–900 ms.

Mean amplitude values were submitted to repeated measures ANOVAs with truth, negation, and electrode as within-subjects factors. All reported p-values for effects with more than one degree of freedom (which was the case in interactions with the factor electrode) were adjusted using the Greenhouse-Geisser correction (Greenhouse & Geisser, 1959). The original degrees of freedom for the F-statistic are reported along with the adjustment factor ε. In the case of a significant interaction between electrode and another factor, four contrasts were computed to assess the shape of the distribution: We tested whether the effect differed between left and right, medial and lateral, frontal and posterior, as well as central and non-central electrode sites.

For pairwise comparisons of the four sentence types, data for each condition were averaged over all electrodes and submitted to t-tests. The derived p-values were adjusted for multiple comparisons (cf. Hochberg, 1988). In general, pairwise comparisons were carried out when a significant truth by negation interaction was found. They were always done for the N400 (200–400 ms), the dependent measure of primary interest.

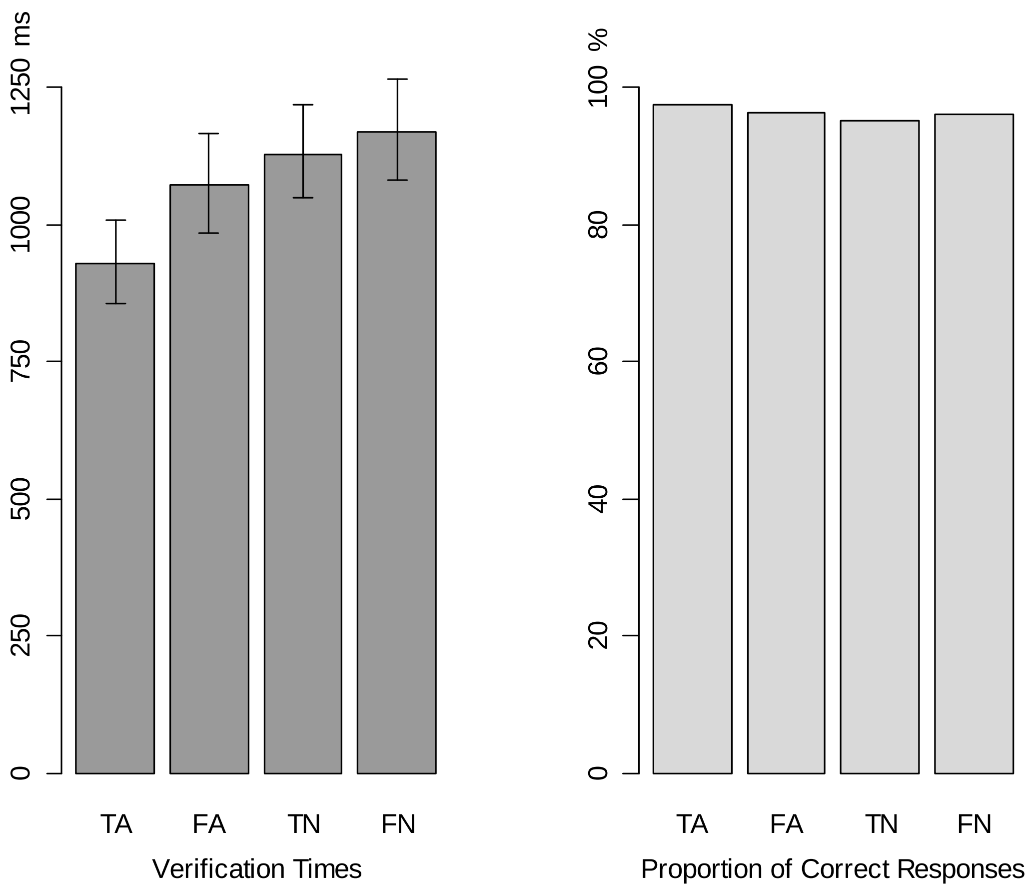

Figure 1: Verification results for Experiment 1. The left panel shows mean response times with 95% confidence intervals. Means and standard errors were computed on by-subject averages of log-transformed data. Back-transformed values are shown. The right panel shows the proportion of correct responses computed over all subjects and items.

Results for the Entire Sample

Subjects scored an average of .176 (*SD* = .065) on the ART and 0.300 (*SD* = .124) on the MRT. These values are notably lower than those reported for larger samples of students by Stanovich and Cunningham (1993, n = 268) and Stanovich et al. (1995, n = 133), who reported mean ART scores of .238 (*SD* = .145) and .327 (*SD* = .14), respectively, and average MRT scores of .486 (*SD* = .162) and .512 (*SD* = .15), respectively. This might indicate that the fourth versions of these culturally sensitive tests were somewhat outdated and therefore not appropriate for the college population that was tested in this experiment. Mean completion times for the Stroop were 36.2 seconds (*SD* = 5.9 s) on the neutral and 56.4 seconds (*SD* = 10.7 s) on the interference version, corresponding to an average interference cost of 56%.

Verification

Accuracy

Accuracy was high with a rate of 96% correct responses overall, and error rates decreased over the course of the experiment (Wald z = 3.02, p = .003). There was not much variability among the sentence types, which is apparent in Figure 1 (above). Neither truth (Wald z = -0.50, p = .519) nor negation (Wald z = 1.59, p = .112) had a significant effect on error rate. There was a marginally significant truth x negation interaction (Wald z = –1.89, p = .059). Post-hoc tests did not reveal any significant differences, although the comparison of TA and TN was significant before adjustment for multiple comparisons.

Response Times

Figure 1 shows RTs that subjects verified affirmative sentences faster than negative ones [$F(1, 3688) = 158.93$, $p < .001$], and true sentences faster than false ones [$F(1, 3688) = 64.06$, $p < .001$]. The significant truth x negation interaction [$F(1, 3688) = 22.47$, $p < .001$] indicated that the RT difference due to truth was larger for affirmative than negative sentences, but both were significant, as were all pairwise comparisons. Also, RTs on later trials were faster than those on earlier ones [$F(1, 3688) = 256.46$, $p < 0.001$].

Event-Related Brain Potentials

Verbs

Figure 2 (below) shows ERPs to the verbs in the final sentences of the stories. The plots indicate that verbs in negative sentences elicit more negative ERPs than verbs in affirmative sentences. Measured in a time-window from 100 to 900 ms, this effect was highly significant [$F(1, 31) = 58.99$, $p < .001$]. Its size differed across electrode sites [$F(25, 775; \varepsilon = .137) = 13.17$, $p < .001$]: the difference was larger on the left than on the right side of the head [$t(31) = -2.42$, $p = .021$], at medial compared to lateral locations [$t(31) = -3.81$, $p < .001$], as well as at central [$t(31) = -2.93$, $p = .006$] and frontal [$t(31) = 3.77$, $p < .001$] compared to non-central and posterior channels, respectively. This negation effect was not affected by the truth of the sentences [truth x negation: $F(1, 31) < 1$; truth x negation x electrode: $F(25, 775; \varepsilon = .188) = 1.99$, $p = .087$], and truth did not have an independent effect, either [main effect and interaction with electrode: both *F*s < 1]. This is expected, as the target word, which rendered the sentence true or false, occurred only after the end of the epoch for almost all sentences; in only seven out of 120 sentences did the target word immediately follow the verb and therefore affect the later part of the ERP.

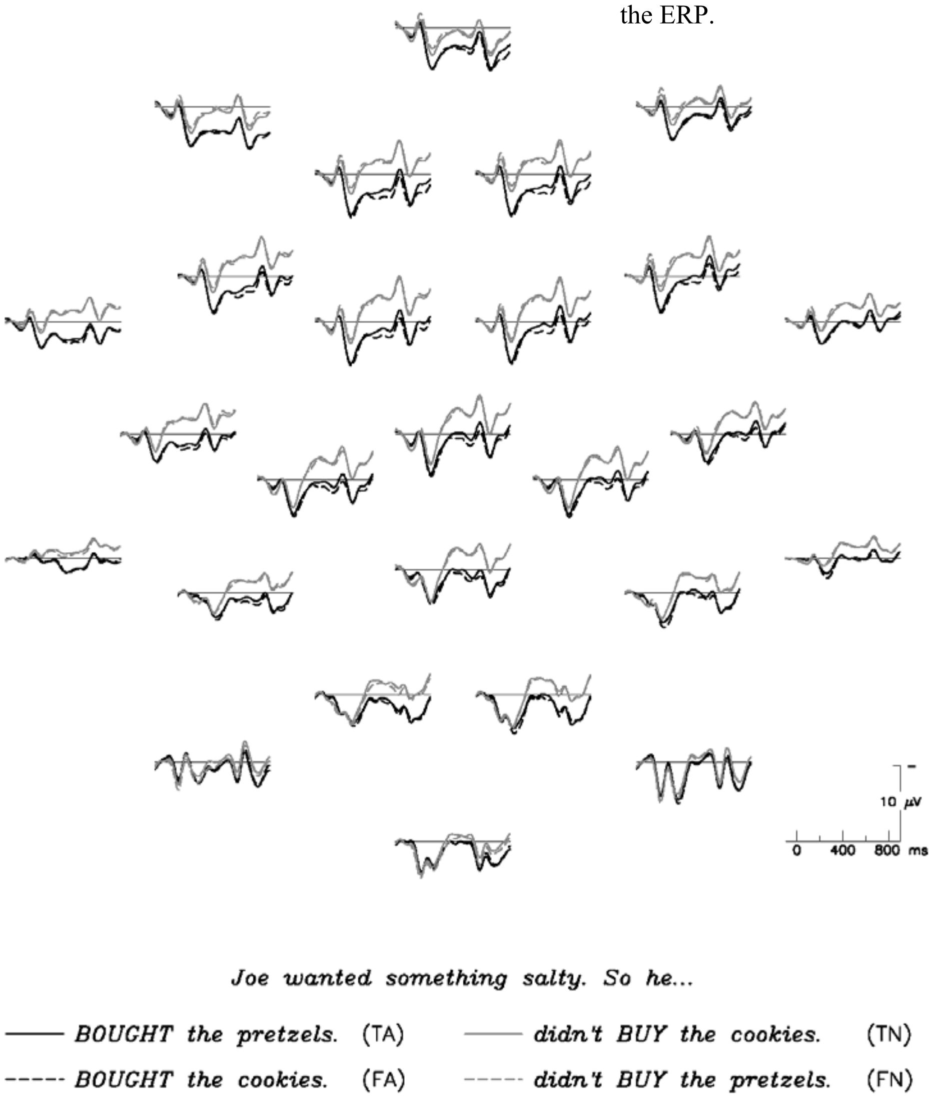

Figure 2. ERPs to final-sentence verbs in Experiment 1.

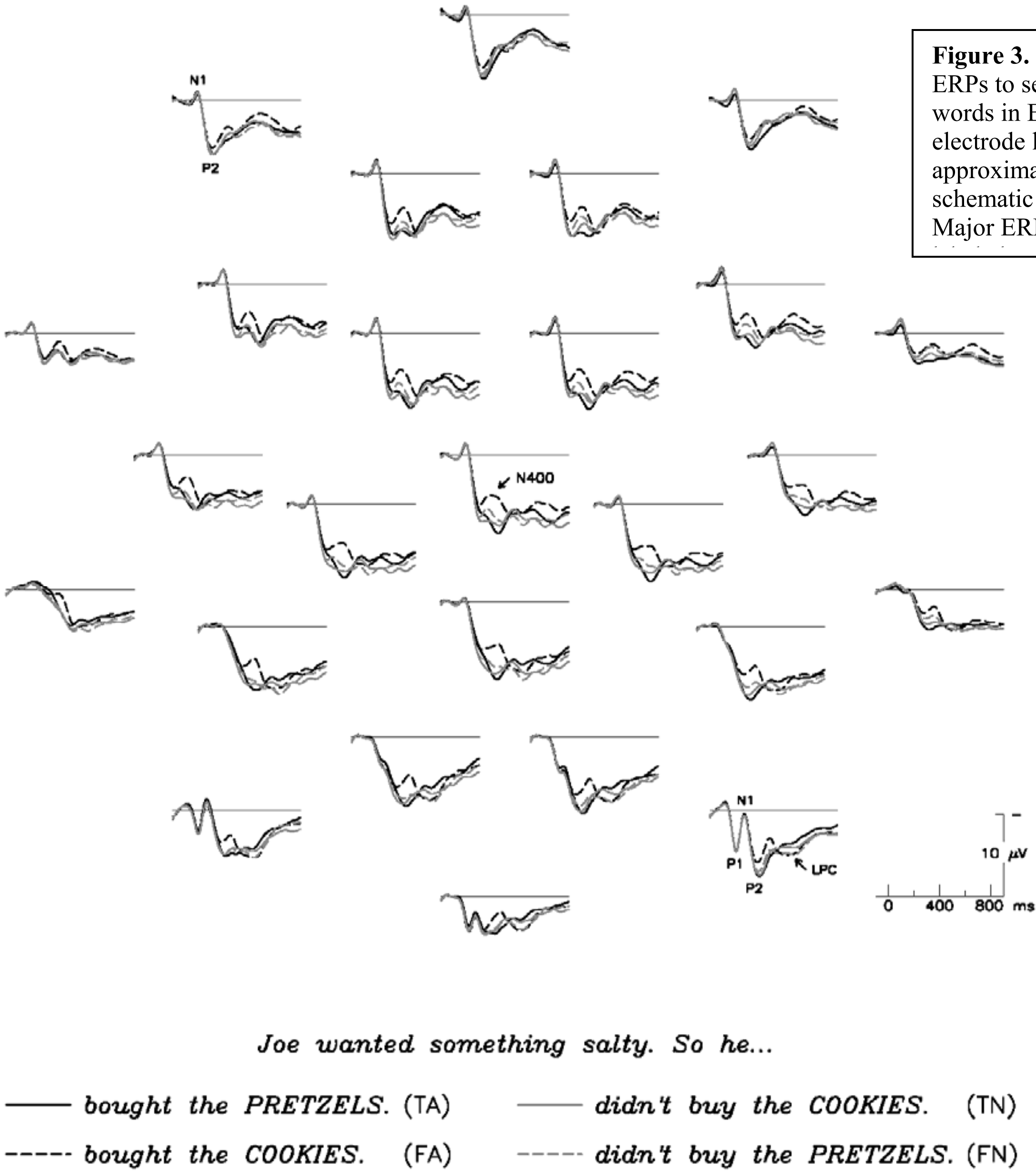

Figure 3. Grand average ERPs to sentence-final target words in Experiment 1. The electrode layout corresponds approximately to the schematic in Figure 3-1. Major ERP components are

Sentence-Final Targets

The grand average ERPs to sentence-final target words are presented in Figure 3 (above). Following early sensory components that were similar for all conditions, the ERPs diverged as a function of sentence type. At fronto-central sites, a uniform negativity peaking around 100 ms (N1) preceded a positivity with a peak around 220 ms (P2) that was larger in negative sentences than in affirmative ones. A posterior positivity and negativity peaking at approximately 100 and 170 ms, respectively (P1-N1 complex), were followed by a positive peak around 290 ms that showed some differentiation among ending types, possibly because of overlap with the following negativity. FA and to a lesser extent FN were associated with a negative going waveform (N400) that peaked around 300 ms at frontal and at approximately 360 ms at more posterior electrode sites. At posterior channels, false endings subsequently showed a positivity (late posterior component; LPC) between about 400 and 600 ms. After 600 ms, targets in affirmative sentence contexts elicited more negative ERPs at central electrodes than targets in negative sentences.

150 – 200 ms

ERPs to the targets showed effects of both truth and negation as early as 150 ms after word onset. Words in negative sentences elicited a larger positivity (P2) than those in affirmative ones [$F(1, 31) = 8.61$, $p = .006$]. The size of this effect varied as a function of electrode site [$F(25, 775;\ \varepsilon = .163) = 3.36$, $p < .001$], with a more pronounced positivity over central [$t(31) = 2.64$, $p = .013$] and medial [$t(31 = 3.03$, $p = .005$] scalp locations. Visually, it also appeared to be larger on the left, but the effect failed to reach significance [$t(31) = 1.67$, $p = .106$]. Truth did not have a significant main effect ($F(1, 31) < 1$], but its effect differed among electrode sites [$F(25, 775;\ \varepsilon = .154) = 1.86$, $p = .007$]: At right scalp locations only, false endings were associated with a larger negativity than true endings [$t(31) = –2\ .376$, $p = .024$], indicating that the onset of the N400 occurred already before 200 ms at these electrode sites. There were no interactions involving truth and negation [truth x negation and truth x negation x electrode: both *F*s < 1].

200 – 400 ms

Based on visual inspection of the grand average ERPs, it was decided to measure the N400 effect between 200 and 400 ms. This is a relatively early time-window for the N400 given that its peak usually occurs around 400 ms, the end of the time-window used in this study. Indeed the N400 in this data set peaked much earlier than in sentence reading studies, however, this is not unusual for a verification experiment. Fischler and colleagues found peak latencies of 320, 340, and 380 ms for N400s in verification studies (Fischler, Bloom, Childers, Arroyo, & Perry, 1984; Fischler et al., 1983; Fischler, Childers, Achariyapaopan, & Perry, 1985).

Statistical analyses revealed main effects of truth [$F(1, 31) = 33.50$, $p < .001$] and negation [$F(1, 31) = 11.09$, $p = .002$], and the two-way interactions with electrode were also significant for both factors [truth: $F(25, 775;\ \varepsilon = .187) = 7.36$, $p < .001$; negation: $F(25, 775;\ \varepsilon =.161) = 4.33$, $p = .002$]. Sentence-final words in negative sentences elicited more positive ERPs than affirmative sentence endings. This was probably due to overlap with the P2 increase for negative sentence targets, which carried over into the N400 time-window. Like the P2 effect, the positivity in the 200-400 ms time-window was larger at central [$t(31) = 3.22$, $p = .003$] and medial [$t(31) = 2.31$, $p = .028$] scalp locations. It was also greater at on the left than on the right [$t(31) = 3.19$, $p = .003$], which resembles the pattern found for the P2, although it did not reach significance there. The main effect of truth reflected the larger negativity associated with false endings compared to true ones. It was more pronounced at medial [$t(31) = 3.99$, $p < .001$] and central [$t(31) = 3.88$, $p < .001$] scalp locations, and it was larger on the right than on the left [$t(31) = -2.65$, $p = .013$].

The size of the truth effect differed between affirmative and negative sentences, which was reflected by the significant truth x negation interaction [$F(1,31) = 6.77$, $p = .014$] that was observed across the scalp [truth x negation x electrode: $F(25, 775;\ \varepsilon = .122) = 1.27$, $p = .289$]. Figure 4 (below) illustrates that the truth effect was larger in affirmative sentences than in negative ones. Indeed, pairwise comparisons revealed that it was significant only for affirmative sentences: FA elicited significantly larger N400s than TA, but N400s to FN and TN did not differ reliably. Post-hoc comparisons showed that FA were different from all other sentence types, which in turn did not differ significantly from each other. Note, however, that comparisons of the N400 to affirmative and negative sentence endings (e.g., TA vs. TN) are problematic and hard to interpret because of the spillover of the P2 difference between affirmative and negative endings into the N400 time window.

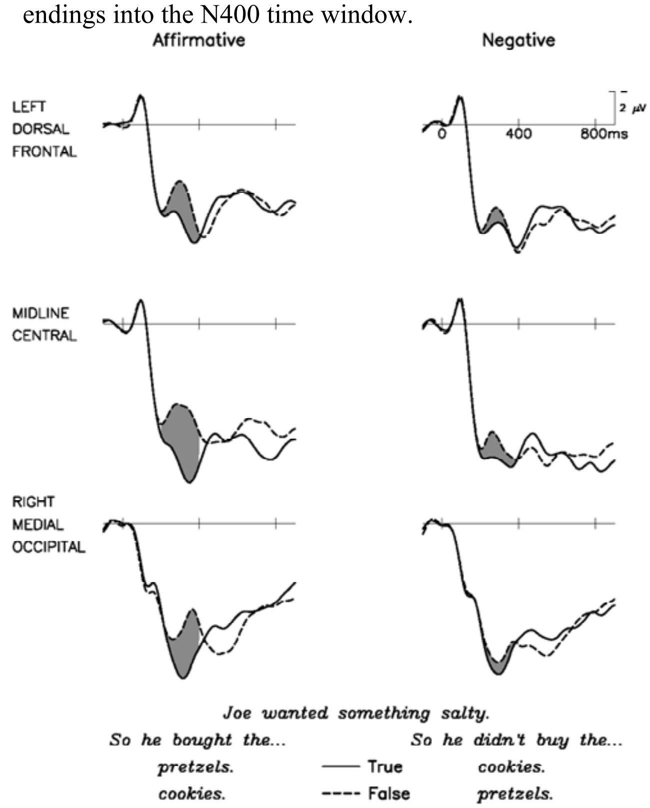

Figure 4. ERPs to true and false endings presented separately for affirmative and negative sentences (Experiment 1). The same three electrode sites are shown for the two sentence modes. The difference between true and false endings is shaded in N400 time-window (200-400

400 – 600 ms

In the time-window following the N400, ERPs to false sentence endings were more positive than those to true endings. This significant truth effect [$F(1, 31) = 4.41$, $p = .044$] varied in size across the scalp [$F(25, 775$; $\varepsilon = .215) = 2.22$, $p = 0.050$]: it was larger over the back than over the front of the head [$t(31) = -2.13$, $p = .041$]. Negation didn't have a reliable effect on mean voltage in this time-window [$F(1, 31) = 2.46$, $p = .127$; negation x electrode: $F(25, 775$; $\varepsilon = .129) = 1.28$, $p = .284$], and the size of the truth effect did not differ between affirmative and negative sentences [truth x negation and truth x negation x electrode: both Fs < 1]. Overall, its sensitivity to truth, a task-relevant factor, as well as its posterior distribution suggest that the positivity was an instance of the late posterior complex (LPC).

600 – 900 ms

Beginning at approximately 600 ms, the ERPs showed a prolonged divergence between affirmative and negative sentences [$F(1, 31) = 15.39$, $p < .001$] that varied in size across the scalp [$F(25, 775$; $\varepsilon = .198) = 3.25$, $p = .008$]. Endings in affirmative sentences were more negative (or less positive) than those in negative sentences, and the difference was most pronounced at central [$t(31) = 3.28$, $p = .003$] and medial [$t(31) = 2.31$, $p = .028$] scalp locations. This effect was not affected by the truth of the sentence [truth x negation: $F(1, 31) = 1.80$, $p = .190$; truth x negation x electrode: $F(25, 775$; $\varepsilon = .183) = 1.52$, $p = .193$], and truth itself did not have an independent effect, either [main effect and interaction with electrode: both Fs < 1].

Summary of Main Results

Subjects verified affirmative sentences faster than negative ones, and true sentences faster than false ones. The RT difference due to truth was larger in affirmative compared to negative sentences. The N400 showed a similar interaction between truth and negation: False sentence endings elicited larger negativities than true ones, but the effect reached significance only for affirmative sentences. Importantly, N400 amplitude was not simply determined by the lexical level relation between bias sentence and target word (as was the case in the Fischler et al., 1983 study), as the truth effect was clearly *not* reversed between affirmative and negative sentences: FN endings, which were related to the bias sentence, elicited N400s that were larger than or similar to the N400 to TN targets, which were not directly related to the content of the bias sentence. Thus, truth had at least as much of an effect on target N400 amplitude as the semantic relation between the target and previously mentioned words. Negation, which changes the truth of a sentence, must therefore have played some role in the processing of the final word of the target sentence.

Truth also affected the LPC, resulting in a larger positivity to false than to true endings. Additional effects distinguished affirmative and negative sentences. They were found on the target as well as on words preceding it. ERPs to negative targets started to show a positivity (P2 effect) around 150 ms after word onset, and beginning around 600 ms after stimulus onset, they were again less negative (i.e. more positive) than those to affirmative targets. Preceding the target, sentence segments following a negation marker were associated with more negative going ERPs than the same words in affirmative contexts.

Subject Groups

The average results presented for the entire subject sample hide an important fact: There was great inter-individual variability, most notably in the N400 data. In some subjects, the N400 truth effect was the same for affirmative and negative sentences, while it was completely reversed in others. In order to explore possible reasons for these diverging data patterns, subjects were sorted into two groups and then compared on a number of measures. One group ('FN > TN', $n = 19$) contained all subjects in whom the N400 to FN was larger than the N400 to TN at the midline central electrode (MiCe). The second grouped contained the remaining subjects ('FN ≤ TN', $n = 13$), in whom the N400 to FN were smaller than or similar to the TN N400.

In a first step, these two groups were compared with respect to all ERP and RT measures that were previously analyzed for the whole sample. Since the N400 data were the basis for the categorization, the groups should certainly differ in their N400 patterns. Given the similarity of RT and N400 data for the whole sample, the groups could also be expected to show divergent RT patterns: Subject with larger N400s to FN should also have longer RTs to FN (compared to TN), and subjects with smaller N400s to FN should also have taken less time to verify FN. No particular hypotheses were formulated for the analyses of the remaining mean amplitude measures as well as accuracy; these were done mainly for exploratory purposes.

In a second step, hypotheses about potential reasons for the diverging N400 patterns were tested. One hypothesis was that the groups differed in language or cognitive abilities. Subjects with higher linguistic ability or better cognitive control might be faster at integrating negation into the sentence context and updating their expectations about the continuation of the sentence. So they should show an advantage (smaller N400) for TN over FN. The groups were therefore compared on their ART, MRT, and Stroop scores.

Another possible explanation is that subjects differed in the way they processed information prior to the target sentence, namely the bias sentence. The inferences derived from the bias sentence are the basis for the expectations about the final sentence. A subject might, for instance, learn that *Joe wanted something salty*. Knowing that he had a choice between pretzels and cookies, she could directly infer that Joe would buy pretzels. Yet, another, less direct conclusion is warranted by the information, namely that Joe would not buy cookies. Experimental evidence suggests that subjects usually draw this kind of logical inference when reading narratives. The result of the inference, however, does not necessarily become or remain activated (Lea & Mulligan, 2002), probably because it is negated and therefore backgrounded. In the context of this experiment, keeping this negative inference active in working memory would be beneficial as it facilitates the verification of negative sentences, and some subjects may therefore have made an effort to do this. Others may not have employed this strategy, and as a result the negative inference would not be (as) available to them during the processing of the subsequent target sentence. These subjects would then not be able to predict the 'true' ending for a negative target sentence, which should lead to a disadvantage (larger N400) for TN compared to FN, whose ending is related to the bias sentence and corresponds to the automatic affirmative inference. Differences in processing the bias sentence may manifest themselves in the time subjects took to process the sentence as well as in the ERP. The 'FN > TN' group might show longer reading times for the bias sentence than the 'FN ≤ TN' group, reflecting the extra time for activating the negative inference. No specific prediction is proposed for the ERPs to the bias sentence; the goal was simply to detect any (*a priori* undefined) difference that might help explain the group differences in N400 patterns.

Data Analysis

Verification and ERPs to Verbs and Sentence-Final Targets

For response times, accuracy, and mean amplitude data, the analyses conducted on the whole sample were repeated with the additional factor group. When a significant interaction involving group was found, separate analyses were carried out for the two subject groups.

Cognitive Tests

Using the Wilcoxon rank-sum test, the groups were compared with respect to their ART, and MRT scores, their times on the neutral and interference Stroop, as well as the relative interference cost on the Stroop test [computed as (interference – neutral)/neutral].

Reading Times for Bias Sentence

Bias sentence reading times were analyzed through a mixed-effects analysis including group, as well as the potentially confounding factors trial and length (in number of words) as fixed effects and Subject and Item as random effects.

Event-Related Potentials to Bias Sentence

In addition to the verb and target word averages, ERPs to the bias sentence were computed for epochs from 200 ms before until 1840 ms after sentence onset, with a 200 ms pre-stimulus baseline. As subjects had been reading entire sentences on the screen (which necessarily involves saccadic eye movements) and had not been instructed to suppress blinks while reading these sentences, artifact rejection would have eliminated every single trial. These ERPs were therefore based on all trials, irrespective of the artifacts that they possibly (and likely) contained. Mean amplitudes were computed for a time-window from sentence-onset to 700 ms after sentence onset. The resulting values were submitted to an ANOVA with group as between-subject and electrode as repeated-measures factor.

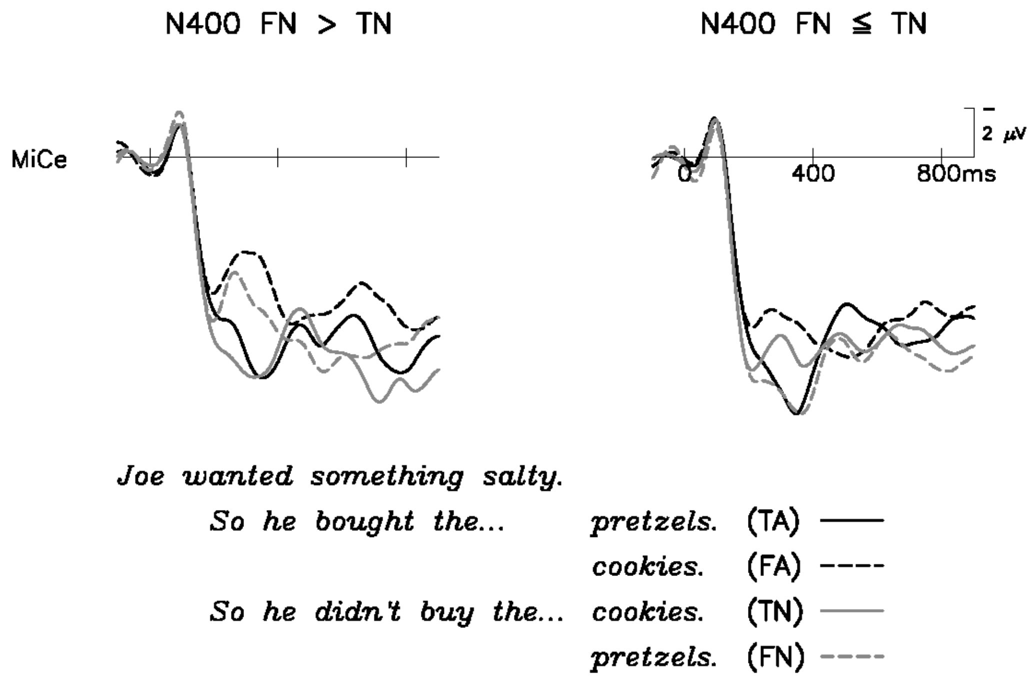

Figure 5. ERPs to sentence-final targets for the two subject groups in Experiment 1. The midline central channel (MiCe) is plotted for both groups.

Event-Related Potentials to Target-Sentence Words

N400 (200 – 400 ms)

Figure 5 (above) shows the ERPs to sentence-final targets for the two subject groups side by side. The groups differed in the ordering of the N400s to FN and TN, which is expected, since the N400 to negative sentence endings was the categorization criterion. The negation effect was similar for both groups across the scalp [group x negation and group x negation x electrode: both Fs < 1]. Group interacted, however, with truth [$F(1, 30) = 13.18$, $p = .001$] as well as with truth and negation [$F(1, 30) = 6.21$, $p = .018$].

In the 'FN > TN' group, the truth effect was significant [$F(1, 18) = 62.52$, $p < .001$], while truth x negation interaction was not [$F(1,18) < 1$]. That is, the truth effect did not differ between affirmative and negative sentences. Both types of false sentence endings elicited larger N400s than both true ending types, but the difference between TA and TN and that between FA and FN did not reach significance.

By contrast, in the 'FN ≤ TN' group, the truth effect was reversed between affirmative and negative sentences. FA elicited significantly larger negativities than TA, but numerically smaller N400s were observed for FN than for TN, although this difference failed to reach significance after adjustment for multiple comparisons. Correspondingly, the truth x negation interaction was significant [$F(1, 12) = 11.65$, $p = .005$], while the main effect of truth was not [$F(1, 12) = 2.16$, $p = .167$].

The distribution of the truth effects also differed between groups [$F(25, 750$; $\varepsilon = .207) = 3.86$, $p = .002$], but the distribution of the truth x negation interaction did not [$F(25, 750$; $\varepsilon = .124) = 1.88$, $p = .137$]. In the 'FN > TN' group, where the truth main effect was significant, its size also varied across the scalp [$F(25, 450$; $\varepsilon=.205) = 10.44$, $p < .001$]. It was larger at right [$t(18) = -3.04$, $p = .007$], medial [$t(18) = 4.59$, $p < .001$], and central [$t(18) = 4.31$, $p < .001$] electrodes. The truth effect in the 'FN ≤ TN' group, which was not significant overall, also did not interact with electrode site [$F(25, 300$; $\varepsilon = .145) = 1.20$, $p = .323$]. Basically, the truth effect varied in size across the scalp only when it was actually present.

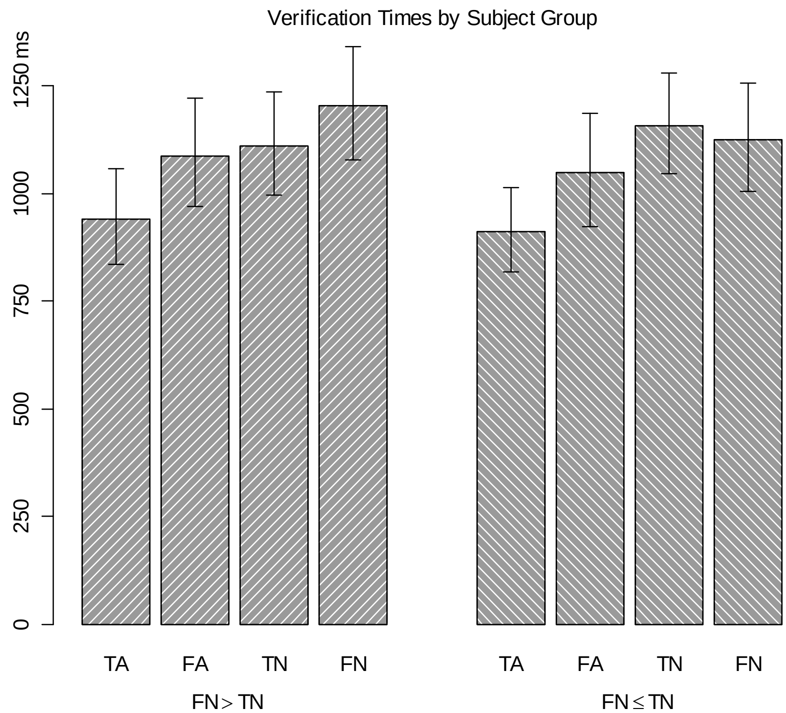

Figure 6. Verification times for the two subject groups in Experiment 1. Error bars represent 95% confidence intervals. Means and standard errors were computed on by-subject averages of log-transformed data. Back-transformed values are shown. The group of subjects with larger N400s to FN compared to TN is shown on the left, the group with the opposite pattern on the right.

The truth effect within affirmatives, which was significant in both groups also showed similar distributions [F(25, 750; $\varepsilon = .188$) < 1]. Likewise, the distribution of the reversed truth effect for negatives in the 'FN ≤ TN' group did not differ significantly from the "regular" truth effect in the 'FN > TN' group [$F(25, 750$; $\varepsilon = .268) = 1.12$, $p = .351$]. So the truth effect distributions were actually similar between the groups. It was only the averaging of a normal with a reversed truth effect in the 'FN ≤ TN' group that eliminated both the truth main effect and its interaction with electrode site.

Other ERP Effects

No interactions involving group were found for the verb and for the sentence-final words in the 150-200 and the 600-900 ms time-windows [all *p*s > .1]. For the LPC (400-600 ms), only the four-way interaction between group, truth, negation, and electrode was significant [$F(25, 750$; $\varepsilon = .149) = 3.21$, $p = .018$; all other *p*s > .1], but no significant effects were found in either group. In the 'TN > FN' group, a non-significant trend toward a smaller truth effect at more frontal channels for affirmative sentences only was observed [truth x negation x electrode: $F(25, 450$; $\varepsilon = .144) = 2.22$, $p = .083$; anteriority contrast: $t(18) = -1.90$, $p = .073$]. For the 'TN < FN' group, there was a non-significant trend toward a truth x electrode interaction [$F(25, 300$; $\varepsilon = .170) = 2.08$, $p = .094$], due to a marginally reduced truth effect at lateral [$t(12) = -1.85$, $p = .089$] and frontal [$t(12) = -2.15$, $p = .053$] channels.

Verification

Error rates were independent of group membership [all *p*s > .1]. Response time patterns, by contrast, varied between groups. As for the N400, group interacted with truth [$F(1, 3684) = 6.361$, $p = .012$] as well as with truth x negation [$F(1, 3684) = 4.91$, $p = .027$], while the negation effect was independent of group [F(1, 3684) < 1]. The main effect of truth [FN > TN: $F(1, 2183) = 61.19$, $p < .001$; FN ≤ TN: $F(1, 1500) = 8.80$, $p = .003$] and the truth x negation interaction [FN > TN: $F(1, 2183) = 4.962$, $p < .026$; FN ≤ TN: $F(1, 1500) = 20.88$, $p < .001$] were significant for both groups. However, the relationship between TN and FN differed between the groups, as Figure 6 (above) shows. For the 'FN > TN' group, the interaction was ordinal: The truth effect in negative sentences was smaller than in affirmative sentences, but it pointed in the same direction. That is, in both affirmatives and negatives, false sentences were verified significantly more slowly than true ones By contrast, a crossover interaction was observed for the 'FN ≤ TN' group. As with the N400 for this group, the truth effect was, at least numerically, reversed between affirmatives and negatives. RTs were significantly longer to FA than to TA, and while FN and TN did not differ significantly, RTs appeared longer for FN. Overall, the group differences in RT patterns closely matched the N400 variations.

Cognitive Tests

Table 2 (below) shows the results of the group comparisons on the ART, MRT, neutral and interference Stroop times, as well as interference cost. No group differences were found to be reliable or even to approach significance.

		W	*p*
Print Exposure	ART	136.0	.640
	MRT	122.5	.977
Stroop	Neutral Time	127.5	.887
	Interference Time	88.5	.185
	Interference Cost	91.0	.219

Table 2. Results of the group comparisons on test scores. The Wilcoxon rank-sum statistic *W* and the exact *p*-value (corrected for the presence of ties) are reported.

Bias Sentence

Reading Times

Reading times for the bias sentence decreased over the course of the experiment [$F(1, 3832) = 337.23$, $p < .001$], but did not depend on the length of the sentences [$F(1, 3832) < 1$]. Importantly, group had a marginal effect [$F(1, 3832) = 3.23$, $p = .073$], with subjects in the 'FN > TN' group tending to take more time to read the bias (3098 ms) sentences than subjects in the 'FN ≤ TN' group (2651 ms).

Event-Related Potentials

Figure 7 (next page) shows the ERPs to the first 1600 ms[1] of the bias sentence. These ERPs looked quite different from ERPs time-locked to single words, as they were recorded while subjects read sentences presented as a whole on the screen. Except for the posterior P1, most sensory components, i.e. the frontal and posterior N1 and P2, appeared to be present however. After first positive peak, the ERPs grew increasingly negative, reaching a maximum around 400 ms at frontal, 500 ms at central, and 800 ms post sentence-onset at posterior sites. This maximum negativity was then followed by a positive-going trend that lasted almost until the end of the epoch (1800 ms), leveling off earlier at frontal than at posterior locations.

Beginning immediately at sentence-onset, the ERPs for the two groups started to diverge, with more negative-going waveforms in the 'FN > TN' group compared to the 'FN ≤ TN' group. In the 0-700 ms time-window, this effect was significant [$F(1, 30) = 4.20$, $p = 0.049$], and it was present across the scalp [group x electrode: $F(25, 750; \varepsilon = .069) = 1.35$, $p = .266$].

Summary of Group Comparisons

In line with the sorting criterion, the two subject groups differed with respect to their N400 patterns: A main effect of truth but no truth x negation interaction was observed for the 'FN>TN' group; i.e., the same truth effect was found for affirmatives and negatives. In the 'FN≤TN' group, by contrast, the truth effect was reversed between affirmative and negative sentences. The RT results paralleled these N400 differences: RTs to FN were longer than those to TN in the 'FN>TN' group, while the opposite pattern was found in the 'FN≤TN' group. As for the overall sample, N400 and RT results went together.

No significant group differences were observed for any of the other ERP measures or Stroop and print exposure scores. The ERPs to the bias sentence, however, differed significantly between the groups, with larger negativities found in the 'FN>TN' group, who also tended to take more time to read the bias sentences.

[1] Approximately 10% of reading times were shorter than 1600 ms, that is, the subject had already stopped reading the sentence by the end of the epoch. A longer epoch would have contained an even higher percentage of such trials (15% at 1800 ms, 22% at 2000 ms, etc).

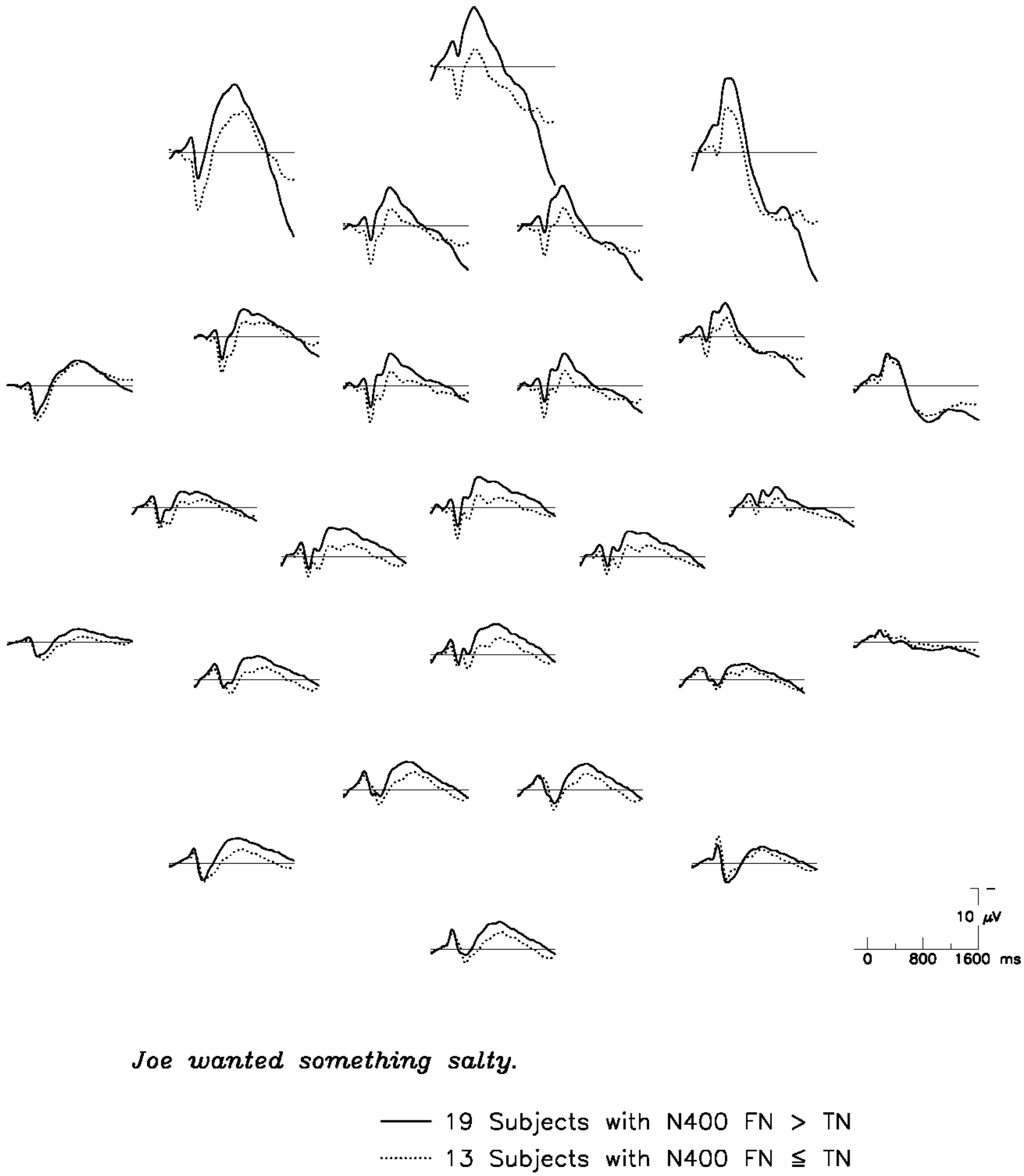

Figure 7. ERPs to bias sentences for Experiment 1. Waveforms for the two subjects groups are plotted on top of each other.

Discussion

If negation was processed only after the embedded affirmative proposition, as Fischler et al. (1983) suggested, N400 amplitude to the sentence-final target should have depended strictly on the relationship between target word and bias sentence, impervious to negation. That is, we should have observed no main effect of truth, but only a truth x negation interaction, with FA eliciting larger N400s than TA, but FN leading to smaller N400s than TN. Both related words (TA and FN) would be associated with smaller N400 amplitudes than both unrelated words (FA and TN). This is the pattern that Fischler and colleagues observed with isolated affirmative and negative sentences, such as *A robin is (not) a bird/vehicle*.

In the present experiment, however, when target sentences were embedded in choice scenarios, thereby allowing for the prediction of plausible and different sentence endings in both affirmative and negative sentences, a very different data pattern emerged: Overall, true sentence endings elicited smaller N400s than false ones, and while the truth effect failed to reach significance for negatives, it was the same direction as for affirmatives. The effect of truth thus dominated over the effect of relatedness with respect to N400 amplitude. Truth, in turn, was dependent on the presence or absence of negation, supporting the conclusion that negation must already have played a role in the processing of the target word as reflected in the N400. Negation must have been integrated in the representation of the target sentence *before* the processing of the embedded proposition – which included the final word – was completed. The hypothesis that pragmatically licensed negation can affect sentence processing early (i.e, almost immediately) by virtue of its modulation of word expectancies – thus was corroborated in the observed pattern of N400 amplitudes. Our findings are in general accord with those of Nieuwland and Kuperberg (2008) who also observed intra-sentential effects of pragmatically-licensed negation on N400 amplitudes, but stopped short of explaining how pragmatic licensing might work. We propose that in their study as in ours pragmatic licensing exerts it influence via expectancy generation. In our experiment, verification times were also consistent with the N400 data, as true sentences had an RT advantage over false ones, whether affirmative or negative.

A closer look at individual data patterns, however, reveals significant variability in the target N400 as well as in verification times for negative sentences: While the same N400 truth effect is found in affirmatives and negatives (i.e. no effect of relatedness whatsoever) in some participants, others show a pattern dominated by relatedness and not truth, with larger N400s to TN than FN. Furthermore, the subject group with a smaller N400 to TN compared to FN shows the same pattern for RTs, while the group with the reversed N400 ordering likewise is characterized by the reversed pattern for verification RTs.

One possible explanation for these different patterns is that one group of participants did indeed integrate negation on-line, as soon as it was encountered, while the other group did not, being characterized by delayed processing of negation, instead. For example, participants with higher cognitive capacity or linguistic skills might find it easier than those with lower abilities to process negation immediately. A comparison of the two groups on a number of test scores, however, did not reveal any reliable group differences in this regard, thus failing to support this hypothesis. Of course, it may be that the tests we administered simply did not measure the relevant abilities. Besides, the difference in negation processing might be a strategy that is not correlated with skill. We therefore cannot rule out the possibility that different participants may integrate negation at different time points during sentence processing.

We did, however, find support for one alternative hypothesis according to which participants differed in the way they processed the bias sentence, which provided the information necessary to evaluate the truth of the target sentence. While all participants presumably inferred the affirmative choice from the bias sentence, not all of them also may have committed the negative inference to memory. Those that did not might not have been able to (quickly) update their expectations about the target sentence ending upon encountering the negation marker. If so, then like Fischler et al.'s participants (1983), they would have no readily available alternative to plausibly complete a negative sentence.

In line with this hypothesis, the ERPs to the bias sentence did differentiate the two participant groups. While the functional interpretation of the sustained voltage difference remains unknown, the significant ERP effect points to some kind of processing difference at this critical point of processing. Furthermore, participants in the group with consistent truth effects for affirmatives and negatives tended to take more time to read and process the bias sentence, perhaps reflecting the increased effort associated with keeping the negative (as well as affirmative) inference active in working memory. These data are consistent with the possibility that not only the information contained directly in the bias sentence but the inferences drawn from it play an instrumental role in the processing of the upcoming target sentence.

Additional ERP effects offer evidence for early effects of negation on subsequent sentence processing. Most notably, the sentence fragment following the negation marker was characterized by significantly more negative going potentials than the same fragment in the affirmative sentence context. At minimum, this indicates that negation must have been registered, affecting subsequent processing to some degree. Its distribution likens this negation effect to the sustained left anterior negativity (LAN) observed in response to sentence fragments argued to place higher demands on working memory resources (Felser, Clahsen, & Münte, 2003; Fiebach, Schlesewsky, & Friederici, 2001; King & Kutas, 1995; Münte, Schlitz, & Kutas, 1998). If this an apt equivalence, then the negativity in the current experiment might reflect an increased working memory load associated with the processing of the negation marker, due for example, to the retrieval of the expected ending for negatives. The outcome of such a negative inference, however, may be less readily available than the affirmative inference, as negative information usually is backgrounded if not suppressed (cf. section 2.2.2).

Preceding the N400, the P2 component to the target word also showed some sensitivity to negation, as target words in negative sentences elicited larger positivities than those in affirmative sentences. While this finding supports the idea that negation affected the processing of the target word, its functional interpretation is unclear. P2 effects have been linked to the matching of visual features, with larger P2s to stimuli containing the target feature (Luck & Hillyard, 1994). In the current experiment, however, the P2 varied only as a function of the context in which the word appeared; it did not depend on any property of the target word or its relationship to the context. A different functional interpretation of the P2 – according to which the P2 reflects the selection of a target and/or the suppression of distractors (Bles, Alink, & Jansma, 2007; Melara, Rao, & Tong, 2002) – may be applicable. On this account, the more competition that exists, the more difficult the selection process, with increased need for suppression – as reflected in increased P2 amplitudes. For the current experiment, an argument can be made that participants had weaker expectations about the ending for negative sentences. Accordingly, the target word had to be selected out of a larger set of candidates, and thus was subject to more competition. The absences of a reliable P2 difference between correct and incorrect targets is consistent with this idea, as both had to be selected from the same set of candidates (equivalent number of competitors).

ERP effects post-N400 did not directly bear on the main hypothesis of the current experiment, because such late components are usually thought to reflect conscious decision-making, rather than online sentence processing. Even so, the pattern of results was at odds with a related claim by Fischler and colleagues (1983): They observed more positive ERPs to negative sentence endings 700 ms post target onset and interpreted this as a delayed late positive component (LPC) to negative sentences, due to a second stage of negation processing following the initial processing of the embedded proposition. The current experiment, however, revealed a much earlier LPC (between 400 and 600 ms), which was sensitive to sentence truth, followed by a slow wave that, like in Fischler et al. (1983), was more positive for negative sentence endings. Given that the correct decisions were reflected in the earlier LPC, it is unlikely that the slow wave in this experiment was an index of a second stage for invoke by the need to process negation. Together with the other ERP and verification data, these findings strengthen the case against a delayed processing of negation.

Experiment 2: Whole-Sentence Verification

The traditional approach to studying negation is based on a timed sentence-picture verification task. An affirmative or negative sentence is presented before, simultaneously with, or after a picture, and the subject is asked to decide whether the sentence is consistent with the picture. In the choice scenarios, the bias sentence together with the introduction fulfills the same function as the picture in the classic paradigm; it serves as the background against which the target sentence is verified. The current experimental paradigm, in which the target is the final sentence, is thus equivalent to presenting the picture before the sentence. This kind of situation has usually produced the RT pattern predicted by Clark's (1976) 'True' model: TA are verified faster than FA, but TN lead to longer RTs than FN. In Experiment 1, however, a different pattern was found, with shorter RTs to true than to false sentences, for *both* affirmatives and negatives. There are two possible explanations for this discrepancy. One option is that the choice scenarios were more likely than pictures to make subjects anticipate the correct affirmative *and* negative sentence completion. The other possibility is that the way in which the final sentence was presented played an important role: In classic verification studies, the whole target sentence appeared at once, while it was presented word-by-word in Experiment 1.

In the current experiment, subjects saw the same choice scenarios as in Experiment 1, but the target sentence was presented as a whole. This experimental setup allowed us to determine whether the change in RT pattern was due to the stimuli alone, or whether the mode of presentation made a difference. If the appearance of a normal truth effect in negatives depended on word-by-word presentation, one would expect the classic RT pattern with a reversed truth effect for negatives if the same sentences were presented as a whole (i.e., in the present experiment). The same truth effect should be found in both affirmatives and negatives, however, if presentation mode did not matter.

Method

Subjects

Sixteen subjects (13 women) with a mean age of 20.8 years (range 18-26 years) participated for academic credit or a cash payment of $7/hour. All were right-handed native speakers of English with normal or corrected to normal vision and no history of neurological disorders.

Design and Materials

The same materials as in Experiment 1 were used.

Procedure

Subjects completed the ART, MRT, and Stroop as described for Experiment 1. During the following experiment, they were seated approximately 50 cm from a computer screen. The instructions were to read the first two screens (containing the introduction and bias sentences) at their own pace, and then to decide as quickly as possible whether the final sentence was consistent with respect to the information in the preceding sentences. Before reading the experimental stories, subjects completed four practice trials.

At the beginning of each trial, a row of ten crosses was presented for 1000 ms. After a 200 ms blank screen, the introductory sentences were shown until the subject pressed a button. Then the introduction disappeared, and 200 ms later the bias sentence was presented, again until a button press by the subject. Two-hundred milliseconds later, a question mark appeared to signal the subject that she had to make a decision on the subsequent sentence. The question mark stayed on for 1000 ms, and after a 200 ms break, the target sentence was presented. It remained on the screen until the subject pressed a button to make the *True*/*False* decision. A new trial started after 1200 ms. The experiment consisted of six blocks. Subjects were told to take breaks between blocks if they felt the need. Typically, a subject completed the experiment in 25 to 40 minutes.

Data Analysis

The analysis of accuracy and response time data followed the same overall strategy as in Experiment 1, with a data loss of about 10% due to elimination of outliers and incorrect responses. Since the final sentence was presented as a whole in this experiment, verification time was confounded with basic reading time, which is affected by sentence length. Negative sentences, which were one to two syllables longer than affirmative ones, should take longer to read and verify simply because of this difference in length. Length (in syllables) was therefore added as a predictor to the mixed effects analysis of RTs, which allowed for the assessment of negation effects after controlling for the number of syllables in a sentence.

Results

With mean values of .181 ($SD = .097$) for the ART and .259 ($SD = .155$) for the MRT, print exposure scores in this study were equally low as those in Experiment 1. Stroop scores were also very similar to those in the previous experiment: Subjects took 36.2 seconds ($SD = 6.1$ s), on average, to complete the neutral version and 55.6 seconds ($SD = 10.1$ s) for the interference version, corresponding to a 55% increase.

Accuracy

The overall percentage of correct responses was 92%. Thus, accuracy was still very high in this whole-sentence paradigm, although it was slightly lower than in Experiment 1, where the final sentence was presented word-by-word. The left panel of Figure 8 shows the rates of correct responses by sentence type. These rates did not increase or

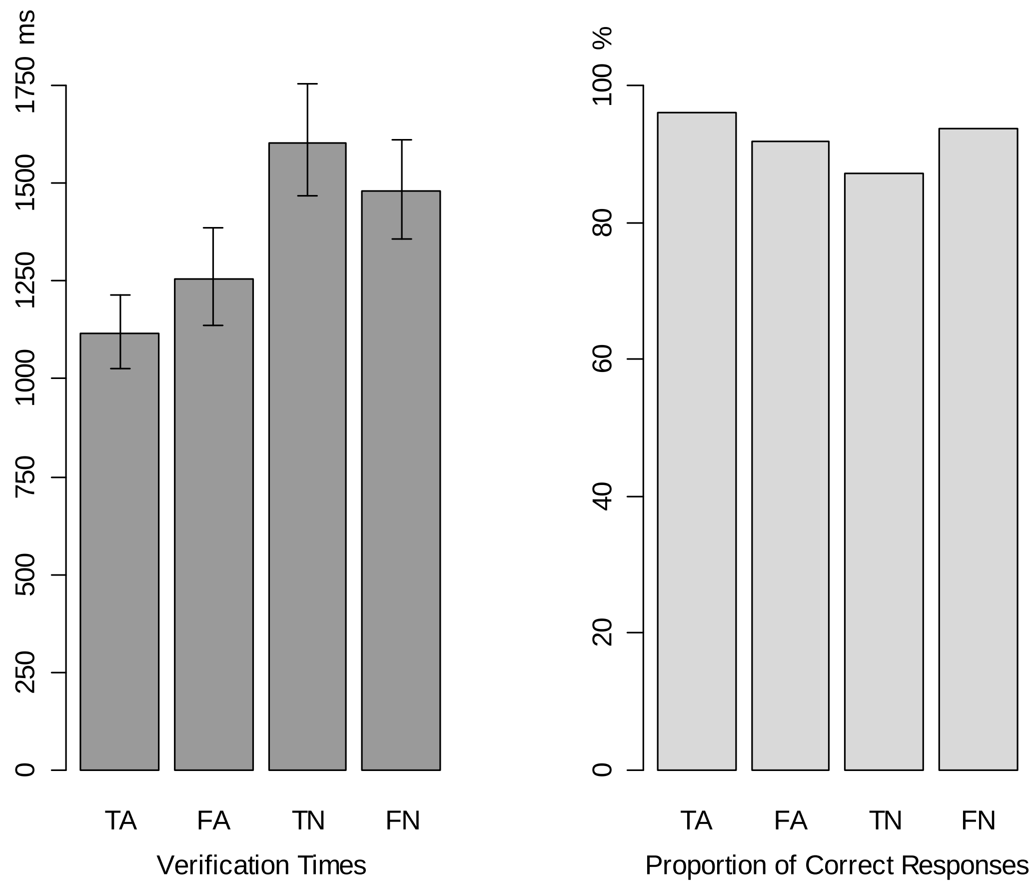

Figure 8. Verification times and accuracy in Experiment 2. The left panel shows mean response times with 95% confidence intervals. Means and standard errors were computed on by-subject averages of log-transformed data. Back-transformed values are shown. The right panel shows the proportion of correct responses computed over all subjects and items.

decrease linearly over the course of the experiment [Wald $z = 0.43$, $p = .656$]. Negation did not have a significant main effect [Wald $z = 1.17$, $p = .242$]. Truth did [Wald $z = 2.71$, $p = .007$], but there was also a significant truth x negation interaction [Wald $z = -4.47$, $p < .001$], indicating that the truth effect differed between affirmative and negative sentences. TA were verified more accurately than FA, but for TN the rate of correct responses was lower than for FN .

Response Times

The right panel of Figure 8 (above) presents descriptive statistics shows RTs by sentence type. Response times decreased over the course of the experiment [$F(1, 1727) = 81.35$, $p < .001$], and they increased with sentence length [$F(1, 1727) = 445.40$, $p < .001$]. Negative sentences took longer to verify than affirmative ones [$F(1, 1727) = 155.05$, $p < .001$]. There was also a main effect of truth [$F(1, 1727) = 4.44$, $p = .035$], but the significant truth x negation interaction [$F(1, 1727) = 68.90$, $p < .001$] indicated that the truth effect differed between affirmative and negative sentences: TA were verified faster than FA, but TN led to significantly longer RTs than FN. This reversed truth effect for negatives was observed in 12 out of 16 subjects.

Discussion

Experiment 2 produced a pattern of verification times that largely conformed to the findings of traditional picture-sentence verification studies and the predictions of the 'True' method (Clark, 1976): Affirmative sentences were verified faster than negative ones, and the two sentence types brought about opposite truth effects, with shorter RTs to TA than FA, but longer RTs to TN compared to FN. This outcome is markedly different from the RT results of Experiment 1, where the target sentences were presented word-by-word. The hypothesis that the mode of presentation was critical in determining negation effects on RTs was therefore corroborated. Only when subjects read the final sentence word-by-word (i.e., in Experiment 1), were they able to update their expectation about the correct sentence ending when encountering a negation marker. The relatively slow and gradual input method probably gave them enough time to do this: Typically, the final word didn't appear until approximately 1500 ms after the onset of *didn't* in word-by-word presentation. Then, upon encountering the final word, subjects verified the sentence in average time of 1034 ms. By contrast, subjects in the whole-sentence paradigm (Experiment 2) read *and* verified the target sentence in 1349 ms on average. That is, they took only little more than 300 ms extra time to read the entire sentence prior to (or while) making the verification decision. This was apparently not sufficient to allow for negation to be (completely) processed prior to making the verification decision.

In addition to mere timing, there were arguably differences in processing load associated with the two presentation modes. Subjects who were shown the final sentence word-by-word could presumably process and integrate all information prior to the final word before verifying the sentence. By contrast, for subjects in the whole-sentence paradigm the task consisted of reading and integrating all the critical parts of the sentence as well as verifying it at more or less the same time, and the resulting multiple-task situation could interfere with the initial processing of negation. Anticipating the correct ending of a negative sentence required retrieving the negative inference made on the basis of the bias sentence. This negative inference was probably less active in working memory than the affirmative option, because negated concepts tend to be suppressed or backgrounded (*cf.* section 2.2.2). If too much attention had to be devoted to the processing of additional information, information could be lost from working memory (Cowan & Morey, 2007), and the backgrounded negative information would be more affected than the affirmative one. Additionally, the multiple-task situation could negatively affect the processing of negation and the retrieval of the negative information, as divided attention or the simultaneous performance of more than one task result in performance decline (cf. Pashler, 1994; Tombu & Joelicoeur, 2003). Both the loss of the negative inference from memory and the failure to retrieve it would bring about the same effect: The only information to which the target sentence could be compared was the affirmative situation, and in this case, the predicted RT pattern is the one found in Experiment 2 as well as numerous sentence-picture verification studies (Carpenter & Just, 1975; Clark & Chase, 1972; Trabasso et al., 1971). As it affects the timing of target sentence processing as well as the resources available for it, the presentation mode could thus have a marked impact on verification times. To the extent that an advantage of true over false negative sentences depends on the possibility to anticipate the correct ending of a negative sentence, finding this advantage appears to require not only the availability of this negative alternative in the context, but also a setting that allows the subject to use that information during the processing of the target sentence.

Experiment 3: Word-by-Word Verification

The comparison of Experiments 1 and 2 demonstrated that the way in which target sentences are presented influences, if not determines the relative verification times for true and false negative sentences. Besides presentation mode, however, the two experiments also differed in other experimental variables. In Experiment 1, the verification task was embedded in an ERP paradigm, which required participants to wear an electrode cap as well as control their body and eye movements. This was likely to affect the amount of attention or cognitive capacity subjects could dedicate to processing and verifying the choice stories. In addition, subjects in the two experiments received different instructions: While subjects in the whole-sentence verification study were asked to respond as quickly as possible, correctness was emphasized in the ERP study, and subjects were not given any timing instructions. In order to draw definite conclusions about the impact of presentation mode, uncontaminated by other experimental variations, it was therefore necessary to conduct a study that differed from Experiment 2 only in the way stimuli are presented, but not in other task demands or instructions. This was the purpose of Experiment 3: Subjects completed a timed verification task without simultaneous EEG recordings, but the target sentence was presented word-by-word. It was expected that the RT pattern in this experiment would largely conform to that observed in Experiment 1, with shorter RTs to TN than FN. Furthermore, as the overall processing load in a verification only paradigm was probably lower than in the ERP setting, it was also likely that the retrieval of the negative inference would be significantly facilitated. If that was the case, the truth effect for negatives might be as strong as that for affirmative sentences.

Method

Subjects

Sixteen subjects (14 women) with a mean age of 20.5 years (range 18-24 years) participated for academic credit or a cash payment of $7/hour. All were right-handed native speakers of English with normal or corrected to normal vision and no history of neurological disorders.

Design and Stimuli

The materials from Experiments 1 and 2 were used here as well.

Procedure

Test administration, instructions, and practice paralleled the procedures described for Experiment 1 and 2.

As in Experiment 2, a trial began with a row of ten crosses that was presented for 1000 ms and followed by a 200 ms blank screen. Next, the subject read the introduction and bias in a self-paced manner; the two screens were separated by a 200 ms break. Starting 1200 ms after the offset of the bias sentence screen, the final sentence was presented word-by-word with an SOA of 500 ms and a stimulus duration of 200 ms. The screen remained blank after the offset of the last word until the subject made his decision. A new trial started after 1200 ms. Subjects typically completed the experiment in 25 to 40 minutes.

Data Analysis

The analysis of accuracy and response time data followed the strategy described for Experiment 1. Approximately 8% of trials were excluded from the statistical analyses because they were outliers or associated with an incorrect response.

Results

Mean scores on the ART and MRT were .169 (SD = .066) and .289 (SD = .127), respectively. Subject completed the neutral version of the Stroop in 35.3 s (SD = 4.7 s) on average. The took an average of 58.1 s (SD = 15.9 s) for the interference version, corresponding to a 65% increase. This sample of subjects was thus quite similar to the subjects who participated in Experiments 1 and 2.

Accuracy

With 95% correct responses, overall accuracy was comparable to Experiment 1, which also used word-by-word presentation. The right panel of Figure 9 (below) presents the rates of correct responses by sentence type. It shows that affirmative sentences were verified correctly more often than negative sentences (Wald z = -2.62, p = .009). Truth did not have a significant main effect on accuracy [Wald z = 1.19, p = .233], and it did not interact with negation, either [Wald z = 0.75, p = .451]. There was also no reliable linear increase or decrease of error rates over the course of the experiment [Wald z = 1.15, p = .252].

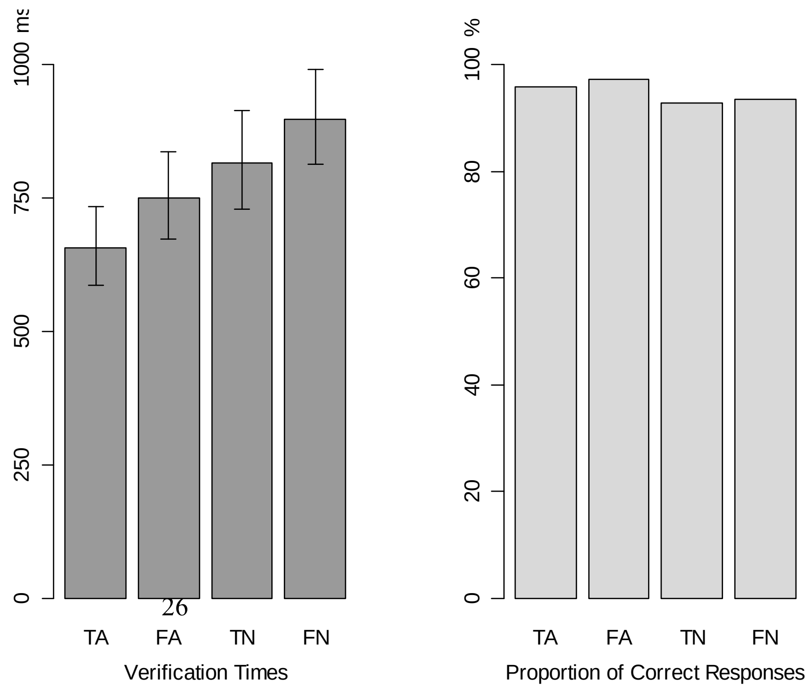

Figure 9. Verification times and accuracy in Experiment 3. The left panel shows mean response times with 95% confidence intervals. Means and standard errors were computed on by-subject averages of log-transformed data. Back-transformed values are shown. The right panel shows the proportion of correct responses computed over all subjects and items.

Response Times

The mixed-effects analysis revealed a significant decrease in RTs [$F(1, 1763) = 84.32$, $p < .001$] over the course of the experiment. There were also reliable independent effects of truth [$F(1, 1763) = 69.88$, $p < .001$] and negation [$F(1, 1763) = 213.28$, $p < .001$]. Affirmative sentences were verified faster than negative ones, and true sentences led to shorter RTs than false ones. The left panel of Figure 9 shows the pattern that was found in 13 out of 16 subjects: TA were verified faster than FA, and TN faster than FN. That is, the direction of the truth effect was the same for affirmatives and negatives. Additionally, the absence of a significant truth x negation interaction [$F(1, 1763) = 2.36$, $p = .125$] indicated that the size of the truth effect did not differ reliably between affirmative and negative sentences.

Discussion

As expected, the pattern of verification times in the current experiment was largely similar to that observed in Experiment 1, which also employed word-by-word presentation: For both affirmative and negative sentences, truth led to shorter RTs. This stands in contrast to the reversed truth effect observed in negative sentences in Experiment 2, where target sentences were presented as a whole. As the current experiment differed from Experiment 2 only in way the target sentence was presented, the difference in relative RTs to true and false negative sentences can clearly be attributed to this difference. Word-by-word presentation gave subjects sufficient time to update their expectations about the correct ending of negative sentences, which led to the RT advantage of true over false sentences.

At the same time, there were also differences between the RT results of the two experiments with word-by-word presentation. In Experiment 3, the truth effects in affirmative and negative sentences did not differ; there was no interaction between truth and negation. Subjects were equally good at anticipating the correct endings of affirmative and negative sentences. This was most likely possible because of the gradual presentation of the target sentence and the relatively low processing load with no interference from the experimental conditions necessitated by the ERP paradigm. By contrast, Experiment 1 produced a significant interaction between truth and negation, as the truth effect for affirmative sentences was larger than for negative sentences. That is, while the retrieval of the correct negative ending was *mostly* successful, it was not as consistent as that of the affirmative ending. With less capacity available for maintaining and retrieving the negative inference from the bias sentence, subjects were somewhat less likely to correctly adjust their expectation about the ending of a negative sentence.

General Discussion

The main goal of this series of experiments was to assess the processing consequences of negation within the same sentence in which the negation occurred and to which it applied. More specifically, we were interested in whether or not negation would affect the fit of a lexical item that occurred later in that negative sentence. An earlier study by Fischler and colleagues (1983) failed to find effects of negation on the processing of a sentence-final word whose plausibility in the sentence (in terms of rendering the sentence true or false) depended on the presence or absence of negation. The same word apparently was facilitated (as inferred from reduced N400 amplitudes) in both sentence modes, although it rendered the affirmative sentence true and the negative one false. These results were taken to imply that negation acted as an embedding operator that was processed only after the embedded affirmative proposition had been understood. We argued, however, that it was Fischler's use of isolated sentences that may have prevented detection of more immediate (intra-sentential) negation effects. Without a context that provided alternatives to be denied, the most plausible ending for a negative sentence was the same one expected for its affirmative counterpart. By contrast, in the current set of experiments (Experiments 1-3) the critical sentences were embedded in contexts such that the most plausible ending between the affirmative and the negative versions differed. We therefore hypothesized that the ending that rendered the corresponding sentence true would be facilitated in both affirmative and negative sentences, because we hypothesized that negation should affect which ending would be more expected.

Experiment 1 largely corroborated this hypothesis, as the N400 to true sentence endings was smaller than that to false endings; this effect was significant for affirmative sentences, but not quite for negative sentences across all participants. Importantly, the results showed that the fit of a word in the sentence *was* modulated by negation. Unlike in Fischler and colleagues' (1983) study, the same word was facilitated when it rendered the affirmative sentence true, but not when it was a false ending for the negative version of the otherwise same sentence. In most participants, the N400 was even larger to FN endings than to TN targets, while the same words showed the opposite pattern when they occurred as TA and FA, respectively, in affirmative sentences. A smaller group of participants, however, showed a pattern more similar to that in Fischler et al.: the same words were facilitated in both affirmative and negative sentences, with lower N400 amplitudes to TA compared to FA, and FN compared to TN. We note that the truth effect for the negative sentences in these two subject groups was correlated with aspects of their processing of the bias sentence. Participants who showed facilitation of TN compared to FN tended to take more time to process the bias sentence than participants with the opposite pattern. ERP differences between the groups also suggested that the processing of the bias sentence may have affected the processing of the targets as reflected in the pattern of target N400s.

Response times in Experiment 1 followed a pattern similar to that for N400 amplitudes, as sentences that elicited larger target N400s led to longer RTs. The RTs showed an additional main effect of negation with longer RTs to negative sentences, but the truth effects within each sentence mode paralleled those found for the N400. This was the case for the data from the entire sample as well as the two participant groups. The participants whose N400s were smaller to true than to false targets in both affirmatives and negatives also verified true sentences faster than false ones in both affirmative and negative sentences. Conversely, the participants with reversed truth effects for negatives on the N400 also tended to show longer RTs to TN than FN. Overall, the factors that determined N400 amplitude, i.e. the combination of sentence truth and mode as well as the processing of the bias sentence, appeared to have similar effects on the corresponding RTs.

Experiments 2 and 3 further demonstrated that RT patterns also depended on the manner in which the target sentence was presented. The advantage of TN over FN was only observed with word-by-word (not whole sentence) presentation, namely, in Experiments 1 and 3. By contrast, when the target sentence appeared in its entirety as in Experiment 2, truth effects were reversed between affirmatives and negatives, with shorter RTs to FN compared to TN. A slower, gradual presentation of the target sentence thus seemed to favor true endings in negative sentences, while a presentation mode that encouraged quick reading of the target sentence favored the ending that rendered the affirmative sentence true (and the negative sentence false). In addition, overall processing load appeared to affect the result patterns, as Experiment 3 produced equal truth effects in both sentences modes, while Experiment 1, which included the ERP procedure, resulted in a weaker truth effect for negative sentences.

In sum, negation can have an effect on the processing of lexical items in its scope. It can reduce the facilitation of a word that would fit in the affirmative version of that sentence and redirect attention to a concept/word that is more appropriate in the negative version. The extent to which these effects can be observed, however, depends strongly on experimental conditions as well as strategies employed by individuals. In the remainder of this discussion, we will attempt to provide a general account of the processes that give rise to the various data patterns observed in this series of experiments.

Classic models of picture-sentence verification (e.g., Carpenter & Just, 1975; Clark, 1976; Trabasso et al., 1971) proposed that response time differences were due to matches or mismatches between the representations of the sentence and picture being compared. For the current paradigm, the equivalent of picture and sentence are the expected sentence-final word and the one that was actually presented. An account of N400 effects, however, does not need to invoke an active serial comparison process typical of the classic semantic verification models. Instead, one can think of the effect as facilitation; the target word is primed by the preceding context and the inferences it affords. Indeed, the basic variable that is thought to give rise to differences in N400 amplitude to the target word is its fit within the context, or how expected a word is or to what extent it is primed by the preceding sentence and discourse context. The strong similarities between N400 and RT data in Experiment 1 suggest that both N400 amplitude and verification RTs are modulated by these priming processes. This is not surprising as primed words are generally processed faster, thereby speeding up RTs to the entire sentence as well. Of course, RTs are also affected by other variables, like the processing of negation itself, which slow verification times. This effect may be located in conscious decision making, which is not part of sentence processing per se, or it may be due to the construction of mental models of the sentence information, which most likely also occurs post-sentence (cf. Kaup, 2006; Kaup & Zwaan, 2003; Kaup, Zwaan, & Lüdtke, 2007). The focus of interest here, however, is the effect of negation on the word processing within the same sentence, i.e., the N400 and RT differences (within one sentence mode) that can be attributed to the fit between the expected and the actual target word. Given that the target words do not differ between conditions, the question is how expectations about the target vary across conditions.

Which word best completes the target sentence depends on the preceding discourse, especially the bias sentence. The bias sentence, which provides information about a character's preferences (e.g., *Joe wanted something salty.*), affords two inferences, one about the option that is likely to be chosen (*pretzels*) and, in turn by not-both elimination, one about the alternative that is not likely to be chosen (*not cookies*). Both inferences can (and perhaps should) be made routinely during reading (Lea & Mulligan, 2002). The result of one of the two inferences completes the sentence so as to render it true; it should therefore be more expected assuming sufficient time and resources. In the case of an affirmative sentence (*So he bought the…*), the affirmative inference (*pretzels*) constitutes the correct completion, while a negative sentence (*So he didn't buy the…*) would be completed correctly by the negative inference (*cookies*).

Under optimal conditions, i.e., if the inferences were made and kept accessible in working memory, and if sufficient time and processing capacity are available, participants should be able to adjust their expectations about the sentence ending and anticipate the content of the appropriate inference whether the target sentence is affirmative or negative. Consequently, true endings should have an equal advantage over false ones for both sentence types. This is, in fact, the result pattern observed for RTs in Experiment 3. In this experiment, target sentences were presented word-by-word, and participants were faced with a single task: to judge the consistency of the stories. Participants apparently had both the time and the processing resources available to retrieve the correct ending from working memory while reading the target sentence, whether it was affirmative or negative. As a consequence, the truth effect was of similar size in both sentence modes.

Under less optimal conditions, retrieving the correct ending may be more difficult. However, an increase in processing load or time pressure will have different effects in affirmative and negative sentences. In general, the affirmative inference is by default more activated than the negative one, as negation directs attention away from a concept to an alternative -- in this case from the negative inference to the affirmative one. If the target sentence is also affirmative, this default does not have to be changed, as the more activated affirmative inference is the correct ending that should be predicted. No particular processing resources are therefore needed in this case, and a higher load should have no effect. When the target sentence turns out to be negative, however, the expectation needs to be updated. This involves retrieving or activating the backgrounded negative inference, which may in fact require more time as well as processing capacity. Without enough time or resources, participants may either fail to retrieve the correct ending for the negative sentence on at least some trials, or activate it only partially on some or all trials. The FN ending, which is the word that would make the affirmative sentence correct, thus remains, at least some times or to some extent, activated. As a result, the TN will be less facilitated and the FN more facilitated. The truth effect in negatives will be smaller than in affirmatives, and it may even be reversed, with an advantage for FN over TN.

These changes in truth effect for negatives were observed in Experiments 1 and 2. In Experiment 2, where the target sentence was presented in its entirety, the truth effect was indeed reversed. As the presentation mode encouraged participants to read the sentence as quickly as possible, time pressure and processing demands increased. We suggest that participants failed to retrieve the correct ending for negative sentences before encountering the target word; as a result the FN was processed faster than the TN. In Experiment 1, the differences in both processing load and truth effect for negatives were more subtle. The target sentence was presented at the same rate as in Experiment 3, but the additional requirements of an ERP paradigm could easily have introduced extraneous task demands, thereby diverting processing resources from the main task, the comprehension and verification of the stories. Consistent with this relatively minor interference, the truth effect in negatives was not reversed, but smaller than in affirmatives. The advantage of TN over FN was not as big as that of TA over FA, which is to be expected if participants fail to retrieve the TN ending on some trials or if they do not completely shift the focus away from the affirmative inference to the correct negative ending.

Even when processing load is not particularly high, as in Experiment 1, some participants simply may not update their expectations about the target as a function of negation. As discussed above, this change in expectation can be viewed as the retrieval of the negative inference from the bias sentence, which tends to be backgrounded due to the suppressive effects of negation. Lea and Mulligan (2002) have shown that the result of such a negative inference may not receive or retain any additional activation.

Remembering the negative inference is useful in the current paradigm, however, as it allows for the prediction of the correct ending of negative sentences and presumably easier verification. Some participants may therefore focus some attention on the negative inference and on maintaining it in working memory. Others may choose not to do so or may be unable to do so. As a consequence, they would not activate the negative inference, making it difficult if not impossible for them to retrieve this information immediately upon encountering the negation marker. Like the participants in the whole-sentence presentation paradigm, these participants should therefore show a reversed truth effect for negatives. In Experiment 1, a subgroup of participants did indeed produce such a pattern for both N400 amplitude and RTs, although they read the stories under the same conditions as the other group, who showed an advantage of TN over FN. They also spent less time processing the bias sentence, which is consistent with the idea that they may not have activated the negative inference in the first place.

In sum, Experiment 1 results demonstrate that negation can lead to differences in the processing of upcoming information, and the experiments overall attest to the fact that the specific pattern of negation effects are impacted by experimental conditions as well as subject variables. The effect of negation on the processing of subsequent words in the sentence can be explained by changes in a people's expectations about what the upcoming words are likely to be. If the actual word matches the expectation, its processing will be facilitated. A change in expectation requires the activation of a suitable alternative, which in the case of negation will usually have to be derived from the discourse context. And, if it is derivable in principle, it must also be accessible to the participant in practice. That is, the appropriate inference has to be made and kept active in memory, and the reader must have sufficient time and processing resources to retrieve the information. Given all these conditions, a negation marker should cause a language user to change his or her expectations about upcoming subsequent sentence elements. The experimental results presented here have established the possibility of detecting these intra-sentential negation effects via behavioral and brain responses. They warrant the conclusion that the processing of negation is not, or at least is not always, delayed until after the end of the negated sentence.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247-264.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials associated with semantic priming. *Electroencephalography and Clinical Neurophysiology, 60*, 343-355.

Bles, M., Alink, A., & Jansma, B. M. (2007). Neural aspects of cohort-size reduction. *Brain Research, 1150*, 143-154.

Bretz, F., Hothorn, T., & Westfall, P. H. (2002). On multiple comparisons in R [Electronic Version]. *R News*, 2(3), 14-17. Retrieved 5/7/2007 from http://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf.

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic model of verification. *Psychological Review, 82*, 45-76.

Carpenter, P. A., Just, M. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1999). Time course of fMRI-activation in language and spatial networks during sentence comprehension. *NeuroImage, 10*, 216-224.

Clark, H. H. (1976). *Semantics and comprehension*. The Hague: Mouton.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3*, 472-517.

Cowan, N., & Morey, C. C. (2007). How can dual-task working memory retention limits be investigated? *Psychological Science, 18*, 686-688.

Crocker, M. W., & Brants, T. (2000). Wide-coverage probabilistic sentence processing. *Journal of Psycholinguistic Research, 29*, 647-669.

De Mey, M. (1971). Negation and attention. *Communication & Cognition, 4*, 187-197.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117-1121.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 a manifestation of context updating? *Behavioral and Brain Sciences, 11*, 357-374.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology, 14*, 456-467.

Faustmann, A., Murdoch, B. E., Finnigan, S. P., & Copland, D. A. (2005). Event-related brain potentials elicited by semantic and syntactic anomalies during auditory sentence processing. *Journal of the American Academy of Audiology, 16*, 708-725.

Federmeier, K., Wlotko, E. W., De Ochoa-Ewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146*, 75-84.

Felser, C., Clahsen, H., & Münte, T. F. (2003). Storage and integration in the processing of filler-gap dependencies. *Brain and Language, 87*.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. *Journal of Psycholinguistic Research, 30*, 321-338.

Fillmore, C. (1985). Frames and the semantics of understanding. *Quaderni de Semantica, 6*, 222-254.

Fischler, I., Bloom, P. A., Childers, D. G., Arroyo, A. A., & Perry, N. W. J. (1984). Brain potentials during sentence verification: late negativity and long-term memory strength. *Neuropsychologia, 22*, 559-568.

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W. J. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology, 20*, 400-409.

Fischler, I., Childers, D. G., Achariyapaopan, T., & Perry, N. W. J. (1985). Brain potentials during sentence verification: automatic aspects of comprehension. *Biological Psychology, 21*, 83-105.

Frege, F. L. G. (1884). *Grundlagen der Arithmetik: Eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Breslau: Verlag von Wilhelm Koebner.

Frege, F. L. G. (1919). Die Verneinung. *Beiträge zur Philosophie des deutschen Idealismus, I*, 143-157.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1*, 183-192.

Giora, R. (2006). Anything negatives can do affirmatives can do just as well, except for some metaphors. *Journal of Pragmatics, 39*, 981-1014.

Giora, R. (2007). "A good Arab is not a dead Arab - a racist incitement": On the accessibility of negated concepts. In I. Kecskes & L. R. Horn (Eds.), *Explorations in Pragmatics: Linguistic, Cognitive, and Intercultural Aspects* (pp. 129-162). Berlin: Mouton de Gruyter.

Giora, R., Balaban, N., Fein, O., & Alkabets, I. (2004). Negation as positivity in disguise. In H. L. Colston & A. Katz (Eds.), *Figurative language comprehension: Social and cultural influences* (pp. 233-258). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Giora, R., Fein, O., Aschkenazi, K., & Alkabets-Zlozover, I. (2007). Negation in context: A functional approach to suppression. *Discourse Processes, 43*, 153-172.

Givón, T. (1979). *On understanding grammar*. New York: Academic Press.

Givón, T. (1984). *Syntax: A functional-typological introduction* (Vol. I). Amsterdam, Philadelphia: John Benjamins.

Givón, T. (1989). *Mind, code and context: Essays in Pragmatics*. Hillsdale, NJ: Erlbaum.

Givón, T. (1993). *English grammar: a function-based introduction* (Vol. I). Amsterdam, Philadelphia: John Benjamins.

Glenberg, A. M., Robertson, D. A., Jansen, J. L., & Johnson-Glenberg, M. C. (1999). Not propositions. *journal of Cognitive Systems Research, 1*.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95-112.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia, 11*, 1518-1530.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika, 75*, 800-802.

Horn, L. R. (1989). *A Natural History of Negation*. Chicago: University of Chicago Press.

Jespersen, O. (1917). *Negation in English and other languages*. Copenhagen: A. F. Høst.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory & Language, 49*, 133-156.

Kaup, B. (1997). The processing of negatives during discourse comprehension. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Conference of the Cognitive Science Society* (pp. 370-375). Mahwah, NJ: Erlbaum.

Kaup, B. (2001). Negation and its impact on the accessibility of text information. *Memory & Cognition, 29*, 960-967.

Kaup, B. (2006). What psycholinguistic negation research tells us about the nature of the working memory representations utilized in language comprehension. In H. Pishwa (Ed.), *Language and Memory: Aspects of Knowledge Representation* (pp. 313-355). Berlin: Mouton de Gruyter.

Kaup, B., Lüdtke, J., & Burkert, M. (2006, August). *Effects of polarity in language comprehension: Is negation really an availability reducing operator?* Poster presented at the AMLaP Conference 2006, University Nijmegen, The Netherlands.

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is that door that is not open mentally closed? *Journal of Pragmatics, 38*, 1033-1050.

Kaup, B., & Zwaan, R. A. (2003). Effects of negation and situational presence on the accessibility of text information. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29*, 439-446.

Kaup, B., Zwaan, R. A., & Lüdtke, J. (2007). The experiential view of language comprehension: How is negation represented? In F. Schmalhofer & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes* (pp. 255-288). Mahwah, NJ: Lawrence Erlbaum.

Kim, K. (1985). Development of the concept of truth-functional negation. *Developmental Psychology, 21*, 462-472.

King, J. W., & Kutas, M. (1995). Who did what when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7*, 376-395.

Klima, E. S. (1964). Negation in English. In J. A. Fodor & J. J. Katz (Eds.), *The structure of language*. Englewood Cliffs, NJ: Prentice-Hall.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology, 38*, 557-577.

Kounios, J., & Holcomb, P. J. (1992). Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General, 121*, 459-479.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience, 15*.

Kutas, M., Federmeier, K., Staab, J., & Kluender, R. (2007). Language. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (3rd ed., pp. 555-580). New York: Cambridge University Press.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 207*, 203-205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161-163.

Lea, R. B., & Mulligan, E. J. (2002). The effect of negation on deductive inferences. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 28*, 303-317.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology, 31*, 291-308.

Lüdtke, J., Friedrich, C. K., De Filippis, M., & Kaup, B. (2008). Event-related potential correlates of negation in a sentence-picture verification paradigm. *Journal of Cognitive Neuroscience, 20*, 1355-1370.

Lüdtke, J., & Kaup, B. (2006). Context effects when reading negative and affirmative sentences. In R. Sun (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1735-1740). Mahwah, NJ: Lawrence Erlbaum.

MacDonald, M. C., & Just, M. A. (1989). Changes in activation levels with negation. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*, 633-642.

Melara, R. D., Rao, A., & Tong, Y. (2002). The duality of selection: Excitatory and inhibitory processes in auditory selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 279-306.

Münte, T. F., Heinze, H.-J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: no evidence for specificity of the syntactic positive shift. *Neuropsychologia, 36*, 217-226.

Münte, T. F., Schlitz, K., & Kutas, M. (1998). When temporal terms belie conceptual order: an electrophysiological analysis. *Nature, 395*, 71-73.

Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth isn't too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, *19*, 1213 - 1218

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97-113.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*, 220-244.

Salmon, N., & Pratt, H. (2002). A comparison of sentence- and discourse-level semantic processing: An ERP study. *Brain and Language, 83*, 367-383.

Sherman, M. A. (1976). Adjectival negation and the comprehension of multiply negated sentences. *Journal of Verbal Learning & Verbal Behavior, 15*, 143-157.

Staab, J. (2007). *Negation in context: Electrophysiological and behavioral investigations of negation effects in discourse processing.* Unpublished doctoral dissertation, San Diego State University & University of California, San Diego.

Stanovich, K. E., & Cunningham, A. (1992). Studying the consequences of literacy within a literate society: the cognitive correlates of print exposure. *Memory & Cognition, 20*, 51-68.

Stanovich, K. E., & Cunningham, A. (1993). Where does knowledge come from? Specific associations between print exposure and information acquisition. *Journal of Educational Psychology, 85*, 211-229.

Stanovich, K. E., West, R. F., & Harrison, M. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology, 31*, 811-826.

Strawson, P. F. (1952). *Introduction to logical theory*. London: Methuen.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 28*, 643-662.

Tombu, M., & Joelicoeur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 3-18.

Trabasso, T., Rollins, H., & Shaughnessy, E. (1971). Storage and verification stages in processing concepts. *Cognitive Psychology, 2*, 239-289.

van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence fromERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 31*, 443-367.

van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: evidence from the N400. *Journal of Cognitive Neuroscience, 11*, 657-671.

van Berkum, J. J. A., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research, 1146*, 158-171.

van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research, 17*, 701-718.

Van Petten, C., Weckerly, J., McIsaac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science, 8*, 238-242.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology, 47A*, 631-650.

Wason, P. C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology, 11*, 92-107.

Wason, P. C. (1961). Response to affirmative and negative binary statements. *British Journal of Psychology, 52*, 133-142.

Wason, P. C. (1965). The contexts of plausible denial. *Journal of Verbal Learning & Verbal Behavior, 4*, 7-11.

Westfall, P. H., Tobias, R. D., Rom, D., Wolfinger, R. D., & Hochberg, Y. (1999). *Multiple comparisons and multiple tests using the SAS system*. Cary, NC: SAS Institute Inc.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*, 1271-1288.